



Universidad de Valladolid



**ESCUELA DE INGENIERÍAS
INDUSTRIALES**

UNIVERSIDAD DE VALLADOLID

ESCUELA DE INGENIERIAS INDUSTRIALES

Grado en Ingeniería de organización industrial

Diagnóstico de fallos en motores con técnicas de Machine Learning.

Autor:

Gordo Martín, Daniel

Tutores:

Fernández Temprano, Miguel.

García Escudero, Luis Ángel.

**Departamento de Estadística e
Investigación Operativa.**

Valladolid, 09/2014.

“Anyone who fights for the future, lives in it today”

Ayn Rand

ÍNDICE DEL TFG

RESUMEN	1
ABSTRACT	1
ESTRUCTURA DE TFG.....	3

CAPÍTULO 1.- INTRODUCCIÓN Y OBJETIVOS 5

1.1 INTRODUCCIÓN Y OBJETIVOS	7
1.2 HERRAMIENTAS UTILIZADAS	8
1.2.1 <i>StatGraphics</i>	8
1.2.2 <i>R</i>	9

CAPÍTULO 2.- MOTORES DE INDUCCIÓN..... 11

2.1 INTRODUCCIÓN	13
2.2 PRINCIPIO DE FUNCIONAMIENTO.....	13
2.3 TIPOS DE ROTOR.....	15
2.3.1 <i>Jaula de ardillas</i>	15
2.3.2 <i>Rotor bobinado</i>	16
2.4 CLASIFICACIÓN DE MOTORES COMERCIALES	16
2.5 FALLOS EN LOS MOTORES DE INDUCCIÓN	17
2.5.1 <i>Fallos del estator</i>	18
2.5.2 <i>Fallos del rotor</i>	21

CAPÍTULO 3.- MACHINE LEARNING 25

3.1 Definición	27
3.2 Tipos de problemas	28
3.2.1 <i>Clasificación</i>	28
3.2.2 <i>Regresión</i>	29
3.2.3 <i>Clustering</i>	29
3.2.4 <i>Asociación</i>	29
3.3 Validación del modelo	29
3.4 Sobreajuste del modelo	31

CAPÍTULO 4.- EXPERIMENTACIÓN33

4.1 DESCRIPCIÓN DE LA EXPERIMENTACIÓN	35
4.2 DESCRIPCIÓN DEL BANCO DE ENSAYOS	36
4.2.1 Motor	37
4.2.2 Freno	38
4.2.3 Sensores de velocidad	39
4.2.4 Transductores.....	40
4.2.5 Variadores de frecuencia.....	43
4.2.6 Sistema de adquisición de datos.....	43
4.3 CARACTERÍSTICAS MEDIDAS	44
4.3.1 Relativas a la carga del motor.....	44
4.3.2 En el dominio del tiempo	45
4.3.3 En el dominio de la frecuencia.....	48

CAPÍTULO 5.- TÉCNICAS UTILIZADAS51

5.1 INTRODUCCIÓN.....	53
5.2 ANÁLISIS DE COMPONENTES PRINCIPALES.....	53
5.2.1 Obtención de la primera componente.....	53
5.2.2 Obtención de la segunda y resto de componentes	54
5.2.3 Criterio de la media aritmética.....	55
5.3 ANÁLISIS DISCRIMINANTE	56
5.4 MÉTODO DE LOS K VECINOS MÁS PRÓXIMOS (KNN).....	57
5.4.1 Variantes de KNN	58
5.4.1.1 KNN por distancia media.....	58
5.4.1.1 Distancia mínima	58
5.4.1.4 Pesado de casos	58
5.4.1.5 Métrica	58
5.5 SUPPORT VECTOR MACHINE (SVM)	59
5.5.1 Caso separable linealmente.....	60
5.5.2 Caso no separable linealmente.....	61
5.5.3 Kernel.....	62
5.5.4 SVM multiclase	63

CAPÍTULO 6.- ANÁLISIS DESCRIPTIVO DE LOS DATOS65

6.1 ANÁLISIS DE LOS DATOS.....	67
6.1.1 Correlaciones.....	67
6.1.2 Normalidad	68

6.1.3 <i>Análisis de componentes principales</i>	70
6.2 ANÁLISIS DE LOS ENSAYOS	72

CAPÍTULO 7.- ANÁLISIS MULTIVARIANTE DE LOS DATOS 77

7.1 INTRODUCCIÓN	79
7.2 DISCRIMINACIÓN R12-R34	79
7.3 DISCRIMINACIÓN R1-R234	81
7.4 DISCRIMINACIÓN R123-R4	83
7.5 DISCRIMINACIÓN R1-R2-R3-R4	85
7.6 ERROR ESTIMADO LDA	88

CAPÍTULO 8.- APLICACIÓN DE LOS ALGORITMOS 91

8.1 INTRODUCCIÓN	93
8.2 K VECINOS MÁS PRÓXIMOS	93
8.2.1 <i>Modelo y error aparente KNN</i>	93
8.2.2 <i>Error estimado KNN</i>	97
8.3 SUPPORT VECTOR MACHINE (SVM)	106
8.3.1 <i>Error estimado SVM</i>	106
8.3.2 <i>Modelo y error aparente SVM</i>	110

CAPÍTULO 9.- CONCLUSIONES 113

9.1 CONCLUSIONES	115
------------------------	-----

ANEXOS 117

A.1. INTRODUCCIÓN	119
A.2. CÓDIGOS	119

BIBLIOGRAFÍA 125

RESUMEN

En el presente trabajo se ha estudiado la aplicación de dos técnicas de machine Learning, K-Nearest Neighbour y Support Vector Machine, al problema del diagnóstico de la rotura de una barra del rotor de un motor de inducción jaula de ardillas con el objetivo de construir varios clasificadores capaces de distinguir entre diferentes estados de rotura en una de las barras midiendo una serie de características. Adicionalmente este trabajo permite comparar el comportamiento de estas técnicas de clasificación de aparición relativamente reciente y que están siendo incorporadas en multitud de aplicaciones prácticas como la que aquí se describe.

Los métodos han sido entrenados a través de una base de datos obtenidos a través de la experimentación con un motor real sometido a diferentes grados de rotura en una de sus barras.

PALABRAS CLAVES

K-Nearest Neighbor, Support Vector Machine, rotura de barra, motor de inducción jaula de ardillas, machine learning.

ABSTRACT

In this paper it has been studied the application of two machine learning methods, K-Nearest Neighbour y Support Vector Machine, to the problem of the diagnosis of a broken bar in a squirrel cage induction motor with the pursuit of build some classifiers able to distinguish between different breakage statements in one of the bars by measuring some features. In addition this work allows contrasting the performance of these relatively recent appearance classification techniques and that are being incorporated in many practical applications.

The methods have been trained through a database precedent from experimentation with a real motor subjected to different degrees of failure in one of its bars.

KEYWORDS

K-Nearest Neighbor, Support Vector Machine, broken bar, squirrel cage induction motor, machine learning.

ESTRUCTURA DEL TFG

Se ha dividido el presente trabajo en los siguientes capítulos:

- **Capítulo 1: Justificación**
Se explica la idea general del TFG y su justificación, junto con una breve descripción de las herramientas utilizadas.
- **Capítulo 2: Motores de inducción**
Se realiza una introducción a los motores de inducción: las partes de que se compone, sus principios de funcionamiento, clasificación de motores y algunos de los fallos por los que se pueden ver afectados.
- **Capítulo 3: Machine Learning**
Machine learning, tipos de problemas que se pueden solucionar con estas técnicas y métodos de validación de los modelos.
- **Capítulo 4: Experimentación**
En este capítulo se analiza la experimentación llevada a cabo en el laboratorio y cada una de las partes que forman el banco de ensayos utilizado para recoger los datos con los que se ha trabajado.
- **Capítulo 5: Técnicas utilizadas**
Se describen las diferentes técnicas que se han utilizado a lo largo del trabajo.
- **Capítulo 6: Análisis descriptivo de los datos**
Análisis descriptivo de la base de datos y de los ensayos.
- **Capítulo 7: Análisis multivariante**
Análisis multivariante de los datos para seleccionar las variables más explicativas, usando el discriminante lineal.
- **Capítulo 8: Aplicación de KNN y SVM**
Aplicación de los métodos KNN y SVM.
- **Capítulo 9: Conclusión**
Exposición de los resultados finales y conclusión.
- **Anexos:**
Parte de los códigos utilizados en R.

Capítulo 1

Introducción y objetivos

1.1. INTRODUCCIÓN Y OBJETIVOS

La idea de este TFG es la de utilizar una base de datos proporcionados por el Departamento de Ingeniería Eléctrica y procedentes de una experimentación, que comentaré en ulteriores apartados, para construir varios clasificadores que permitan determinar la “salud” de las barras de un motor de inducción eléctrico tipo jaula de ardillas.

Para ello se aplicarán dos técnicas estadísticas de machine learning al problema de clasificar el estado de un motor eléctrico midiendo una serie de características. Las técnicas elegidas son: los vecinos más próximos (K-Nearest Neighbor) y las máquinas de vectores de soporte (Support Vector Machine).

En la tabla 1.1 se muestran las características que se utilizaron para entrenar los clasificadores, que serán explicadas con más detalle en el capítulo 4.

	Características
En el dominio del tiempo	Skewness
	Momento
	Curtosis
	Cumulante
	Valor cuadrático medio
	Pico absoluto
	Promedio de valores absolutos
	Valor eficaz o RMS
	Factor Cresta
	Factor Forma
En el dominio de la frecuencia	Banda superior 1
	Banda superior 2
	Banda inferior 1
	Banda inferior 2
	Banda superior 5
	Banda inferior 5
	Banda superior 7
	Banda inferior 7

Tabla 1.1. Características.

Además de la construcción de los clasificadores este trabajo permite comparar el comportamiento de estos dos métodos con el del análisis discriminante lineal, que es otra técnica muy utilizada por su simpleza y su

menor coste computacional, para dar una idea de hasta qué punto mejoran sus resultados en aplicaciones como esta.

Por otra parte, el objeto de este trabajo es de suma importancia práctica ya que los motores eléctricos de inducción son ampliamente utilizados en la industria, representan el 85% del total de las cargas eléctricas, y tienen múltiples usos: bombas, ventiladores, compresores, cintas transportadoras, aplicaciones en barcos, trenes eléctricos, etc. Se hace imprescindible el estudio de diferentes métodos para conocer el estado de los mismos sin necesidad de pararlo y llevarlo a un taller especializado para estudiarlo con la problemática económica que esto supone.

1.2. HERRAMIENTAS UTILIZADAS

1.2.1. Statgraphics

Uno de los programas utilizados para la realización del trabajo ha sido StatGraphics, especializado en el análisis estadístico de datos, un software bastante conocido, intuitivo y relativamente sencillo de manejar.

Algunas de las acciones que permite son: análisis descriptivo de una o varias variables, intervalos de confianza, contraste de hipótesis, análisis de regresiones, análisis multivariantes o técnicas para el control de calidad.

La versión más reciente, la que ha sido utilizada, se llama StatGraphics Centurion XVI, cuya interfaz se muestra en la siguiente figura. Para aprender su manejo se ha utilizado [PER13].

	Alimentacion	Estado	Estado2	Estado3	Estado4	Frecuencia	Velocidad	Tension	Corriente	Deslizamiento	Velocidad dc	BandInf	Bar
1	Red	R1	R1	R12	R123	50	1441	233,5	2,46	0,038285714	1442,5714-69,52977-69		
2	Red	R1	R1	R12	R123	50	1491	233	2,03	0,004571429	1493,1428-57,53272-52		
3	Red	R1	R1	R12	R123	50	1448	233,5	2,33	0,033285714	1420,0714-69,07949-70		
4	Red	R1	R1	R12	R123	50	1460	234,4	2,21	0,026571429	1460,1428-69,07893-73		
5	Red	R1	R1	R12	R123	50	1480	233,7	2,03	0,013714286	1479,4285-69,02093-68		
6	Red	R1	R1	R12	R123	50	1493	233,9	2,02	0,004285714	1493,5714-69,47553-49		
7	Red	R1	R1	R12	R123	50	1439	233,9	2,52	0,039571429	1440,6428-70,38024-73		
8	Red	R1	R1	R12	R123	50	1445	233,6	2,43	0,036428571	1445,3571-67,08811-69		
9	Red	R1	R1	R12	R123	50	1460	233,4	2,19	0,02687143	1459,7142-66,05824-69		
10	Red	R1	R1	R12	R123	50	1467	233	2,09	0,021285714	1468,0714-65,34239-65		
11	Red	R1	R1	R12	R123	50	1489	232,6	2,01	0,006142857	1490,7857-69,12138-58		
12	Red	R1	R1	R12	R123	50	1469	232,5	2,06	0,01987143	1470,2142-64,00064-64		
13	Red	R1	R1	R12	R123	50	1437	233,6	2,51	0,041428571	1437,8571-72,35172-70		
14	Red	R1	R1	R12	R123	50	1447	233,4	2,35	0,03587143	1446,2142-66,36224-70		
15	Red	R1	R1	R12	R123	50	1456	233,5	2,19	0,026338391	1457,5714-69,84014-66		
16	Red	R1	R1	R12	R123	50	1467	233,7	2,09	0,020326367	1466,5714-64,27354-64		
17	Red	R1	R1	R12	R123	50	1468	233,4	2,06	0,020142857	1469,7857-61,62624-61		
18	Red	R1	R1	R12	R123	50	1492	233,5	2,02	0,005	1492,5-57,56737-53		
19	Red	R1	R1	R12	R123	50	1444	234,5	2,46	0,03787143	1443,2142-66,04944-66		
20	Red	R1	R1	R12	R123	50	1436	234,5	2,49	0,041571429	1437,6428-69,40611-72		
21	Red	R1	R1	R12	R123	50	1463	235,1	2,19	0,023285714	1465,0714-63,75772-66		
22	Red	R1	R1	R12	R123	50	1494	230,5	1,94	0,003571429	1494,6428-49,79486-30		
23	Red	R1	R1	R12	R123	50	1450	230,6	2,24	0,032571429	1451,1428-71,86787-68		
24	Red	R1	R1	R12	R123	50	1469	230,5	1,99	0,021428571	1467,8571-67,36392-68		
25	Red	R1	R1	R12	R123	50	1439	231,2	2,48	0,03987143	1440,2142-73,85656-71		
26	Red	R1	R1	R12	R123	50	1450	231,3	2,3	0,033428571	1449,8571-70,26413-75		
27	Red	R1	R1	R12	R123	50	1455	231,4	2,19	0,028714286	1456,9285-70,31246-71		
28	Red	R1	R1	R12	R123	50	1469	231,6	2,03	0,020285714	1469,5714-64,37282-61		
29	Red	R1	R1	R12	R123	50	1493	231,8	1,97	0,004285714	1493,5714-57,57074-54		
30	Red	R1	R1	R12	R123	50	1467	232,1	1,99	0,021571429	1467,6428-69,23787-69		
31	Red	R1	R1	R12	R123	50	1432	228,5	2,43	0,044428571	1433,3571-64,88151-67		
32	Red	R1	R1	R12	R123	50	1438	228,6	2,34	0,04	1440-74,93333-74		

Figura 1.1. Captura del espacio de trabajo de Statgraphics.

1.2.2. R.

Para el análisis en profundidad de la base de datos y la resolución del problema se ha utilizado un software estadístico llamado R. Es un software que permite efectuar análisis estadísticos en general más avanzados que los que se realizan con Statgraphics. Este programa es manejado a través de una consola en la que se introduce el código oportuno de su propio lenguaje para realizar las operaciones que queramos, figura 1.2.

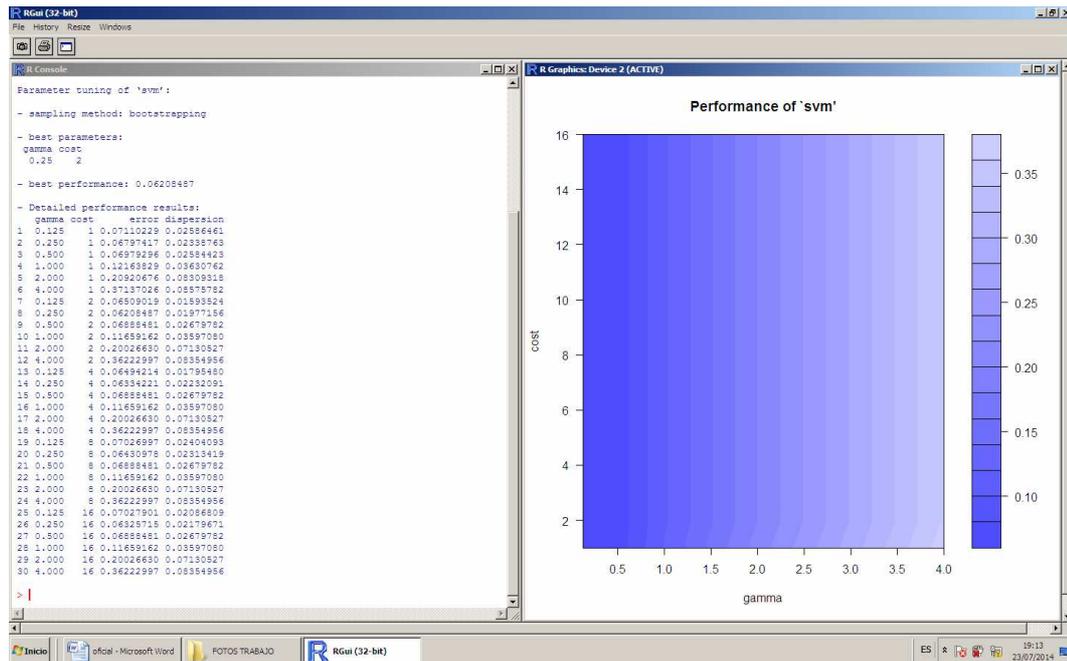


Figura 1.2. Captura del espacio de trabajo de R.

R fue inicialmente diseñado por Robert Gentleman y Ross Ihaka, miembros del Departamento de Estadística de la Universidad de Auckland, en Nueva Zelanda. Aunque R se considera la versión libre de otro programa, llamado S, desarrollado por los Laboratorios Bell. La principal característica de este programa es que es de software libre y gracias a eso hay un gran número de personas que trabajan voluntariamente en su desarrollo, por ejemplo, creando algoritmos que puedan ser usados por todo el mundo. La página web principal de R se llama <http://cran.r-project.org/>, en ella se pueden encontrar gran cantidad de manuales como el que se ha usado [PAR10], información y se pueden descargar los paquetes de trabajo.

Para el lector interesado en aprender su manejo se recomienda el uso del curso “Data Analysis” de la plataforma COURSERA.ORG.

Capítulo 2

Motores de inducción

2.1. INTRODUCCIÓN

Los motores de inducción son ampliamente utilizados en la industria debido a su versatilidad, poco peso, robustez, su elevado ratio potencia/peso y su precio.

Esta clase de motores, como se comentó en la introducción, representan alrededor del 85% del total de las cargas eléctricas [DUQ05].

Son parte crítica de muchas plantas industriales por lo que se hace especialmente importante las investigaciones sobre el mantenimiento y el diagnóstico de los mismos para evitar, por ejemplo, parones en una cadena de producción que puedan acarrear coste económico. Debido a esto desde hace décadas se viene investigando en estos campos con el objetivo de reducir los costes de mantenimiento y las pérdidas de tiempo por averías.

Aunque el presente trabajo no es eminentemente sobre ingeniería eléctrica sino que entra más en el campo de la estadística se ha investigado sobre estos motores y sus características.

2.2. PRINCIPIO DE FUNCIONAMIENTO

El motor de inducción o motor asíncrono (ver figura 2.1 y 2.2) está formado básicamente por dos partes, el estator y el rotor. El estator es la parte fija y el rotor, que se encuentra en el interior del estator, es la que gira, lo hace en torno a un eje que pasa por su centro y que está apoyado en unos rodamientos para facilitar el giro. Los dos están formados por devanados fabricados con algún tipo de conductor eléctrico y protegidos por aislantes. El aire que se encuentra entre el rotor y el estator se denomina entrehierro, además sobre el estator se coloca una carcasa para proteger los devanados.

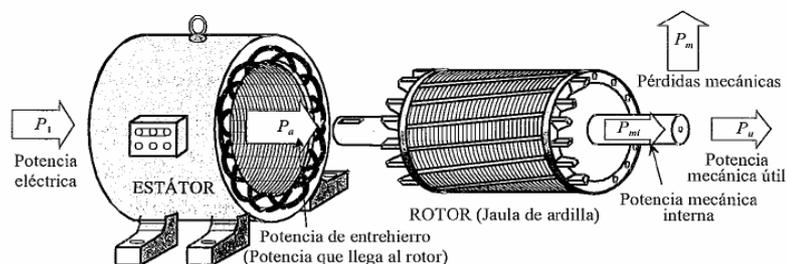


Figura 2.1. Esquema de un motor de inducción con rotor jaula de ardillas [FRA08].



Figura 2.2. Corte de un motor de inducción.

El funcionamiento es simple [FRA08], los devanados del estator son alimentados por corriente alterna sinusoidal, esta corriente crea un campo magnético que atraviesa los conductores del devanado del rotor, induciendo una corriente en este. Debido a la ley de Lenz, en el rotor se generan fuerzas electromotrices que va a tender a seguir al campo giratorio del estator, por lo que el rotor comienza a girar en el mismo sentido.

Para que esto ocurra tiene que existir un movimiento relativo entre el campo y los conductores inducidos, por lo tanto, el rotor nunca va a alcanzar la misma velocidad que el campo (velocidad síncrona), porque si lo hiciera se anularían las fuerzas electromotrices. Dependiendo del par resistente al que se vea sometido el rotor su velocidad se incrementara más o menos. Por ejemplo, en el vacío la velocidad del rotor sería casi igual. La figura 2.3 muestra las diferentes fuerzas electromagnéticas que actúan sobre el rotor de un motor jaula de ardillas.

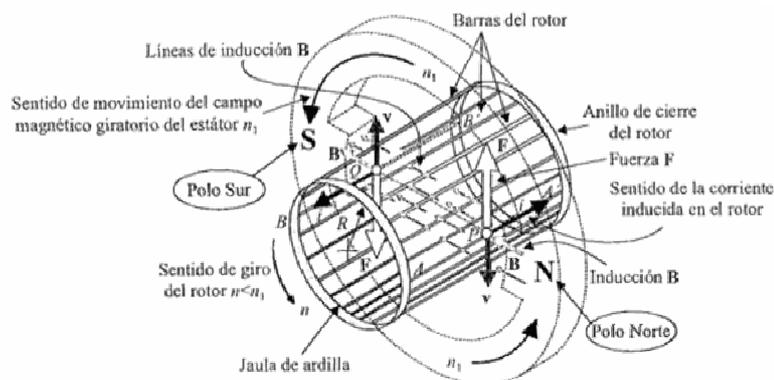


Figura 2.3. Campos magnéticos que actúan sobre el rotor jaula de ardillas [FRA08].

2.3. TIPOS DE ROTOR

Los rotores tienen dos configuraciones básicas [DUQ05] que dan lugar a dos tipos diferentes de motor de inducción, el rotor de jaula de ardillas y el rotor bobinado.

2.3.1. Jaula de ardillas:

De este tipo es el rotor del motor utilizado para la realización de los experimentos de este trabajo, está formado por una serie de barras dispuestas de tal manera que forman un cilindro y que están unidas por sus extremos a unos anillos, que se denominan anillos de cortocircuito, como indican las figuras 2.4 y 2.5.

Esta configuración se utiliza sobre todo en motores pequeños y medianos siendo la más frecuente en la industria ya que es: más fácil de construir, tiene menor peso, menor volumen, menos coste y es más robusto.

Su desventaja principal es que para determinados rangos es difícil controlar su velocidad y se hace necesario el uso de equipos electrónicos, como los variadores de tensión.

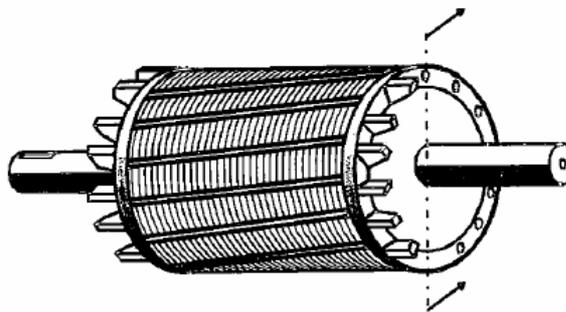


Figura 2.4. Rotor Jaula de ardillas [FRA08].



Figura 2.5. Rotor Jaula de ardillas de 33 barras [JAR09].

2.3.2. Rotor bobinado

El rotor bobinado, figura 2.6, está formado por un devanado trifásico distribuido uniformemente. Los terminales de este devanado van conectados a unos anillos rozantes que giran con el rotor y van conectados a unas escobillas, como indica la figura 2.7.

Su principal desventaja es que precisa un mantenimiento mayor que el rotor de jaula de ardillas, además es más caro y voluminoso.

Por otra parte, tiene como ventajas que se le pueden modificar desde el exterior algunas características como el par de arranque.

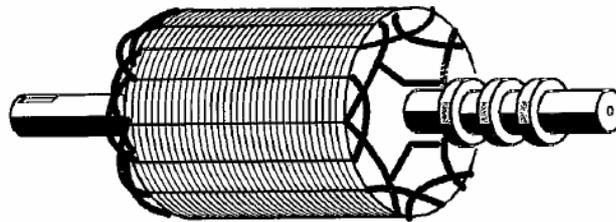


Figura 2.6. Esquemas Rotor bobinado [FRA08].

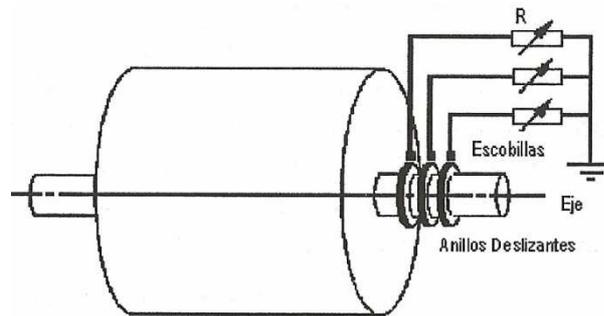


Figura 2.7. Esquemas Rotor bobinado [DUQ05].

2.4. CLASIFICACIÓN DE LOS MOTORES COMERCIALES

La norma CEI-60034-12 establece una clasificación para los motores de jaula de ardillas atendiendo a características del par y de la corriente.

- **Categoría N:** Par y corriente de arranque normal, bajo deslizamiento. Constituyen la mayor parte de los motores comerciales.
- **Categoría H:** Par de arranque alto mientras que la corriente de arranque es normal, bajo deslizamiento. Se usan para cargas con inercias grandes que requieren un par de arranque elevado.
- **Categoría D:** Par de arranque alto, corriente de arranque normal y deslizamiento mayor del 5%.

Otra clasificación es la que hace la "National Electrical Manufacturers Association".

- **Clase A:** motores con par de arranque y corriente de arranque elevados, buena regulación de velocidad, par máximo alto y factor de potencia elevado.
- **Clase B:** Par de arranque alto y menor corriente de arranque los de clase A. A esta clase pertenecen la mayor parte de los motores comerciales de jaula de ardilla.
- **Clase C:** alto par de arranque. Corriente de arranque inferior a los A y a los B. Mala regulación. Su uso se limita a aplicaciones con difíciles arranques.
- **Clase D:** Diseño apropiado para servicios con arranques pesados, pero con potencias en régimen relativamente baja. No se recomienda para aplicaciones con múltiples arranques debido a su pequeña sección transversal y su mala disipación del calor.

2.5. FALLOS EN LOS MOTORES DE INDUCCIÓN

A pesar de la gran robustez y fiabilidad de estos motores, al igual que otras máquinas industriales están sujetos a diversos fallos que normalmente son consecuencia del sometimiento a esfuerzos mayores que aquellos para los que están diseñados y que provocan una reducción de la vida útil del motor o averías graves.

Estudios realizados por la Electric Power Research Institute (EPRI) y el Institute of Electrical and Electronics Engineers (IEEE) muestran la estadística de la figura 2.8 sobre porcentaje de las zonas del motor de inducción en las que suceden los fallos.

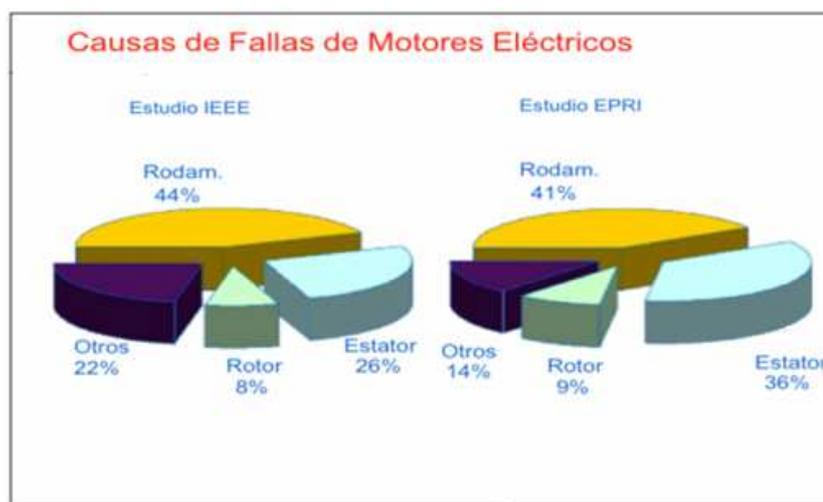


Figura 2.8. Causas de fallos

Para diagnosticar un fallo hay 2 opciones, dependiendo de si se altera o no el funcionamiento normal de la planta:

- **Diagnóstico en línea**
Consiste en efectuar el análisis del motor sin sacarlo de la línea de producción, trabajando en condiciones normales.
- **Diagnóstico fuera de línea**
Consiste en sacar el motor de su zona de trabajo habitual para realizar los estudios pertinentes, este diagnóstico está relacionado con el mantenimiento preventivo.

En este trabajo se estudian dos modelos para intentar detectar el fallo en los motores debido a rotura en las barras, pero además de este fallo existen muchos más, por lo que se pasa a explicar otros modos de fallo incipiente que se dan en los motores jaula de ardilla.

Se dividirán estos fallos según las partes del motor a las que afecten: estator y rotor, como se puede ver en [DUQ05].

2.5.1. Fallos en el estator

Se ha dividido este apartado en varios subapartados dependiendo de la naturaleza del fallo.

2.5.1.1. Solicitaciones eléctricas

El estator puede presentar los siguientes problemas de carácter eléctrico:

- **Descargas parciales.**
Cuando el aislante pierde capacidad, comienza a deslaminarse o se produce la evaporización de elementos volátiles contenidos en este se forman pequeños huecos de aire en el aislamiento.
La aplicación de tensión al estator da lugar a que en estos huecos se creen campos magnéticos, que pueden llegar a ocasionar descargas eléctricas en su interior. El fallo empieza cuando estas descargas son capaces de perforar el aislamiento de las zonas afectadas.
- **Transitorios**
Los transitorios de tensión pueden afectar a todos los elementos del sistema eléctrico, su consecuencia es la aparición de sobretensiones de efectos negativos. Las causas de estos transitorios pueden ser debidas a:

- Cortocircuitos
 - Reencendidos repetidos
 - Interrupción de la corriente
 - Apertura y cierre del interruptor del circuito
 - Desconexión de condensadores
 - Descargas eléctricas
- **Efecto corona**
- Este efecto está causado por la ionización del aire circundante alrededor de los conductores, se produce porque las partículas del aire se ionizan y son capaces de conducir la corriente eléctrica. El efecto corona se puede apreciar visualmente como una capa de color que rodea los conductores o de manera audible y se puede impedir con los denominados anillos anticorona. Las consecuencias del efecto corona pueden ser la rotura de los aislantes y la formación de un arco eléctrico.
- **Efecto “tracking”**
- Con los aislantes deteriorados por el efecto corona se pueden producir caminos paralelos superficiales sobre el aislante que permiten la circulación de corriente hacia masa. La consecuencia de este fallo es la formación de puntos calientes sobre los aislantes que contribuyen a su destrucción.

2.5.1.2. Solicitaciones térmicas

Los materiales aislantes presentan temperaturas límite de funcionamiento, por encima de la cual empiezan a perder sus propiedades. Esta temperatura se usa como base para la división de los aislantes en diferentes clases, como ilustra la tabla 2.1.

Clase	A	E	B	F	H	200	220	250
Temperatura Max (°C)	105	120	130	155	180	200	220	250

Tabla 2.1. Clases térmicas según la norma UNE 21-305-90.

- **Envejecimiento del aislamiento**
- El factor básico por el que el aislamiento envejece es la temperatura, por cada aumento de 10 °C se reduce su vida útil a la mitad. Para evitar el envejecimiento se puede aumentar la clase del aislante y vigilar la temperatura de operación.

– **Sobrecargas térmicas**

Disminuyen la vida útil de los aislamientos y son debidas a: variaciones en la tensión de alimentación, desequilibrio de tensión entre fases, múltiples arranques, sobrecargas, mala ventilación, incremento de la temperatura ambiente, etc.

2.5.1.3. Solicitaciones mecánicas

– **Movimiento relativo de las bobinas**

Las elevadas intensidades que se originan en los motores de inducción de media y alta tensión, sobre todo en los arranques o conmutaciones producen grandes fuerzas electromagnéticas. Debido a estas fuerzas es posible que las sujeciones de las bobinas se aflojen y estas se desplacen alterando el funcionamiento correcto del motor o la ruptura de los aislantes.

– **Roce entre estator y rotor**

Las causas más comunes por las cuales se produce este fallo son:

- Fallo en los rodamientos.
- Eje doblado
- Desalineamiento entre el rotor y el estator.
- Campo magnético desequilibrado.

– **Fatiga producida por vibraciones**

Las propias vibraciones de un motor en movimiento o las de máquinas adyacentes pueden ser la causa de la separación y rotura de las láminas de mica que forman el sistema aislante y del rozamiento de las bobinas con otros elementos del motor.

– **Introducción de objetos extraños.**

Este fallo afecta tanto al estator como al rotor, consiste en la introducción de elementos extraños, tanto del exterior del motor como del interior (tornillos sueltos, arandelas sueltas..etc), en el entrehierro del motor de inducción, normalmente a través del sistema de ventilación de este.

Este fallo produce daños en el aislamiento, en las bobinas, barras, anillos de cortocircuito o chapas del núcleo magnético.

2.5.1.4. Solicitaciones de origen ambiental.

– **Contaminación**

La contaminación ambiental tal como partículas de suciedad, agua, aceite, grasas o sus vapores pueden quedar impregnadas en los aislantes de los devanados del motor haciendo que la capacidad para disipar el calor se vea afectada reduciéndose la vida de los aislamientos, que se formen caminos alternativos de la corriente hacia la masa y que se degrade la grasa de los rodamientos.

– **Condensación**

Otro problema de índole ambiental es la condensación de la humedad en los aislamientos lo que al igual que antes provoca derivaciones a masa y pérdida de capacidad para disipar calor.

Este fallo puede solucionarse secando el interior del motor cuando este lleve mucho tiempo sin funcionar o en el caso de motores importantes se puede mantener la temperatura del interior del motor por encima de la temperatura ambiente mientras este no funcione mediante el uso de estufas o similares.

– **Funcionamiento en altitud elevada**

Por encima de determinada altitud el enrarecimiento del aire provoca una disminución de la capacidad de los motores para disipar el calor. Por lo tanto hay que disminuir la potencia nominal de la máquina para que no se produzcan averías.

De todas formas los fabricantes de motores suelen aportar graficas que relacionan el porcentaje de reducción de la potencia con la altura.

2.5.2. Fallos en el rotor

Según la naturaleza de los fallos.

2.5.2.1. Solicitaciones térmicas.

– **Sobrecargas térmicas.**

En determinadas situaciones el rotor se puede ver sometido a sobrecargas térmicas que darán lugar a fallos en el rotor y en el conjunto del motor, algunas de estas situaciones son:

- Elevado número de arranques consecutivos que harán aumentar la temperatura de las barras y de los anillos de cortocircuito.

- Bloqueos del rotor por carga elevada.
 - Barras rotas.
 - Ventilación insuficiente.
 - Roce entre el rotor y el estator.
- **Desequilibrio térmico.**
Este fallo puede causar rotura de barras o de anillos en el rotor, pueden ser debido a condiciones en su diseño o a condiciones de trabajo por encima de los límites especificados por el fabricante.
Las causas más comunes son:
- Desigual transferencia en las barras del rotor y el núcleo del rotor.
 - Puntos calientes en la superficie del rotor.
 - Diferencias de temperatura por una desigual ventilación en el rotor.
 - Frecuentes arranques que originan diferencias de temperatura en las barras del rotor.
- **Puntos calientes y pérdidas.**
Hay muchas variables que pueden ser causa de pérdidas o de formación de puntos calientes en el rotor como:
- Uniones deficientes entre las barras y los anillos.
 - Chispas en la superficie del rotor.
 - Falta de simetría.
 - Las barras del rotor están desigualmente sujetas.
 - Diseño inadecuado de la laminación.
 - Acortamiento irregular de las barras del rotor.
- **Chispeos en el rotor.**
En los arranques la corriente que circula por el rotor puede llegar a superar en cinco veces su valor nominal, originando caídas de tensión que pueden exceder en seis veces su valor en régimen de funcionamiento. Se crean fuerzas magnéticas que hacen vibrar las barras y en ocasiones se pueden producir desprendimientos de partículas de hierro de las barras en el entrehierro que son llevadas a la incandescencia por la corriente.
Este chispeo puede ocasionar la rotura de las barras.

2.5.2.2. Solicitaciones electromagnéticas

- **Efecto electromagnético**

Las fuerzas electromagnéticas son proporcionales al cuadrado de la corriente y son unidireccionales. Los esfuerzos que producen estas fuerzas sobre las barras del rotor provocan la deflexión de las mismas, sobre todo en los momentos en que son sometidas a mayor fatiga como en los arranques.

- **Campos electromagnético desequilibrados**

En un motor ideal el estator y el rotor están centrados, de tal manera que las fuerzas magnéticas se equilibran, en cambio en un motor real aparecen fuerzas resultantes debido a que el rotor muestra cierta excentricidad con respecto al estator causado por el propio peso, desgaste de los rodamientos o alineación del motor entre otros.

Si estas resultantes actúan durante el tiempo suficiente el eje puede llegar a doblarse y que el rotor choque con el estator. Este efecto es más importante cuanto mayor corriente circule por los devanados como en los arranques.

2.5.2.3. Solicitaciones dinámicas

- **Fuerzas centrífugas**

La mayoría de fallos debido a sollicitaciones dinámicas se deben a someter al rotor a velocidades mayores que aquellas para las que fue diseñado. Normalmente los motores tienen mecanismos para no sobrepasar la velocidad máxima y evitar así desplazamientos del eje y del paquete de chapas o de la jaula respecto al núcleo del rotor.

- **Esfuerzos cíclicos**

Si el rotor está sometido a esfuerzos cíclicos pueden producirse fallos por fatiga. Las causas que originan este tipo de esfuerzos como: desalineamiento con el eje accionado, ejes doblados, fluctuaciones de la carga, correas de transmisión sobretensadas, etc.

- **Solicitaciones ambientales.**

Como también ocurre con el estator, elementos externos que entren dentro del motor pueden producir abrasión, corrosión o impedir el correcto funcionamiento del sistema de ventilación del motor provocando calentamientos.

Capítulo 3

Machine Learning

3.1. DEFINICIÓN

Machine Learning es una rama de la inteligencia artificial, ligada a la estadística y a la minería de datos, que se centra en desarrollar métodos para obtener conocimiento a través de la experiencia, por ejemplo, detectando patrones en los datos que permitan predecir las preferencias de los clientes de una empresa.

Esto se consigue mediante algoritmos que pueden aprender, entrenados con un conjunto de datos de entrenamiento de forma que ellos sean capaces de generalizar con nuevos datos que se le introduzcan, como se expone en la figura 3.1.

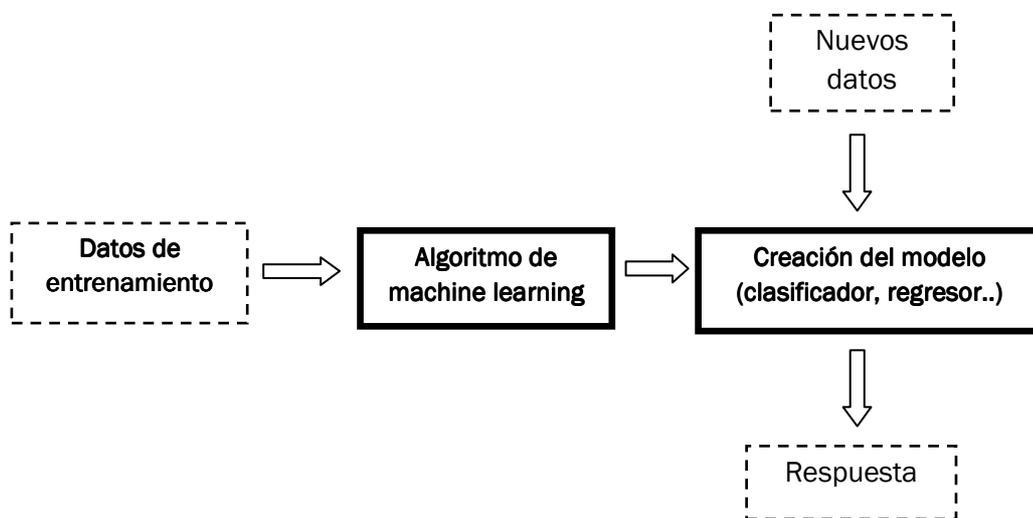


Figura 3.1. Pasos del aprendizaje.

Por citar un caso famoso: la compañía Walmart, que fue de las empresas pioneras en recopilar datos de sus clientes en los años 90, gracias a este tipo de técnicas, descubrió que las ventas de determinada barrita energética aumentaban hasta 10 veces los días previos a un tornado.

Se suelen considerar dos tipos principales de aprendizaje [SIE06]: Supervisado y no supervisado.

– Supervisado

En este caso el conjunto de entrenamiento está formado por un conjunto de objetos para los que se conocen una serie de características (también denominadas variables explicativas o predictores) englobadas en un vector X y una etiqueta que denominaremos Y que contiene el valor de la variable a predecir o

variable respuesta para ese objeto. De este modo el conjunto de entrenamiento puede representarse como $D = \{(X_i, Y_i)_{i=1, \dots, n}\}$.

– **No supervisado**

Solo tenemos el vector de características sin ninguna etiqueta $D = \{(X_i)_{i=1, \dots, n}\}$ y el objetivo suele ser encontrar patrones en los datos.

También es común dividir los métodos en:

– **Métodos lazy (perezoso)**

Estos métodos como los k-vecinos más próximos no crean el modelo matemático hasta que se les consulta, por eso se denominan lazy porque esperan el máximo antes de crear el modelo.

Estos métodos aprenden de forma rápida pero clasifican de manera lenta.

– **Métodos eager (ávidos)**

Con los datos de entrenamiento crean un modelo, es decir, aprenden, y a partir de este modelo son capaces de hacer generalizaciones.

Tienen la ventaja de que clasifican rápido pero la creación del modelo es lenta, por ejemplo, redes neuronales o Support Vector Machine.

3.2. TIPOS DE PROBLEMAS

Se explicará brevemente algunos de los problemas que se pueden resolver con Machine Learning, como se puede ver en [HER07], añadiendo ejemplos de aplicaciones reales.

3.2.1. Clasificación

Este es un problema supervisado en el que la variable Y es una variable categórica. Se dispone de un conjunto $D = \{(X_i, Y_i)_{i=1, \dots, n}\}$, donde $Y \in \{1, \dots, C\}$ siendo C el número de clases (si $C=2$ se llama clasificación binaria, si $C>2$ se denomina clasificación multiclase). El objetivo es, dado este conjunto de entrenamiento D ser capaz de clasificar un nuevo objeto X del que se desconoce su clase.

Algunas aplicaciones reales son:

- Los algoritmos que detectan y eliminan los correos basura (spam), lo hacen representando cada correo como un vector de características, tales como el número de veces que se repiten determinadas palabras como “barato” o “comprar” y aplicándoles distintos algoritmos entrenados con otros correos de los que se conoce la clase.
- Programas de reconocimiento de escritura o reconocimiento facial.

- Fondos de inversión que intentan predecir la tendencia del mercado de valores (alcista o bajista) usando datos correspondientes a determinadas características de los “tuits” escritos en un determinado país.

3.2.2. Regresión

Similar a la clasificación, pero aquí la clase o variable respuesta es una variable continua, por ejemplo:

- Intentar predecir el precio de las viviendas con arreglo a grandes bases de datos del mercado inmobiliario.

3.2.3. Clustering

Este es un procedimiento no supervisado. Consiste en intentar descubrir la existencia de grupos en los datos. Así objetos que presentan características similares quedarían agrupados en un mismo cluster o grupo, ejemplo:

- Agrupar a nuestros clientes en diferentes clusters, de tal forma que podamos personalizar la publicidad atendiendo a sus preferencias.

3.2.4. Asociación

Su objetivo consiste en identificar relaciones entre las características o variables explicativas, un ejemplo:

- Análisis de las cestas de la compra por parte de los supermercados, asignando en una matriz la cantidad de productos comprados (columnas) para cada compra (transacción), de tal manera que podamos encontrar correlaciones entre diferentes productos comprados, lo que será útil a la hora de cambiar la distribución del supermercado o lanzar ofertas.

3.3. VALIDACIÓN DEL MODELO

En problemas de clasificación como el que aquí se trata lo habitual es medir la bondad del modelo a través de la probabilidad de clasificación de errores o tasa de error verdadero del clasificador, como se puede ver en [PAJ10], [SIE06], que es la probabilidad de que el modelo no clasifique a un nuevo objeto Z cualquiera en la clase correcta Y . Si denotamos el clasificador como f podemos escribir esta probabilidad como $P(f(Z) \neq Y_Z)$. Esta probabilidad depende de las distribuciones de las variables en cada una de las clases. Como estas distribuciones son habitualmente desconocidas en la práctica no suele ser posible calcular esta probabilidad por lo que suelen calcularse estimadores de la misma.

El estimador más intuitivo es la tasa de error aparente, se define como la proporción de objetos de la muestra de entrenamiento que no son correctamente clasificados:

$$err(f, D) = \frac{1}{n} \sum_{i=1}^n (f(X_i)) \neq Y_i$$

Esta tasa es optimista pues los objetos utilizados en un modelo se suelen adaptar mejor a él que aquellos que se clasificarán en el futuro y con los que el clasificador no ha sido construido, conociéndose este hecho como sobreajuste u over-fitting.

Como este estimador no es bueno se construyen otras tasas de error estimadas que reducen el sesgo del estimador anterior, generalmente basados en dividir el conjunto de datos en dos partes, una con la que se entrena el algoritmo (conjunto de entrenamiento) y otra con la que se evalúa la bondad del mismo (conjunto de test). Algunos métodos para llevar a cabo esto son los siguientes:

- **Hold-out:** es el más sencillo, se divide la muestra disponible en una de entrenamiento y otra de test, para evaluar.
- **Cross-Validation:** Se divide el conjunto de datos en K partes, donde K-1 partes se convierten en el conjunto de entrenamiento, la parte restante se utilizará como conjunto de test. Se aplica el modelo K veces, siendo la tasa de error igual a la media de las tasas en cada iteración. Un caso particular es el “leave-one-out”, en el cual el parámetro K es igual al número de objetos.
- **Bootstrapping:** Se escoge una muestra aleatoria con reemplazamiento del mismo tamaño como conjunto de entrenamiento, y aquellas que no están incluidas en este conjunto de entrenamiento constituirán la muestra de test. Como los objetos se eligen aleatoriamente con reemplazamiento, la probabilidad de que tras N muestras un objeto no sea escogido es $(1-(1/N))^N \approx e^{-1} \approx 0,368$ y el número esperado de objetos distintos en la muestra se acerca a $0,632N$, así la estimación de la tasa de error después de realizar el proceso de selección b veces:

$$est_{boot} = \frac{1}{b} \sum_{i=1}^b (0,632\alpha_i + 0,368 err(f, D))$$

Donde α_i es la proporción de elementos mal clasificados en la i -ésima iteración del proceso.

3.4. SOBREAJUSTE DEL MODELO

El sobreajuste es un aspecto importante a tener en cuenta [SIE06], en la figura 3.2 se muestra un ejemplo en el que el clasificador utilizado (línea verde) se adapta excesivamente bien a los datos de entrenamiento, pero seguramente tenga problemas a la hora de clasificar nuevos datos, esto se corrige mediante el clasificador creado por la línea negra, más robusto.

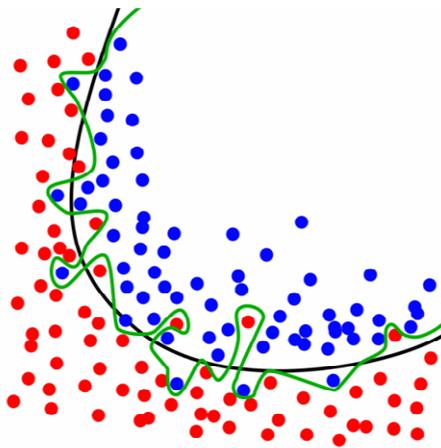


Figura 3.2. Sobreajuste. Línea verde.

Una manera de saber si este fenómeno se da en nuestro modelo es comparar el error aparente y el error estimado, si el error aparente no deja de disminuir mientras que el estimado aumenta, cuando la complejidad del modelo aumenta (por ejemplo, cuando se añaden más variables explicativas del mismo) es posible que exista sobreajuste, como ocurre en la figura 3.3.

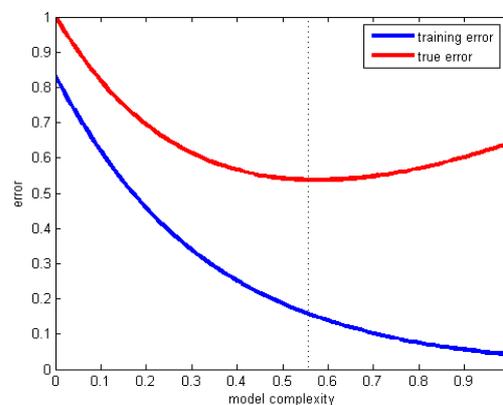


Figura 3.3. Sobreajuste al aumentar la complejidad del modelo.

Capítulo 4

Experimentación

4.1. DESCRIPCIÓN DE LA EXPERIMENTACIÓN

El experimento ha sido realizado por diversos profesores del Departamento de Ingeniería Eléctrica de la Escuela de Ingenierías Industriales de la Universidad de Valladolid, consiste en la reproducción de una serie de ensayos sobre un motor de inducción al que se le han ido haciendo diferentes modificaciones. Estas modificaciones consisten en hacer roturas de diferentes profundidades en una de las barras del rotor. Para ello, con un taladro se hizo un agujero en la barra, dependiendo de la profundidad de la rotura se definieron 4 estados:

- Los ensayos con etiqueta **R1** corresponden a un motor sano.
- Los ensayos con etiqueta **R2** corresponden a un motor con fallo incipiente. Se abre el motor, se saca el rotor y se realiza un agujero ligeramente profundo en una de sus barras.
- Los ensayos con etiqueta **R3** corresponden a un motor con una severidad de rotura ligeramente mayor que **R2**. Se aumentó la profundidad del primer agujero.
- Los ensayos con la etiqueta **R4** corresponden a un motor con la barra totalmente perforada, la máxima severidad de fallo.

En esta tabla se recogen las diferentes profundidades del agujero para cada estado:

	Estado	Profundidad
R1	Sano	0mm
R2	Fallo leve	5,7mm
R3	Fallo medio	12,1mm
R4	Barra rota	18mm

Tabla 4.1. Profundidad para cada estado.

En total para este trabajo se utilizó una base de datos correspondiente a 239 ensayos, con el motor conectado directamente a la red y a una frecuencia de 50 Hz.

En la tabla 4.2 se recoge el número de repeticiones para cada estado.

Estado	Nº ensayos
R1	110
R2	40
R3	42
R4	47

Tabla 4.2. Número de ensayos.

Además en cada ensayo pueden variarse una serie de parámetros como son: la fuerza de un par resistente, con el que se intenta simular una situación de funcionamiento real y la alimentación del motor, que puede provenir directamente de la red eléctrica o de variadores de frecuencia.

En el presente trabajo todos los ensayos han sido obtenidos con el motor directamente conectado a la red, con una frecuencia de 50 Hz.

4.2. DESCRIPCIÓN DEL BANCO DE ENSAYOS

En los siguientes apartados y subapartados se describen los diferentes aparatos de los que se compone el banco de ensayos, aunque de algunos de ellos no se haya hecho uso en este trabajo en concreto.

Los elementos utilizados para los ensayos son: motor de inducción, freno, sensores de velocidad, transductores, variadores de frecuencia.

Además se contó con un sistema de adquisición de datos para medir los valores de la corriente, que se conectó vía USB 2.0 a un ordenador portátil. Se usó el software MatLab para hacer los cálculos.



Figura 4.1. Banco de ensayo completo.

4.2.1. Motor

El motor eléctrico utilizado es el de la figura 4.2, siendo de la marca Siemens, con las siguientes características:

Características técnicas	
Potencia nominal	0,75 Kw
Tensión nominal	3x230/400V
Conexión del estator	Estrella
Frecuencia nominal 50 Hz	50 Hz
Corriente nominal	1,9 A
Número de pares de polos	2
Número de barras del rotor	28
Velocidad nominal	1395 rpm
Denominación	3 Mot 1la7083-4aa10
Número de serie	UD 0601/7009477-12
Eficiencia	IE1
Índice de protección	IP55
Tamaño	80M IM B3
Clase Térmica	ThcL F
Factor de servicio	1,1

Tabla 4.3. Características técnicas del motor utilizado en la experiencia.

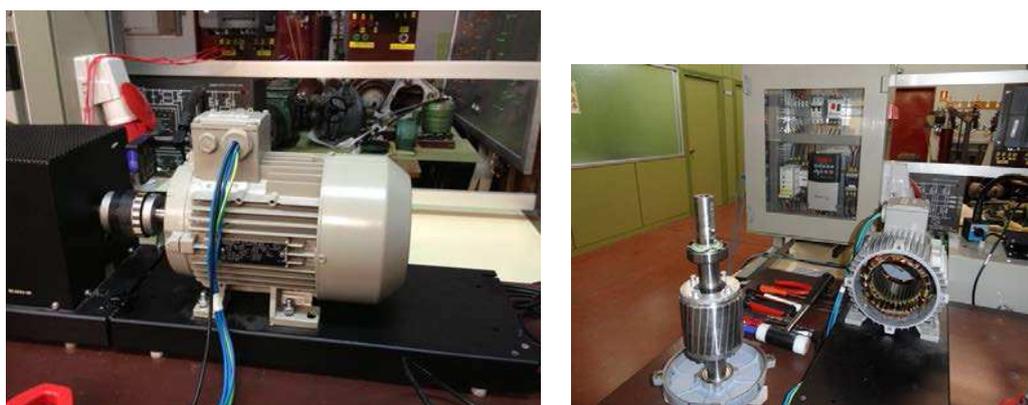


Figura 4.2. Diferentes perspectivas del motor utilizado.

4.2.3. Sensores de velocidad.

Los tacómetros o sensores de velocidad sirven para medir la velocidad de giro del motor. El de la figura 4.5 es de tipo digital, de la marca datalogic, concretamente el S60W y mide la velocidad mediante el contraste blanco-negro producido por una banda de esos mismos colores instalada en el eje que se puede apreciar perfectamente en la figura 4.6.

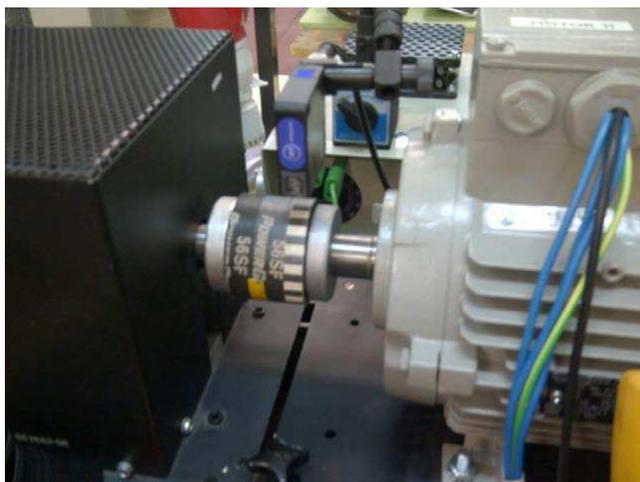


Figura 4.5. Sensor S60W

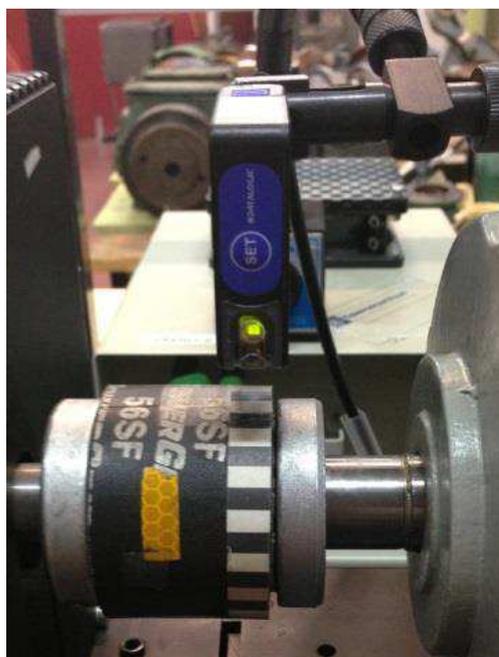


Figura 4.6. Frontal del sensor S60W

Este sensor tiene las siguientes especificaciones técnicas:

Características técnicas	
Tensión de alimentación	10-30Vdc
Rizado	2Vpp max
Consumo de corriente	30Ma
Outputs	PNP o NPN 30Vdc
Corriente de salida	100Ma
Voltaje de saturación de salida	$\leq 2V$
Tiempo de respuesta	100 μ s
Frecuencia de conmutación	5kHz
Indicadores	LEDs
Ajustes	Boton de ajuste
Memoria de datos	EEPROM
Distancia de funcionamiento	19 \pm 2mm

Tabla 4.5. Características técnicas.

4.2.4. Transductores

Los transductores son unos dispositivos que proporcionan una representación manipulable de la variable que llevan asociada. En el banco de ensayo se han acoplado a una placa que fue diseñada y construida en el laboratorio. En este experimento se ha contado con dos tipos de transductores:

- **Los transductores de intensidad:** convierten la corriente alterna en tensión alterna proporcional a la intensidad.
- **Los transductores de tensión:** convierten una tensión de 380V en otra más pequeña, $\pm 5V$ o $\pm 10V$.

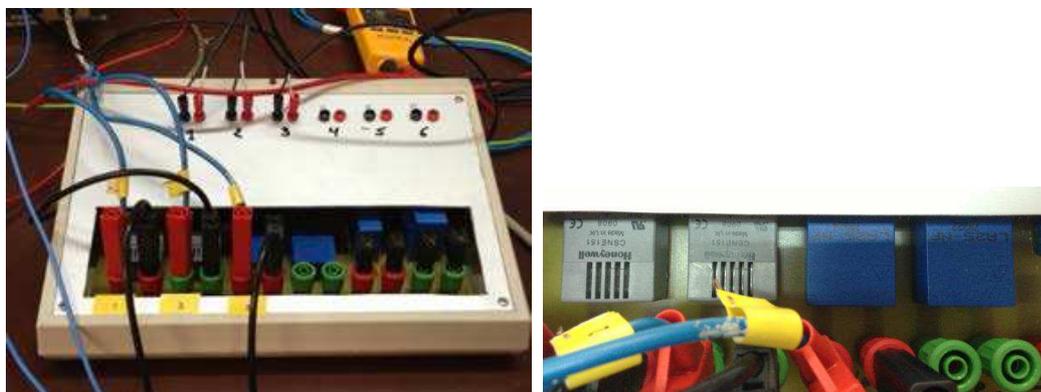


Figura 4.7. Transductor de intensidad a la izquierda. Detalle del transductor a la derecha.



Figura 4.8. Transductor de tensión a la izquierda. Detalle del transductor a la derecha.

- Transductores de corriente.
Se han utilizado dos transductores de corriente Honeywell y un tercero de LEM.
El de la figura 4.9. es el transductor de la marca Honeywell, el de la figura 4.10 es de tipo LEM.



Figura 4.9.. Transductor de corriente Honeywell

Características técnicas	
Denominación	CSNE151
Corriente primaria (valor máximo en continuo)	$\pm 36\text{A}$
Tensión primaria (AC y DC)	10-500V.
Corriente secundaria nominal	25mA (eficaces)
Tensión de alimentación	$\pm 15\text{Vdc}$
Consumo de corriente	10($\pm 15\text{V}$)+Corriente secundaria mA.
Ratio de conversión	2500:1000
Precisión	$\pm 0,5\%$
Tiempo de respuesta	1 μs
Empaquetamiento	PCB

Tabla 4.6. Características técnicas transductor Honeywell.

- Transductor LEM de corriente:

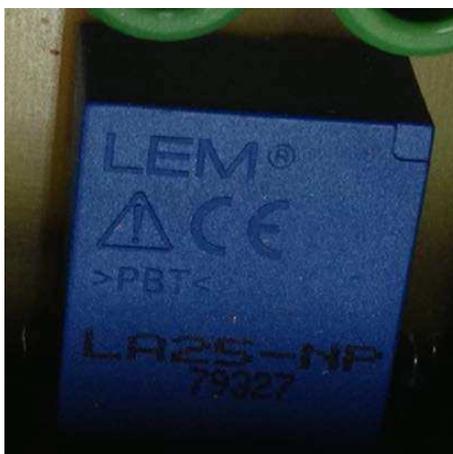


Figura 4.10. Transductor de corriente LEM

Características técnicas	
Denominación	LA25-NP
Corriente primaria nominal	25mA(eficaces)
Rango de medida	0.. ±36A
Corriente secundaria nominal	25mA(eficaces)
Tensión de alimentación	±15V.
Consumo de corriente	10+Corriente secundaria mA.
Precisión	±0,5%
Tiempo de respuesta	1 μs
Ando de Banda (-1 dB)	DC.. 150kHz
Empaquetamiento	PCB

Tabla 4.7. Características técnicas transductor de corriente LEM.

- Transductor de tensión LEM:
Se utilizó el transductor LEM de la figura 4.11



Figura 4.11. Transductor de tensión

Características técnicas	
Corriente primaria nominal	10 mA (eficaces)
Tensión primaria (AC y DC)	10-500 V.
Corriente secundaria nominal	25 mA (eficaces)
Tensión de alimentación	±12-15 V.
Consumo de corriente	10 (±15V)+ Corriente secundaria mA.
Ratio de conversión	2500:1000
Tiempo de respuesta a un escalón del 90% de la corriente primaria nominal	40 μ s
Empaquetamiento	PCB

Tabla 4.8. Características técnicas transductor de tensión LEM.

4.2.5. Variadores de frecuencia

Los variadores de frecuencia se utilizan para controlar la velocidad de un motor de inducción mediante la variación de la frecuencia de la corriente de alimentación.

Aunque en este trabajo no se han recopilado datos donde fueran utilizados variadores de frecuencia, pues se trabajó conectado a red a una frecuencia de 50 Hz.

En el laboratorio se encuentran variadores de frecuencia comerciales que se pueden conectar al banco de ensayos y alimentar con ellos el motor para realizar la experimentación.

4.2.6. Sistema de adquisición de datos.

Para transferir los datos relativos a la corriente desde los distintos transductores y sensores al ordenador se utilizó una plataforma de adquisición de datos de National Instruments.



Figura 4.12. Portátil con el que se trabajó a la izquierda. Detalle del sistema de adquisición de datos a la derecha.

4.3. CARACTERÍSTICAS MEDIDAS

En la base de datos que se usó para entrenar los clasificadores se recogieron para cada ensayo 27 características.

Aquellas relativas a la carga del motor no se tuvieron en cuenta a la hora de añadir variables a los clasificadores, como se explica más adelante.

Otras, en cambio fueron calculadas porque se piensa que pueden ser explicativas del estado del motor, así, se consideraron las medidas en el dominio de la frecuencia del espectro de corriente, que corresponden a las amplitudes de los armónicos indicativos de fallo alrededor de los armónicos 1, 5 y 7. Están expresados en dB y medidos respecto al 0 tras aplicar la FFT (Fast Fourier Transform) y el PSD (Power Spectral Density), ver los trabajos [JAR09] y [EST14].

Otras características se midieron en el dominio del tiempo aplicando varios estadísticos a la señal de la corriente estatórica.

4.3.1. Relativas a la carga del motor.

Estas variables no proporcionan información directa sobre el estado del motor, solo sirven para conocer el estado de la carga de este durante los ensayos y algunas de estas variables se usarán posteriormente para facilitar la visualización de otras que pensamos si son más importantes en el diagnóstico de fallos. Así, por ejemplo, el momento 1 respecto de la velocidad o del deslizamiento. Por esta razón estas variables no se han incluido en los análisis del capítulo 7 ni del capítulo 8.

- Corriente
Valor eficaz de la corriente consumida por el motor durante el ensayo.
- Velocidad del Motor
Denotada como velocidadM durante los ensayos, es la velocidad a la que gira el rotor, en revoluciones por minuto y esta medida es tomada por el sensor de velocidad.
- Deslizamiento
El deslizamiento s se corresponde con el valor relativo de la velocidad del rotor.

$$s = \frac{(n_1 - n)}{n_1}$$

Siendo n_1 la velocidad de sincronismo, es decir, la velocidad del campo creado por las corrientes trifásicas del motor y n la velocidad del motor.

4.3.2. Características en el dominio del tiempo

Se han aplicado los siguientes estadísticos a la onda temporal de la corriente estática.

Un estadístico es cualquier función aplicada a las observaciones de una muestra aleatoria. Por ejemplo, la media aritmética de una muestra.

- Momento

Dada una distribución bidimensional de variables (x, y) , dos números naturales (r, s) y un par de números reales (c, d) se define el momento respecto al par (c, d) de orden r y s como:

$$m(c, d) = \frac{\sum_{i=1}^n \sum_{j=1}^m (x_i - c)^r (y_i - d)^s n_{ij}}{n}$$

Donde:

n_{ij} es la frecuencia del par (x_i, y_i) .

n es el total de observaciones de la muestra.

Los momentos que más aportan son los que se calculan cuando el par c, d es el formado por las medias o cuando el par está formado por el origen $(0, 0)$.

Se ha calculado hasta el momento de orden cuatro.

El momento de primer orden se calculó respecto del origen, corresponde con la media de los valores de la corriente:

$$m_1 = \frac{1}{n} \sum_{i=1}^n x_i$$

El de orden dos con respecto de la media se corresponde con la varianza de los valores medidos:

$$m_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Los momentos restantes, calculados respecto de la media, hasta el de orden cuatro, se calculan atendiendo a la siguiente fórmula, donde M es el orden.

$$m_M = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^M$$

- Skewness

Es un estadístico que nos da una medida de la simetría que presenta la muestra de una variable aleatoria respecto de su media.

Hay varias formas de calcular este estadístico, se ha utilizado el coeficiente de Skewness de Pearson que se corresponde con el tercer momento normalizado.

$$Skew = \frac{m_3}{(\sqrt{m_2})^3}$$

Si Skewness es próxima a cero la muestra es razonablemente simétrica, si es mayor que cero es asimétrica positiva o a la derecha y si es menor que cero es asimétrica negativa o hacia la izquierda.

- Kurtosis

Este estadístico representa una medida de la forma que sigue la distribución de la muestra.

$$Kurt = \frac{m_4}{(\sqrt{m_2})^3}$$

Aunque se suele utilizar la siguiente notación, de forma que la curtosis de una distribución normal sea igual a tres.

$$G = \frac{m_4}{(\sqrt{m_2})^3} - 3$$

Dependiendo del valor del estadístico la distribución puede ser:

- Leptocúrtica: $Curt > 3$, colas más anchas que la normal y presenta mayor apuntalamiento.
- Platicúrtica: $Curt < 3$, colas menos anchas y menor apuntalamiento.
- Mesocúrtica: $Curt = 3$, corresponde a la distribución normal.

- Cumulante

Si tomamos la serie de Taylor de la expresión del momento de orden $r=0,1,\dots$ con respecto del origen y elegimos un número r de coeficientes obtenemos el cumulante de orden r .

En la tabla 4.9 se representan los cumulantes utilizados.

Orden del cumulante	Expresión
1 ^{er}	m_1
2 ^{do}	$m_2 - m_1^2$
3 ^{er}	$m_3 - 3m_1m_2 + 2m_1^3$
4 ^{to}	$m_4 + m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4$

Tabla 4.9. Cumulantes.

- Valor cuadrático medio (x_r)

Está definido por:

$$x_r = \left(\frac{1}{n} \sum_{i=1}^n \sqrt{|x_i|} \right)^2$$

- Pico Absoluto (x_p)

Es el máximo valor absoluto que toma la corriente estadística.

$$x_p = \max_{i=1,\dots,n} |x|$$

- Promedio de valores absolutos

Se calcula la media de los valores absolutos.

$$\overline{|x|} = \frac{1}{n} \sum_{i=1}^n |x|$$

- Valor Eficaz o RMS

Se corresponde con

$$X_{\text{rms}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

El valor eficaz de una corriente variable se corresponde con el valor constante de esa corriente que es necesaria para producir los mismos efectos caloríficos a través de una resistencia.

Esta característica se corresponde con el valor de la corriente incluida en el apartado de características relativas a la carga del motor, se utilizará para calcular los dos siguientes estadísticos.

- **Factor Cresta**
Se define como el cociente entre el pico absoluto y el valor eficaz de la corriente.

$$FC = \frac{x_p}{x_{rms}}$$

Para una señal sinusoidal el factor de cresta es igual a $\sqrt{2}$.

Este factor se utiliza para detectar picos absolutos excepcionales con respecto al valor eficaz.

- **Factor Forma**
Se define como el cociente entre el valor eficaz de la corriente y la media de los valores absolutos.

$$FF = \frac{x_{rms}}{|x|}$$

4.3.3. Características en el dominio de la frecuencia

Para calcular las características en el dominio de la frecuencia lo primero que hay que hacer es transformar la onda de la corriente en el dominio temporal a otra en el dominio de la frecuencia.

En determinadas condiciones resulta útil la transformada de Fourier (FT) para esta tarea, aunque en este caso se ha utilizado la Transformada rápida de Fourier (FFT) ya que requiere de un menor número de operaciones [JAR09].

Una vez obtenida la onda en el dominio de la frecuencia se calculó sobre esta el PSD (Power Spectral Density) normalizado [ROD13] y se midió el valor de la amplitud de las bandas laterales alrededor de determinados armónicos: el primero, el quinto y el séptimo.

La figura 4.13 representa el PSD normalizado del ensayo nº 206 del motor conectado a la red, que corresponde al ensayo nº 13 del motor con la barra totalmente perforada (estado R4). Las medidas de la amplitud que aparecen en la misma figura son las de la banda inferior y la banda superior del primer armónico.

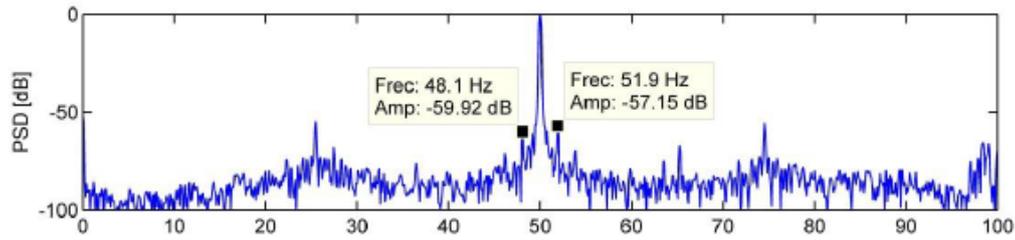


Figura 4.13. Espectro del PSD normalizado

La tabla 4.10 recoge la frecuencia en la que se encuentra cada banda lateral, como se puede ver en [DUQ05]. En total se midieron ocho bandas, cuatro alrededor del primer armónico, dos alrededor del quinto y otras dos alrededor del séptimo

Banda lateral	Frecuencia
1 ^{er} armónico, banda superior 1	$f_{bs}=(1+2s)f_1$
1 ^{er} armónico, banda superior 2	$f_{bs2}=(1+4s)f_1$
1 ^{er} armónico, banda inferior 1	$f_{bi1}=(1-2s)f_1$
1 ^{er} armónico, banda inferior 2	$f_{bi2}=(1-4s)f_1$
5 ^{to} armónico, banda superior	$f_{b5s}=(1+2s)5f_1$
5 ^{to} armónico, banda inferior	$f_{b5s}=(1-2s)5f_1$
7 ^{mo} armónico, banda superior	$f_{b7s}=(1+2s)7f_1$
7 ^{mo} armónico, banda inferior	$f_{b7s}=(1-2s)7f_1$

Tabla 4.10. Bandas laterales.

Donde s corresponde al desplazamiento y f_1 es la frecuencia fundamental.

Capítulo 5

Técnicas utilizadas

5.1. INTRODUCCIÓN

En este apartado se describen las diversas técnicas estadísticas usadas para la realización del trabajo.

En el capítulo de análisis descriptivo se utilizó el análisis de componentes principales (ACP) para conocer a nivel descriptivo la relación que existe entre las variables. Aunque en un primer momento se pensó que también se podrían aplicar los algoritmos KNN y SVM a los resultados del ACP se desechó la idea pues este método no reduce las variables a medir sino que trabaja con una combinación de todas ellas como se verá posteriormente y su interpretación sería complicada.

Por último, los métodos de clasificación elegidos que son el vecino más próximo (K-NN), las máquinas de vectores de soporte (SVM) y el discriminante lineal (LDA).

5.2. ANÁLISIS DE COMPONENTES PRINCIPALES (ACP)

El análisis de componentes principales es una técnica estadística multivariante que se utiliza para reducir la dimensionalidad de un grupo de variables explicativas mediante combinaciones lineales adecuadas de estas variables [PEÑ02]. El máximo de componentes principales que se pueden obtener es igual al número de variables originales, pero seleccionarlas todas para trabajar no aportaría nada por lo que se elige un número menor de ellas que expliquen lo suficientemente bien la variabilidad de las variables.

5.2.1. Obtención de la primera componente

Para aplicar este método es habitual tipificar las variables originales. Suponiendo que se tienen los valores $\{(X_{i1}, \dots, X_{ip})_{i=1, \dots, n}\}$ que toman p variables X_1, X_2, \dots, X_p para una muestra de tamaño n , se busca una combinación lineal Z_1 (primera componente principal) con coordenadas $Z_{1i} = u_{11}X_{1i} + \dots + u_{1p}X_{ip}$ y tal que Z_1 tenga la máxima variabilidad posible.

Simplificando, usando notación matricial, podemos escribir:

$$Z_1 = Xu_1$$

Donde u_1 corresponde al vector de pesos de la combinación lineal y X la matriz de datos cuyas columnas son los valores que toman las variables X_1, \dots, X_p .

La varianza de Z_1 será:

$$V(Z_1) = \frac{\sum_{i=1}^n Z_{1i}^2}{n} = \frac{1}{n} Z_1' Z_1 = \frac{1}{n} u_1' X' X u_1 = u_1' \left[\frac{1}{n} X' X \right] u_1 = u_1' V u_1$$

La primera componente principal se va a obtener de forma que su varianza sea máxima, es decir que los datos estén lo mas separados posible, que la suma de los pesos u_{ij} al cuadrado sea igual a la unidad.

Para resolver este problema de optimización se aplica el método de los multiplicadores de Lagrange con la siguiente función de Lagrange:

$$L = u_1'Vu_1 - \lambda(u_1'u_1 - 1)$$

Derivando respecto a u_1 .

$$\frac{\partial L}{\partial u} = 2Vu_1 - 2\lambda u_1 \Rightarrow (V - \lambda I)u_1 = 0$$

Que tiene solución si $|V - \lambda I| = 0$, que es equivalente a decir que λ es un valor propio de la matriz V , así tendríamos n raíces $\lambda_1, \lambda_2, \dots, \lambda_n$, que se podrán ordenar de mayor a menor y tomando el λ mayor se maximiza la varianza de Z_1 .

Por lo tanto, el vector propio asociado al valor propio λ normalizado ($u_1'u_1 = 1$) será el vector de óptimo de pesos $u_1 = (u_{11}, \dots, u_{1p})$ y, así, obtenemos la primera componente principal.

$$Z_1 = Xu_1$$

Siendo λ la varianza que recoge la primera componente principal.

5.2.2. Obtención de la segunda y resto de componentes

Para calcular la segunda componente principal $Z_2 = Xu_2$ se haría igual pero pidiendo que las componentes Z_1 y Z_2 estén incorreladas. Es decir,

$$0 = E[Z_2'Z_1] = E(u_2'X'Xu_1) = u_2'E(X'X)u_1 = u_2'Vu_1$$

Queremos hallar, entonces Z_2 maximizando la varianza $V(Z_2) = u_2'u_1$ sujeta a las restricciones $u_2'u_2 = 1$ y $u_2'Vu_1 = 0$.

Usando, como antes, el método de Lagrange:

$$L = u_2'Vu_2 - 2\mu(u_2'Vu_1) - \lambda(u_2'u_2 - 1)$$

Derivando, se tiene:

$$\frac{\partial L}{\partial u} = 2Vu_2 - 2\mu Vu_1 - 2\lambda u_2 = 0$$

A partir de esta ecuación obtendríamos como antes:

$$(V - \lambda I)u_2 = 0$$

Si multiplicamos por u_2' a la derecha se obtiene que $u_2'Vu_2 = \lambda \Rightarrow V(Z_2) = \lambda$, por lo que para maximizar la varianza de Z_2 hay que elegir el segundo máximo valor propio de la matriz V y tomar como vector de pesos el vector propio normalizado al segundo vector propio mayor, obteniendo así la segunda componente principal.

$$Z_2 = Xu_2$$

De esta forma se van calculando todas las componentes principales, es decir, para la componente j -ésima el vector de pesos normalizado u será igual al vector propio asociado al j -ésimo valor propio de la matriz V .

Siendo, además, la proporción de variabilidad explicada por la componente principal j -ésima la siguiente

$$\frac{\lambda_j}{\sum_{j=1}^p \lambda_j} = \frac{\lambda_j}{\text{Traza}(V)}$$

Y la proporción de la variabilidad explicada por las k primeras componentes es:

$$\frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^p \lambda_j} = \frac{\lambda_j}{\text{Traza}(V)}$$

Siendo esto último muy útil para elegir el número de componentes que nos interesa.

5.2.3. Criterio de la media aritmética.

Existen diferentes métodos para elegir un número óptimo de componentes principales, el más común es el que se muestra a continuación. No obstante existen otros como el contraste de raíces características o la prueba de Anderson.

En este se eligen las primeras componentes cuyo λ supere a la media.

$$\lambda_h > \bar{\lambda} = \frac{\sum_{j=1}^p \lambda_h}{p}$$

Si las variables están tipificadas se elegirían las componentes cuyo $\lambda_h > 1$.

5.3. ANÁLISIS DISCRIMINANTE

Métodos paso a paso

Lo que se intenta con los métodos paso a paso es determinar cuáles son las variables que más aportan a la discriminación entre los grupos [PEÑ02], eliminando las que menos aporten o las que aporten la misma información que otras. Existen 2 modalidades: paso a paso hacia delante (Forward) o hacia atrás (Backward).

- Forward: La primera variable que entra es la que maximice la separación entre los grupos, posteriormente se van formando parejas formadas por esta variable y las demás, una a una, de tal forma que la siguiente variable sea la que más ayude a la separación, y así, sucesivamente.
- Backward: En este método se comienza con todas las variables y se van quitando consecutivamente, una a una, las que menos aporte a la discriminación.

Para comprobar que variable discrimina mejor en cada paso es preciso utilizar algún criterio, de los que destacan los siguientes:

- Criterio basado en la discriminación de la lambda de Wilks:
En cada paso se selecciona la variable que haga más pequeño el valor del estadístico lambda de Wilks, definido por:

$$\lambda = \frac{|S|}{|T|}$$

Donde S es la suma de cuadrados intragrupo y T es la suma de cuadrados total.

- Criterio basado en la distancia de Mahalanobis:
La distancia de Mahalanobis mide la separación entre los grupos, en cada paso se añade al modelo la variable que maximice esta distancia. La distancia de Mahalanobis entre dos grupos a y b se define como:

$$D_{ab}^2 = (n - g) \sum_{i=1}^p \sum_{j=1}^p w_{ij} (\bar{X}_{ia} - \bar{X}_{ib})(\bar{X}_{ja} - \bar{X}_{jb})$$

Donde n es el número de casos, p es el número de variables, g es el número de grupos, \bar{X}_{ia} es la media del grupo a en la i -ésima variable independiente, \bar{X}_{jb} es la media del grupo b en la j -ésima variable independiente, w_{ij} es el elemento correspondiente de la matriz inversa de covarianzas intragrupos.

- Criterio basado en la F intergrupos:
La F es un estadístico que se calcula a partir de la distancia de mahalanobis (D_{ab}^2). En cada paso se va a elegir la variable que produce el mayor valor de F .

$$F = \frac{(n - 1 - p)n_a n_b}{p(n - 2)(n_a + n_b)} D_{ab}^2$$

Donde n_a y n_b son el número de casos en cada grupo.

5.4. MÉTODO DE LOS VECINOS MÁS PRÓXIMOS (KNN)

Los vecinos más próximos son una de las técnicas de clasificación más simples e intuitivas, como se puede ver en [SIE06], pertenecen al aprendizaje supervisado y son de tipo lazy, como se explicó en el apartado 3.1 de este trabajo. La idea que hay detrás es que los objetos de una misma clase, posiblemente, se encuentren cerca en el espacio de representación.

Por tanto, asignaremos a un nuevo objeto sin clasificar a la clase mayoritaria de los k -vecinos más próximos a él como se ve en la figura 5.1. Esto se puede hacer de varias maneras como se comentará después (voto mayoritario, métrica de la distancia etc).

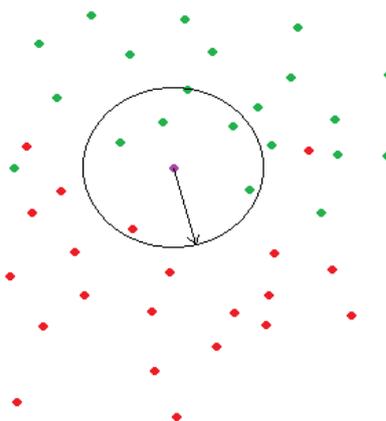


Figura 5.1. Ejemplo de 5-vecinos más próximos.

5.4.1. Variantes de KNN

Se comentarán algunas modificaciones que se pueden hacer al KNN original con las que se puede trabajar para intentar aumentar su rendimiento.

5.4.1.1. KNN por distancia media

Se da mayor importancia al parámetro distancia que al voto, de tal manera, que se calcula la distancia media de los k vecinos más próximos por clase y se le asigna al nuevo objeto la clase de la distancia media menor.

5.4.1.2. Distancia mínima

A cada clase del conjunto de entrenamiento se le asigna un vecino representante, en el caso del NCN (Nearest Centroid Neighbour) este representante es el centroide de cada clase, de forma que los nuevos objetos serán clasificados a aquella clase a la que pertenezca el vecino representante más cercano.

5.4.1.3. Pesado de casos

En esta variante el voto de cada K vecino más cercano va a ser ponderado con un peso W_i , $i=1,\dots,K$. Este peso puede tener distintas formas, por ejemplo se puede tomar:

- El voto es inversamente proporcional a la distancia.

$$W_i = \frac{1}{Dist(x, x_i)}$$

- Voto fijo: los pesos vienen por defecto y de tal forma que los más cercanos aporten más.

5.4.1.4. Métrica

Por último, también es posible jugar con la métrica, ya que dependiendo de una serie de factores, como la variabilidad de los datos, nos puede resultar interesante utilizar una u otra.

- Euclidea: es la que se ha usado en este trabajo y se define por:

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i^2 - y_i^2)}$$

- Mahalanobis: Fue propuesta por el científico indio Mahalanobis y difiere de la euclídea en que esta sí tiene en cuenta la correlación entre las variables.

$$d_m(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T \Sigma^{-1} (\vec{x} - \vec{y})}$$

Donde Σ es la matriz de covarianza.

5.5. SUPPORT VECTOR MACHINE (SVM)

El Support Vector Machine (SVM) es un algoritmo de aprendizaje supervisado. Para el problema de la clasificación biclase, la idea principal consiste básicamente en encontrar un hiperplano que logre separar las dos clases. La particularidad del SVM se basa en que busca el hiperplano de “margen” máximo.

El margen del hiperplano es la distancia del hiperplano a los vectores de soporte, siendo estos las observaciones de diferentes clases más cercanas entre sí, de tal manera que para hallar dicho hiperplano solo van a ser de utilidad los vectores de soporte, las demás observaciones se vuelven irrelevantes. La figura 5.2 permite comprender mejor la idea.

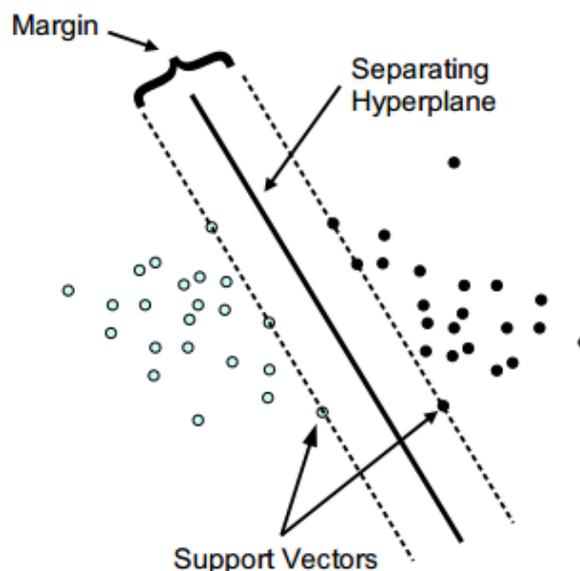


Figura 5.2. Hiperplano separando 2 conjuntos de datos [KAR06].

NOTA:

Los hiperplanos en R^n son de la forma:

$$H = \{ x \in R^p / w^t x = \alpha \} \text{ donde } w = (w_1, \dots, w_p)$$

Es decir, se tiene:

$$w^t x = w_1 x_1 + w_2 x_2 + \dots + w_p x_p = \alpha$$

En R^2 , por ejemplo, los hiperplanos serán rectas:

$$w_1 x_1 + w_2 x_2 = \alpha$$

y en R^3 serán planos:

$$w_1 x_1 + w_2 x_2 + w_3 x_3 = \alpha$$

5.5.1. Caso separable linealmente

Se tiene un conjunto de datos de entrenamiento con dos clases que supondremos linealmente separables (es decir, separables por un hiperplano):

$$(x_1, y_1), \dots, (x_n, y_n), x \in R^p, y_i \in \{+1, -1\}, i = 1, \dots, n$$

La frontera de separación será el hiperplano:

$$D(x) = w^T x + b$$

Donde w es un vector de dimensión p perpendicular al hiperplano y b es la separación con respecto al origen, como son linealmente separables y no hay ningún punto donde $w^T x_i + b = 0$, se cumple que cada clase está separada de la otra y dependiendo en qué lado se encuentre cada muestra con respecto al hiperplano de separación su etiqueta será:

$$\begin{aligned} w^T x_i + b &\geq 0 \text{ si } y_i = 1 \\ w^T x_i + b &\leq 0 \text{ si } y_i = -1 \end{aligned}$$

La mínima distancia entre el hiperplano de separación y los datos más cercanos se llama margen y correspondería a $M/2$ en la figura 5.3.

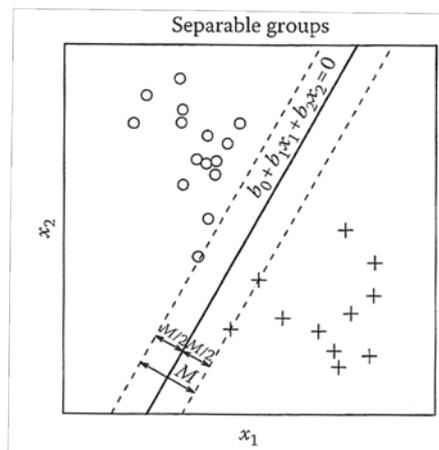


Figura 5.3. Hiperplano de margen máximo [FIL09].

El objetivo será por lo tanto encontrar los parámetros w y b con los cuales el margen se haga máximo. No se expondrá aquí la demostración matemática pues excede del alcance de este trabajo pero se puede encontrar en [ACH07] y en [PAJ10].

5.5.2. Caso no separable linealmente

Se puede dar el caso de que las muestras de entrenamiento no puedan ser separadas linealmente sin errores por un hiperplano, es decir, por lo menos una de las muestras aparece en el lado equivocado del hiperplano de separación como indica la figura 5.4.

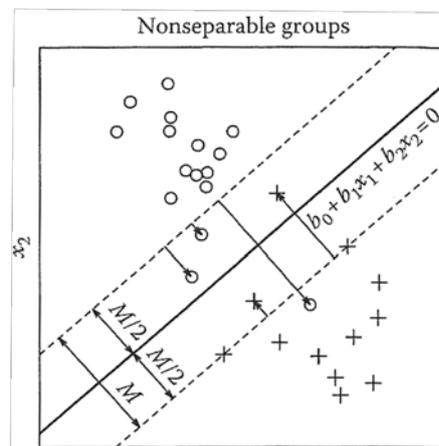


Figura 5.4. Hiperplano de margen máximo [FIL09].

Aunque no se puedan separar linealmente sí podemos encontrar un hiperplano que maximice el margen y minimice las clasificaciones incorrectas mediante la introducción de un parámetro C , denominado coste, que es un valor ajustable por el usuario que como se verá en el capítulo 8 se puede variar para obtener un mejor comportamiento del algoritmo a la hora de clasificar.

Este apartado tampoco contendrá la demostración matemática para la obtención de los parámetros w y b cuando los datos no son separables linealmente adecuados por exceder el alcance del trabajo, se vuelve a hacer referencia a [ACH07] y [PAJ10] donde se encontrará la demostración matemática oportuna.

5.5.3. Kernel

Para evitar el problema de que los datos de entrenamiento no sean separables linealmente se pueden transformar los datos del espacio original a un espacio de mayor dimensión en el que sí sean separables, que se denomina espacio de características.

Para hacer esto se utiliza una función no lineal llamada función Kernel, $g(x)=(g(x_1),\dots,g(x_n))$, que transforma los datos de un espacio p -dimensional a otro de dimensión l , con $l > p$. La función de decisión quedaría:

$$D(t) = w^T g(x) + b$$

Algunos kernels que se utilizan en SVM son:

Kernel polinomial

El kernel polinomial de grado d , siendo d un número natural, esta dado por:

$$H(x, x') = (x^T x' + 1)^d$$

Por ejemplo, para $d=2$ y $p=2$:

$$H(x, x') = 1 + 2x_1x'_1 + 2x_2x'_2 + 2x_1x'_1x_2x'_2 + x_1^2x_1'^2 + x_2^2x_2'^2 = g^T(x)g(x)$$

Donde:

$$g(x) = (1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2)$$

En este caso, la función kernel polinomial transformaría los datos en un espacio original de dimensión dos a un espacio de características de dimensión seis.

Kernel de base radial

Este kernel está dado por:

$$H(x, x') = \exp(-\gamma \|x - x'\|^2)$$

Donde el parámetro gamma puede modificarse para controlar el radio de la base radial.

Kernel de red neuronal

Está dado por la expresión:

$$H(x, x') = \tanh(c_1 x_i^T x_j + c_2)$$

En la figura 5.5 se muestran las funciones de decisión después de usar diferentes tipos de funciones Kernel.

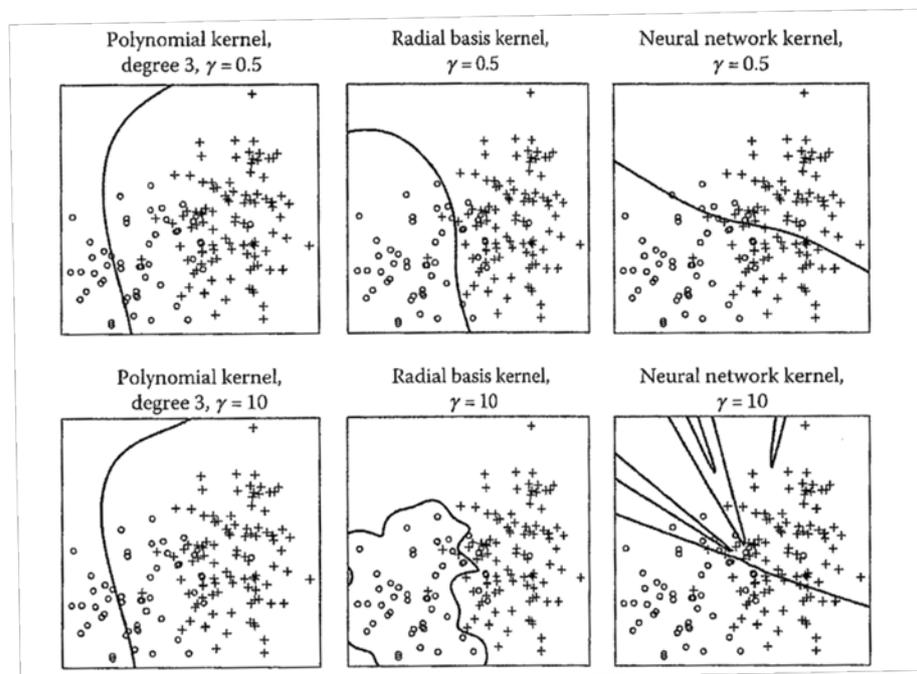


Figura 5.5. Frontera separación para distintos kernels [FIL09].

5.5.4.SVM multi-clase.

Para resolver el problema del SVM con C clases lo que se hace es dividirlo en pequeños problemas binarios. Existen múltiples opciones:

- **one-vs-all:** se clasifica cada clase con respecto al resto de clases que se agruparan en una única clase que llamaremos "all". Así, tendremos que hay que construir C clasificadores que definen una serie de hiperplanos que separan cada clase de las restantes o clase all. Para que una clasificación sea considerada de forma satisfactoria solo debe ser asociada a una clase diferente a la clase all en un clasificador y en los otros clasificadores ser asociada a la clase all.

- **one-vs-one:** con este método se construyen exactamente $C(C-1)/2$ clasificadores, enfrentando cada clase dos a dos con todas las demás, obteniendo un hiperplano separación para cada problema de dos clases. Se somete al nuevo objeto a todos los $C(C-1)/2$ clasificadores y se le añade un voto por cada clase propuesta.

Capítulo 6

Análisis descriptivo de los datos

6.1. ANÁLISIS DE LOS DATOS

Para conseguir un mejor rendimiento a la hora de aplicar los algoritmos de clasificación KNN y SVM a los datos se hace de vital importancia un buen pre-procesamiento de los mismos, con lo que conseguiremos reducir el ruido y retener la mayor cantidad de información posible, para ello se intentará reducir el número de características, ya que es posible que algunas no aporten ninguna información a la clasificación y puede que otras estén relacionadas entre sí y aporten la misma.

6.1.1. Correlación

Ante la sospecha de que puede existir colinealidad (algunas variables son combinación lineal exacta de otras) se ha hecho un estudio de las correlaciones entre todas las variables.

Mediante un análisis multivariado con StatGraphics se obtienen los siguientes pares de variables tienen un coeficiente de correlación igual a 1.

(Momento2, Cumulante2)

(Cumulante1, momento1)

Por lo que se ha eliminado del análisis **momento2, momento1**.

Esto era obvio ya que si se revisa con detalle la definición de las variables se puede ver que momento1 es igual a cumulante1 y momento2 es equivalente a cumulante2.

Ahora vamos a considerar la colinealidad aproximada, cuando el coeficiente de correlación es ≈ 1 .

	Coef.Corre
Cumulante4, momento4	-0,9993
Pico absoluto, momento4	0,97
Xr, momento4	0,98
Picoabosuluto,cumulante4	-0,975
Xr,cumulante4	0,9872
Xr, pico absoluto	0,9927
Cumulante1, cumulante3	-0,9763
Xr, Cumulante 2	0,9963
Xr, Abs.Mean	0,9994

Tabla 6.1. Coeficiente de correlación entre diferentes variables.

La tabla 6.1 contiene aquellos pares de variables en los que los coeficientes de correlación son en valor absoluto mayor que 0,96. Como consecuencia se

tomó la decisión de eliminar las variables: **cumulante4**, **momento4**, **picoabsoluto**, **cumulante1**, **cumulante2**, **abs.mean**.

La tabla 6.2 muestra datos estadísticos básicos obtenidos con Statgraphics de las variables que no han sido eliminadas.

	VelocidadC	BandInf	BandSup	BandInf2	BandSup2	Band5Inf
Recuento	239	239	239	239	239	239
Promedio	1462,25	-63,3339	-62,3	-72,7908	-72,2927	-74,2432
Desviación Estándar	18,642	7,10998	8,27944	5,32752	5,80873	6,47498
Coeficiente de Variación	1,27%	-11,23%	-13,29%	-7,32%	-8,04%	-8,72%
Mínimo	1425,86	-81,4724	-77,0907	-101,793	-86,8387	-98,7451
Máximo	1495,5	-48,4597	-30,8056	-57,2895	-56,2197	-59,6666
Rango	69,6429	33,0127	46,2851	44,5031	30,619	39,0785
Sesgo Estandarizado	1,05386	1,62027	2,72604	-4,06929	-0,335305	-2,33964
Curtosis Estandarizada	-3,14317	-2,94549	-1,10628	12,402	-2,29016	0,523592

	Band5Sup	Band7Inf	Band7Sup	Momento3	Cumulante2	Cumulante3
Recuento	239	239	239	239	239	239
Promedio	-78,168	-79,9086	-82,9526	0,00677908	4,61907	-0,0874648
Desviación Estándar	6,63959	5,39803	7,31433	0,00834627	0,883592	0,166621
Coeficiente de Variación	-8,49%	-6,76%	-8,82%	123,12%	19,13%	-190,50%
Mínimo	-91,1233	-97,0971	-106,833	-0,02196	3,25953	-0,503179
Máximo	-38,8166	-54,9874	-34,7662	0,030366	7,37256	0,228916
Rango	52,3067	42,1098	72,067	0,052326	4,11304	0,732095
Sesgo Estandarizado	9,51434	4,08566	11,3582	-0,974381	4,89971	-2,86822
Curtosis Estandarizada	18,7006	6,06416	28,2306	2,92541	-0,615256	-1,98055

	Cumulante6	Skewness	Kurtosis	xr	FactorCresta	FactorForma
Recuento	239	239	239	239	239	239
Promedio	2,53307	0,00072814	1,5059	1,7613	1,4632	1,11091
Desviación Estándar	0,0565837	0,00083447	0,0167848	0,176014	0,0295979	0,00505366
Coeficiente de Variación	2,23%	114,60%	1,11%	9,99%	2,02%	0,45%
Mínimo	2,4259	-0,002151	1,48253	1,46901	1,38981	1,10084
Máximo	2,67783	0,002973	1,54301	2,26119	1,52914	1,12746
Rango	0,251937	0,005124	0,060483	0,792181	0,139326	0,026623
Sesgo Estandarizado	3,00683	-1,02677	3,31971	3,74294	-2,88961	3,1585
Curtosis Estandarizada	-1,05295	1,13236	-2,98863	-1,78685	-0,674556	-0,247834

Tabla 6.2. Diferentes estadísticos de cada variable.

6.1.2. Normalidad

Para que los métodos de selección de características sean más eficaces suele ser conveniente que las variables tengan distribución aproximadamente normal. Después de hacer test de normalidad a las 18 variables con Statgraphics, las únicas con distribución aproximadamente normal son: band7inf, momento 3 y Skewness. A continuación se han aplicado transformaciones (logarítmica, raíz cuadrada e inversa) a todas las variables para las que se rechazó la normalidad. Se consiguió normalidad aproximada

para la transformación logarítmica de la variable BandInf2 por lo que a partir de ahora se considerará siempre esta variable transformada logarítmicamente.

La figura 6.1 muestra la distribución de la Variable BandInf2 antes de la transformación, no sigue una distribución aproximadamente normal.

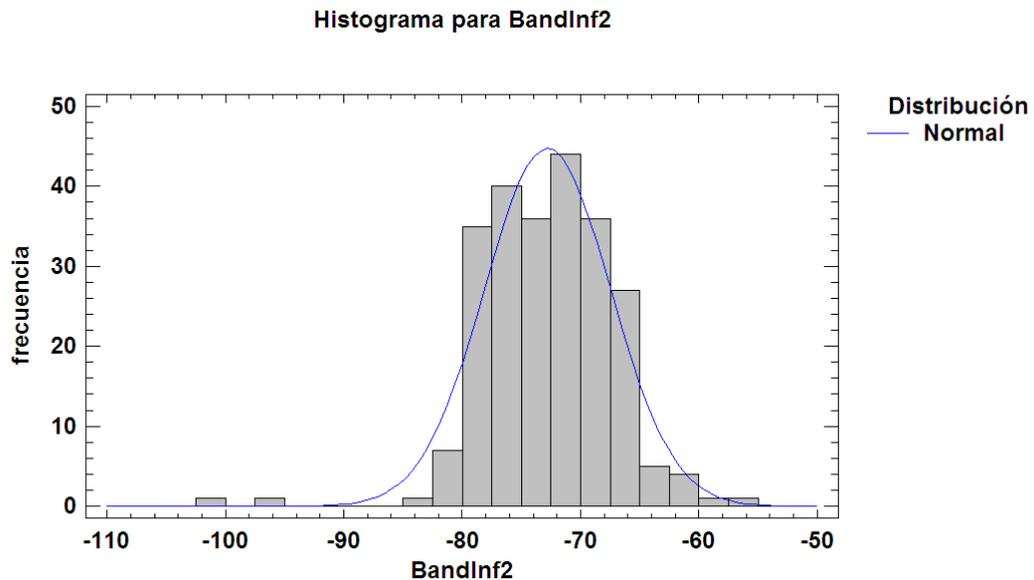


Figura 6.1. Histograma de BanInf2 antes de la transformación.

Después de la transformación, en la figura 6.2, se observa que la variable se aproxima mejor a la normalidad.

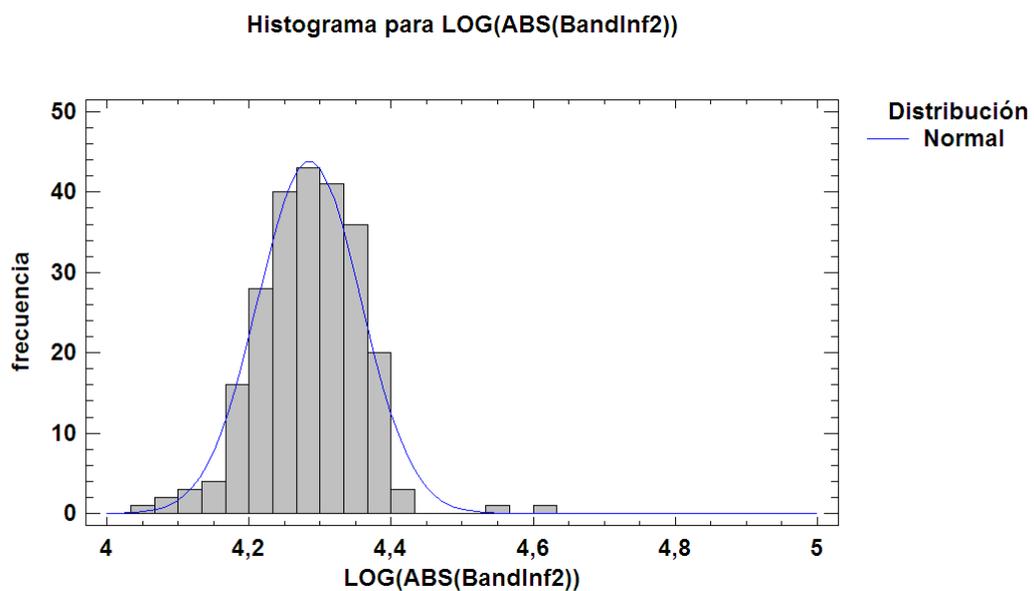


Figura 6.2. Histograma de BandInf2 después de la transformación logarítmica.

6.1.3. Análisis de componentes principales

Tras realizar un análisis de componentes principales con StatGraphics se obtuvieron los siguientes resultados. En la figura 6.3 se observa como existe un grupo grande de variables muy relacionadas entre si (las que ocupan la parte derecha del gráfico) y que además están muy relacionadas con la primera componente principal. Por otra parte las variables Kurtosis, FactorForma y xr parecen estar más relacionadas con la segunda componente principal.

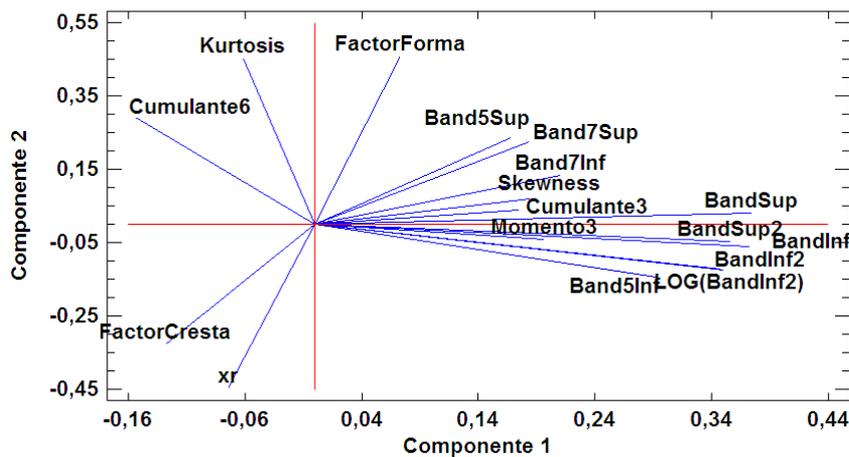


Figura 6.3. Dos primeras componentes principales

Teniendo en cuenta el gráfico de la figura 6.4 y que estamos haciendo un análisis normado parece razonable, utilizando el criterio habitual de extraer aquellas componentes con autovalor mayor que uno que ya mencionamos en el capítulo 5, retener las 5 primeras componentes principales.

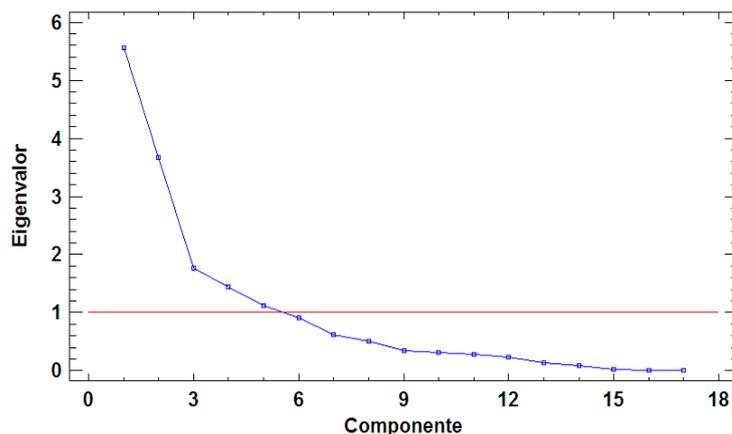


Figura 6.4. Autovalores para cada componente

La tabla 6.3 recoge el porcentaje de variabilidad expresado para cada componente y el acumulado, con 5 componentes principales se recoge el 80% de la variabilidad de los datos. Las figuras 6.5 y 6.6 muestran como se

distribuyen las dos primeras componentes principales para los distintos grupos que consideramos a la hora de clasificar en los capítulos posteriores. En 6.5 se muestra el diagrama de dispersión para las 2 primeras componentes principales discriminando entre R1-R234 mientras que en 6.6 se muestra el diagrama de dispersión para las dos primeras componentes principales discriminando entre R1-R2-R3-R4. Parece claro que estas dos primeras componentes son insuficientes a la hora de clasificar las observaciones ya que no se observa que separen a los distintos grupos considerados.

Componente	Eigenvalor	% de varianza	% acumulado
1	5,55473	32,675	32,675
2	3,66695	21,57	54,245
3	1,76298	10,37	64,616
4	1,4499	8,529	73,144
5	1,1269	6,629	79,773
6	0,905753	5,328	85,101
7	0,62397	3,67	88,772

Tabla 6.3. Autovalores y porcentaje explicado por cada componente

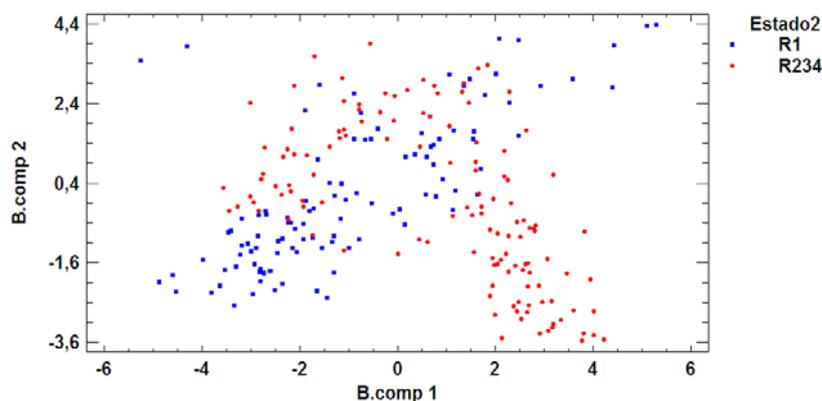


Figura 6.5. Dos primeras componentes principales para el caso R1-R234.

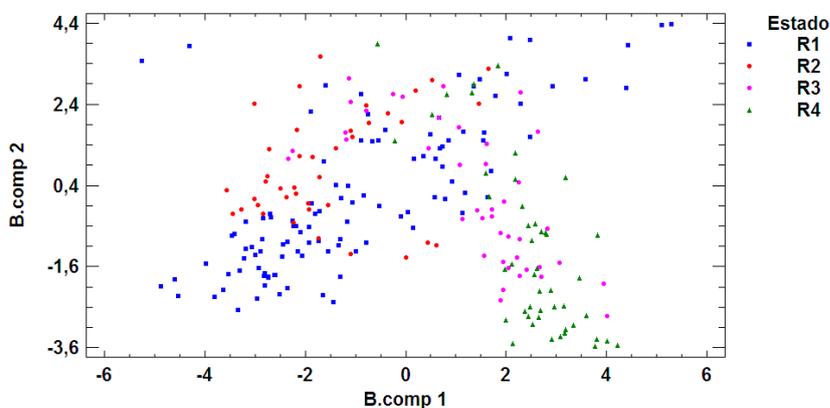


Figura 6.6. Dos primeras componentes principales para el caso R1-R2-R3-R4.

6.2. DESCRIPCIÓN DE LOS ENSAYOS

Se va a estudiar la capacidad de discriminación que tienen algunas de las características medidas, así como su relación con la carga del motor.

Debido al gran número de variables no se estudiarán todas, sino que se han seleccionado unas que, como se verá en el capítulo siguiente tienen gran poder discriminante.

Como variable de la carga del motor se ha seleccionado el deslizamiento.

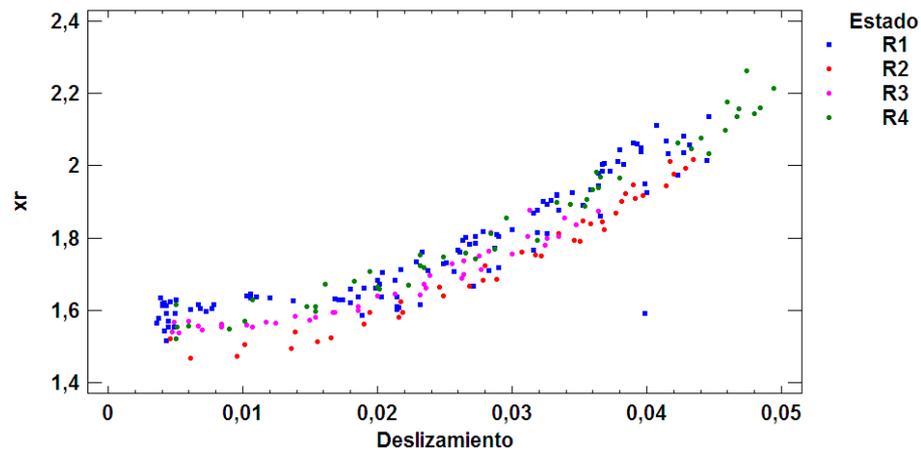


Figura 6.7. Ensayos con el motor. Características xr.

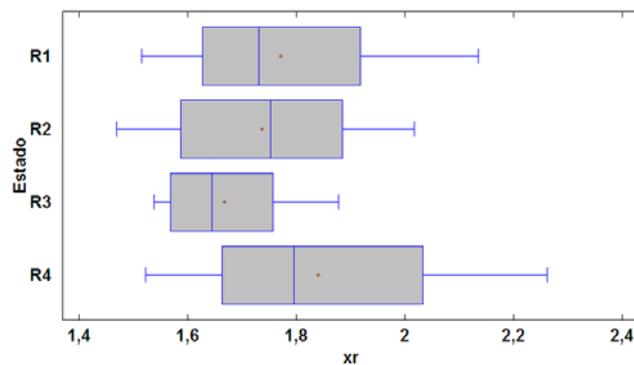


Figura 6.8. Boxplot para xr.

En la figura 6.7 los cuatro estados se encuentran bastante solapados, siendo los más alejados entre sí los estados R2 y R4, lo que serviría para discriminar entre estos dos.

Además como se aprecia en la figura la variable xr tiene una relación casi lineal con la carga, aumentando cuando deslizamiento se hace mayor. Un punto correspondiente al estado R1 parece un outlier.

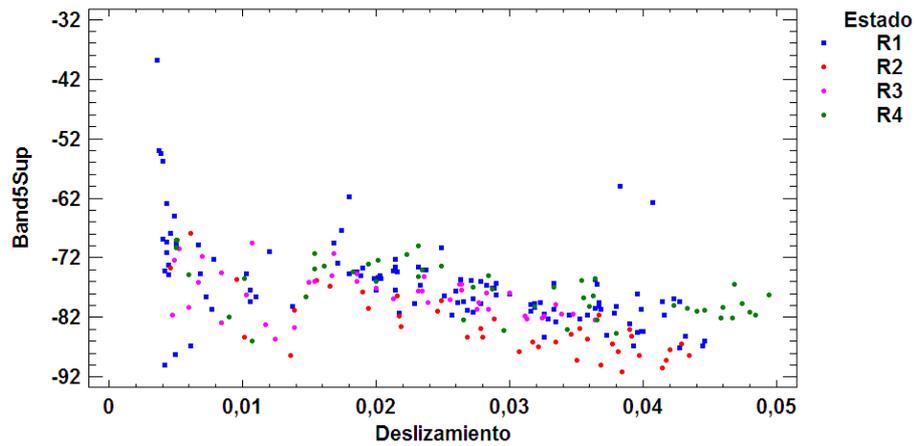


Figura 6.9. Ensayos con el motor. Características Band5Sup.

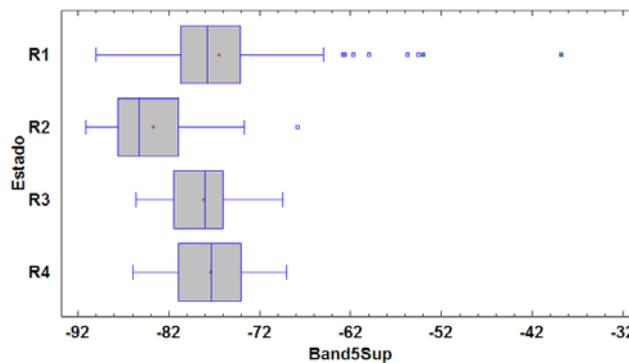


Figura 6.10. Boxplot para band5Sup.

En la figura 6.9 como con la característica anterior los puntos de la Band5Sup se encuentran muy solapados, por lo que sería difícil discriminar.

Por debajo del nivel de deslizamiento de 1,6% aproximadamente la relación de la variable con el deslizamiento es no lineal, en cambio para valores superiores se aprecia una relación lineal, de forma que cuando el deslizamiento aumenta la Band5Sup disminuye.

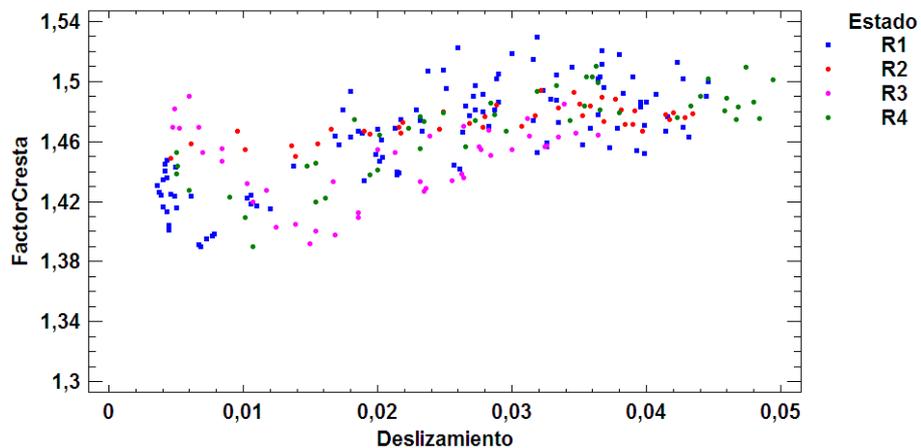


Figura 6.11. Ensayos con el motor. Características Factor Cresta.

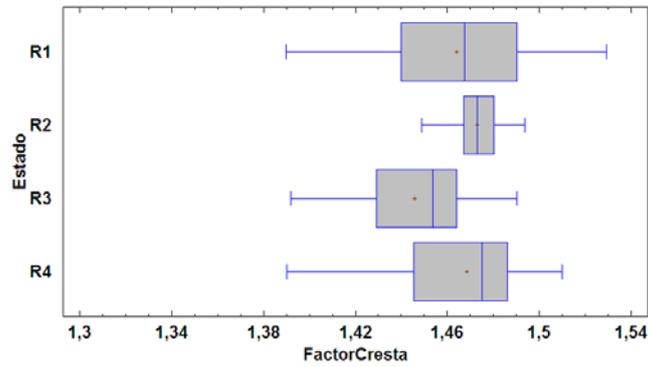


Figura 6.12. Boxplot para Factor Cresta.

En la figura 6.11 se aprecia otro solapamiento de los datos, en este caso no hay ninguna clase que se pueda separar.

El coeficiente de correlación es de 0,68, no se puede decir que haya una relación lineal clara.

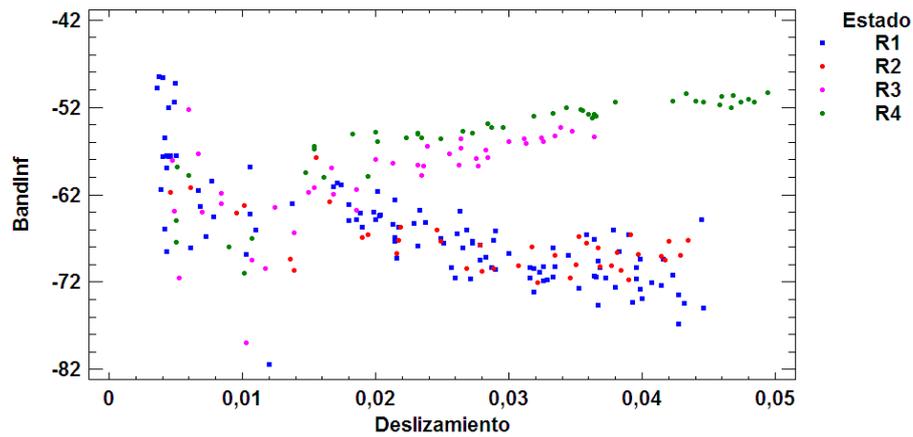


Figura 6.13. Característica BandInf.

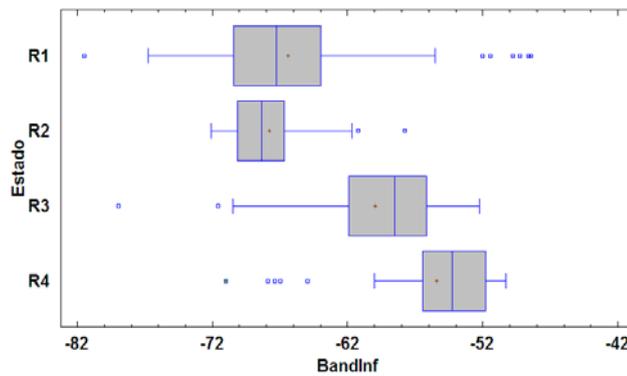


Figura 6.14. Boxplot para BandInf..

La figura 6.13 muestra el diagrama de dispersión de la variable BandInf , a partir del nivel de deslizamiento correspondiente al 1,6% resulta sencillo discriminar entre los niveles R12 y R34, e incluso entre R3 y R4, además para valores superiores al 1,6% parece que hay una relación lineal entre la carga y la BandInf.

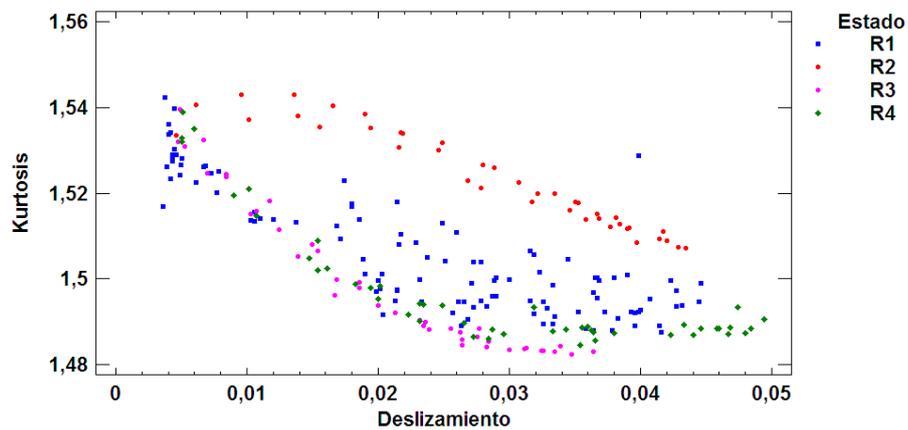


Figura 6.15. Característica Kurtosis.

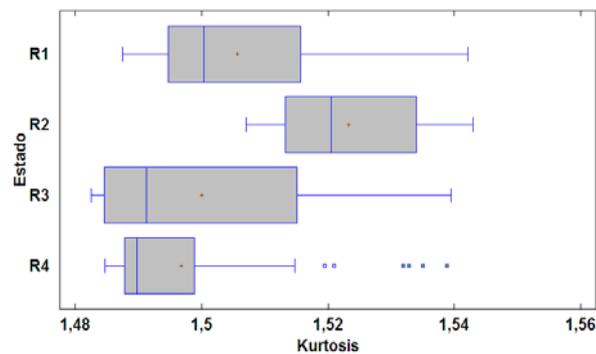


Figura 6.16. Característica Kurtosis.

La variable kurtosis de la figura 6.15 disminuye su valor conforme aumenta el desplazamiento del motor, aunque cada estado lo hace con una pendiente diferente, no existe una relación lineal clara entre las dos variables.

El estado R2 aparece muy alejado de los otros, pero esto no proporciona mucha información.

Capítulo 7

Análisis multivariante

7.1. INTRODUCCIÓN

Lo que se pretende conseguir es la construcción de clasificadores que consigan discriminar entre las siguientes combinaciones de estados:

- **R12-R34**
Diferenciar entre la barra sana o con fallo leve y la barra con fallo medio o rotura grave.
- **R1-R234**
Diferenciar entre una barra sana y una barra con cualquier tipo de fallo o rotura grave.
- **R123-R4**
Diferenciar una barra con rotura grave y todas las demás.
- **R1-R2-R3-R4**
Diferenciar entre los 4 tipos de fallos que se han definido.

Se usarán los métodos paso a paso para intentar conseguir las características con mayor poder discriminante para cada combinación de estados. Posteriormente sobre estas características obtenidas se aplicarán los algoritmos KNN y SVM.

7.2. DISCRIMINACIÓN R12-R34

Método Backward

La selección de las variables más discriminantes en cada paso se hizo de acuerdo al criterio del estadístico F, como se vio en el apartado 5.3, con los parámetros que vienen por defecto en StatGraphics, F para introducir y para eliminar iguales a cuatro.

Si consideramos como medida de la bondad de ajuste el error aparente obtenemos una clasificación correcta en el 91,63% de los casos (no obstante esta es una medida, como ya se comentó en el capítulo 3 excesivamente optimista).

La tabla 7.1 muestra los coeficientes estandarizados. Cuando mayor valor absoluto tenga el coeficiente mayor poder discriminante posee esa característica.

BandInf	0,485216
BandSup2	0,270205
Band5Sup	-0,33921
Cumulante3	0,401941
Cumulante6	-6,66168
Skewness	0,295187
Kurtosis	8,59684
FactorCresta	1,38599
FactorForma	-3,58884

Tabla 7.1. Coeficientes estandarizados para el caso R12-R34.

En los sucesivos se muestra en diagramas de dispersión como la figura 7.1 la relación existente entre las dos características más explicativas.

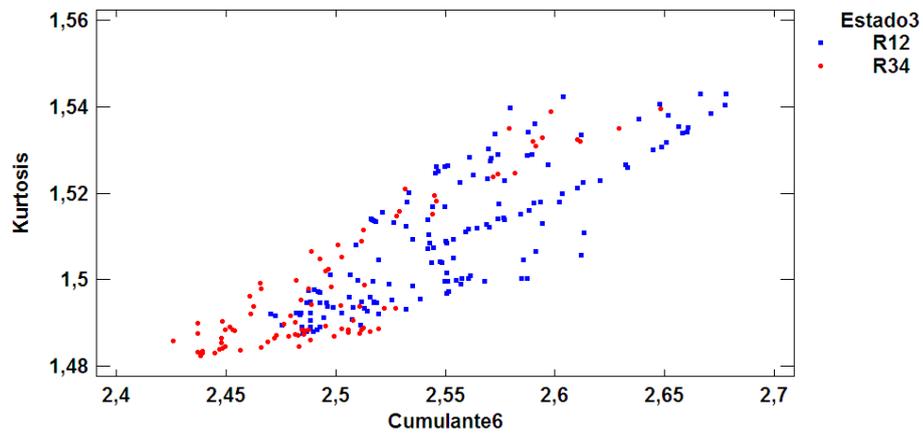


Figura 7.1. Diagrama de dispersión de las dos variables más significativas.

La tabla 7.2 detalla más la clasificación obtenida. Esta tabla es de gran utilidad ya que se podría dar el caso de que el discriminante solo clasifique bien una clase, en este caso los porcentajes de bien clasificados de las dos son similares.

	Tamaño	R12	R34
R12	150	138	12
		92,00%	8,00%
R34	89	8	81
		8,99%	91,01%

Tabla 7.2. Porcentaje de bien/mal discriminados para el caso R12-R34.

Método Forward

Con este método se obtiene exactamente la misma función discriminante por lo que no detallaremos de nuevo los resultados.

7.3. DISCRIMINACIÓN R1-R234

Método Backward

Con los parámetros que vienen por defecto en StatGraphics, F para introducir y para eliminar igual a cuatro.

Si consideramos como medida de la bondad de ajuste el error aparente obtenemos una clasificación correcta en el 87,45% de los casos (no obstante esta es una medida excesivamente optimista).

La tabla 7.3 muestra los coeficientes estandarizados que se obtuvieron, cuanto mayor valor absoluto del coeficiente tenga una característica mayor será su poder discriminante.

BandSup	-1,10684
Band5Sup	0,928338
Band5Inf	0,27623
Momento3	-0,413331
xr	0,594608
FactorForma	0,407863

Tabla 7.3. Coeficientes estandarizados para el caso R1-R234.

La figura 7.2 muestra el diagrama de dispersión para las dos características más explicativas.

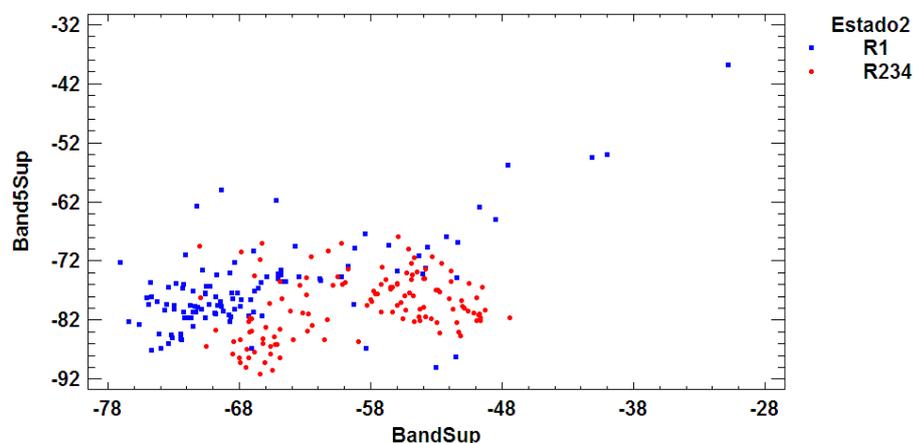


Figura 7.2. Características más explicativas.

La tabla 7.4 corresponde a la tabla de contingencia obtenida. Se comprueba que no existe ningún sesgo apreciable en la discriminación.

	Tamaño	R1	R234
R1	110	98	12
		89,09%	10,91%
R234	129	18	111
		13,95%	86,05%

Tabla 7.4. Bien/mal clasificados para cada clase para el caso R1-R234.

Método Forward

Con los parámetros que vienen por defecto en StatGraphics, F para introducir y para eliminar igual a cuatro.

Tomando como medida de bondad de ajuste el error aparente obtenemos una clasificación correcta el 84,94% de los casos.

La tabla 7.5 muestra los coeficientes estandarizados que se obtuvieron, cuanto mayor valor absoluto del coeficiente tenga una característica mayor será su poder discriminante.

BandSup	-1,09106
Band5Sup	0,883506
Band5Inf	0,255549
Skewness	-0,478891

Tabla 7.5. Coeficientes estandarizados para el caso R1-R234.

El diagrama de dispersión para el método Forward sería también el de la figura 7.2. La tabla 7.6 representa la tabla de contingencia, donde no existe ningún sesgo apreciable.

	Tamaño	R1	R234
R1	110	95	15
		86,36%	13,64%
R234	129	21	108
		16,28%	83,72%

Tabla 7.6. Bien/mal clasificados para el caso R1-R234.

7.4. DISCRIMINACIÓN R123-R4

Método Forward

Con los parámetros que vienen por defecto en StatGraphics, F para introducir y para eliminar iguales a cuatro.

Tomando como medida de bondad de ajuste el error aparente obtenemos una clasificación correcta el 86,61% de los casos.

La tabla 7.7 muestra los coeficientes estandarizados que se obtuvieron, cuanto mayor valor absoluto del coeficiente tenga una característica mayor será su poder discriminante.

FactorForma	-0,632102
BandSup2	0,571566
BandInf	0,461623
Momento3	0,216844
Cumulante3	0,300352
Kurtosis	0,436752
xr	0,35787

Tabla 7.7. Coeficientes estandarizados para el caso R123-R4.

La figura 7.3 muestra el diagrama de dispersión para las dos características más explicativas.

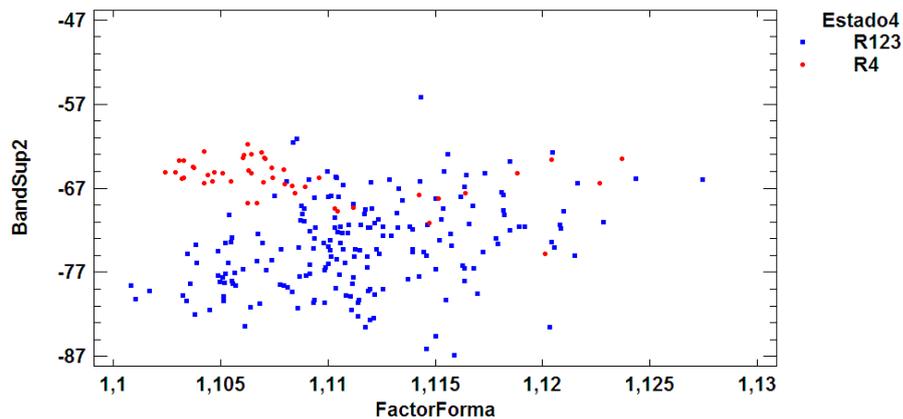


Figura 7.3. Características más explicativas

En la tabla 7.8 sí que se aprecia una diferencia significativa, clasificando mejor los R123 que los R4.

	Tamaño	R123	R4
R123	192	172	20
		89,58%	10,42%
R4	47	12	35
		25,53%	74,47%

Tabla 7.8. Bien/mal clasificados para el caso R123-R4.

Método Backward

Con los parámetros que vienen por defecto en StatGraphics, F para introducir y para eliminar iguales a cuatro. Tomando como medida de bondad de ajuste el error aparente obtenemos una clasificación correcta el 86,61% de los casos.

La tabla 7.9 muestra los coeficientes estandarizados que se obtuvieron, cuanto mayor valor absoluto del coeficiente tenga una característica mayor será su poder discriminante.

BandInf	0,478593
BandSup2	0,560417
Cumulante3	0,304937
Skewness	0,250204
Kurtosis	0,433534
xr	0,431755
FactorForma	-0,611744

Tabla 7.9. Coeficientes estandarizados para el caso R123-R4 .

La figura 7.4 muestra el diagrama de dispersión para las dos características más explicativas.

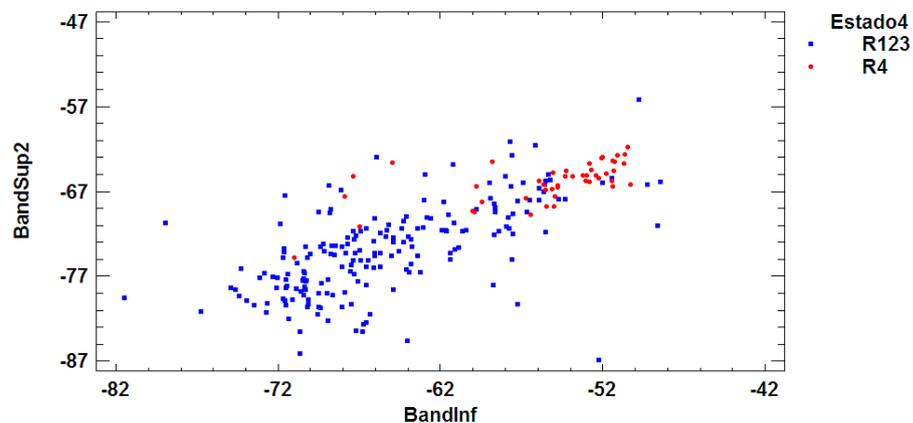


Figura 7.4. Características más explicativas

En la tabla 7.10 se comprueba que el discriminante tiene problemas a la hora de clasificar el estado R4, fallando un 23,4% de las veces.

	Tamaño	R123	R4
R123	192	171	21
		89,06%	10,94%
R4	47	11	36
		23,40%	76,60%

Tabla 7.10. Bien/mal clasificados para el caso R123-R4.

7.5. DISCRIMINACIÓN R1-R2-R3-R4

Método Backward

Con los parámetros que vienen por defecto en StatGraphics, F para introducir y para eliminar iguales a cuatro.

Tomando como medida de bondad de ajuste el error aparente obtenemos una clasificación correcta el 85,36% de los casos.

La tabla 7.11 muestra los coeficientes estandarizados que se obtuvieron, cuanto mayor valor absoluto del coeficiente tenga una característica mayor será su poder discriminante.

Kurtosis	9,23885	0,507058	-1,93651
Cumulante6	-6,99634	-0,950987	1,98891
BandSup	-0,21583	-0,729582	-0,552852
BandInf	0,560902	0,0491163	0,135506
Band5Sup	0,372213	0,843727	0,817747
BandSup2	0,253687	-0,224958	0,466376
Band5Inf	-0,231816	-0,017259	0,494966
Band7Sup	-0,360587	-0,0699328	-0,30525
Momento3	0,0177937	-0,39813	-0,0385572
Cumulante3	0,449776	-0,0268165	0,141852
xr	-0,145865	0,132206	0,981366
FactorCresta	1,78978	0,28691	-0,635286
FactorForma	-3,93993	0,298561	0,493788

Tabla 7.11. Coeficientes estandarizados para el caso R1-R2-R3-R4.

En la tabla 7.12 sorprende que haya un 8,51% de R4 que hayan sido clasificados como R1, cuando se supone que son los estados más diferentes.

	Tamaño	R1	R2	R3	R4
R1	110	96	6	4	4
		87,27%	5,45%	3,64%	3,64%
R2	40	0	40	0	0
		0,00%	100,00%	0,00%	0,00%
R3	42	3	1	36	2
		7,14%	2,38%	85,71%	4,76%
R4	47	4	2	9	32
		8,51%	4,26%	19,15%	68,09%

Tabla 7.12. Coeficientes estandarizados para el caso R1-R2-R3-R4.

La figura 7.5 muestra el diagrama de dispersión para las dos características más explicativas.

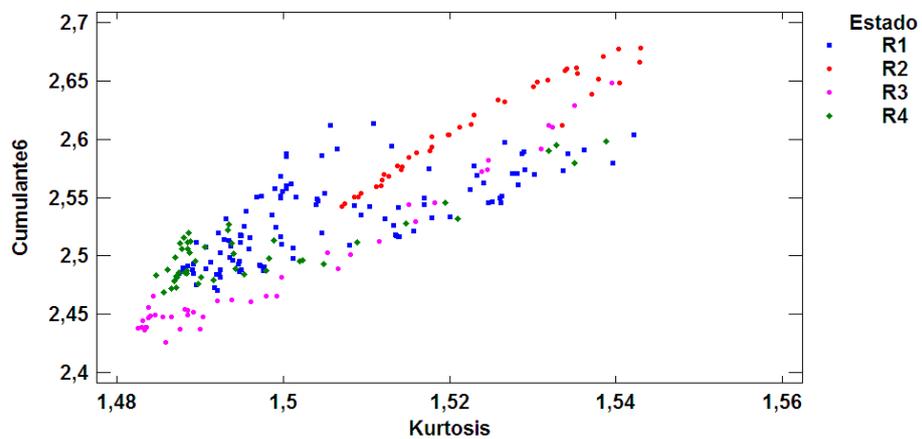


Figura 7.5. Características más explicativas

La tabla 7.6 muestra las funciones discriminantes obtenidas, es curioso que se encuentre más cerca de los estados R3 y R4 la del R1 y no R2, cuando sería previsible que estuviera más alejada por existir mayores diferencias.

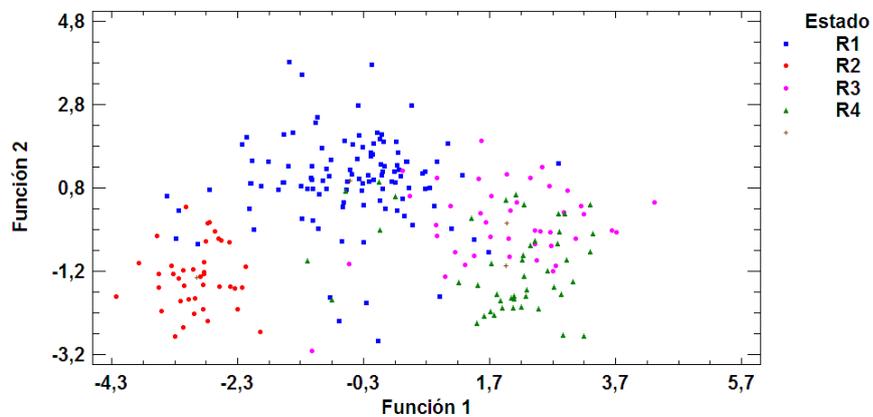


Figura 7.6. Funciones discriminantes.

Método Forward

Con los parámetros que vienen por defecto en StatGraphics, F para introducir y para eliminar iguales a cuatro

Tomando como medida de bondad de ajuste el error aparente obtenemos una clasificación correcta el 85,77% de los casos.

La tabla 7.13 muestra los coeficientes estandarizados que se obtuvieron, cuanto mayor sea el valor absoluto mayor poder discriminante tiene esa variable.

Kurtosis	9,13013	0,898356	-1,93016
Cumulante6	-6,91211	-1,22849	1,97853
BandSup	-0,211376	-0,727325	-0,563574
BandInf	0,566542	0,0269399	0,134451
Band5Sup	0,36498	0,851218	0,83084
BandSup2	0,250089	-0,218952	0,46259
Band5Inf	-0,230303	-0,0453435	0,492677
Band7Sup	-0,360782	-0,0611578	-0,30516
Cumulante3	0,447581	-0,0408173	0,1384
Skewness	0,0774376	-0,394833	-0,0265413
xr	-0,127304	0,0171303	0,976219
FactorCresta	1,76346	0,337113	-0,636196
FactorForma	-3,90271	0,0953707	0,49163

Tabla 7.13. Coeficientes estandarizados para el caso R1-R2-R3-R4.

	Tamaño	R1	R2	R3	R4
R1	110	96	7	5	2
		87,27%	6,36%	4,55%	1,82%
R2	40	0	40	0	0
		0,00%	100,00%	0,00%	0,00%
R3	42	2	1	37	2
		4,76%	2,38%	88,10%	4,76%
R4	47	4	2	9	32
		8,51%	4,26%	19,15%	68,09%

Tabla 7.14. Bien/mal clasificados para el caso R1-R2-R3-R4.

La tabla 7.15 muestra un resumen de los diferentes errores aparentes que se han calculado.

	R12-R34	R1-R234	R123-R4	R1-R2-R3-R4
Método Backward	0.0837	0.1255	0.1339	0.1464
Método Fordward	0.0837	0.1506	0.1339	0.1423

Tabla 7.15. Resumen errores aparentes.

7.6. ERROR ESTIMADO LDA

Estos resultados han sido obtenidos con el software R, considerando como medida de la bondad de ajuste el error estimado por remuestreo se ha programado el método de validación bootstrapping en R (ver anexos), tomando un 80% (139 muestras) de los datos como entrenamiento y el restante 20% (50 muestras) como test, todo de manera aleatoria, durante 200 iteraciones.

En la tabla 7.16 se indica la tasa de error estimado que fue calculada con bootstrapping.

	R12-R34	R1-R234	R123-R4	R1-R2-R3-R4
Método Backward	0.0837	0.1467	0.1102	0.1712
Método Fordward	0.0837	0.1509	0.1048	0.1687

Tabla 7.16. Errores estimados.

La figura 7.7 muestra los diagramas de cajas y bigotes de los errores estimados durante las 200 iteraciones para cada clasificación y para cada grupo de variables, como se indica debajo de cada diagrama. Los mejores errores estimados son a su vez los que mejor dispersión presentan después de las 200 iteraciones.

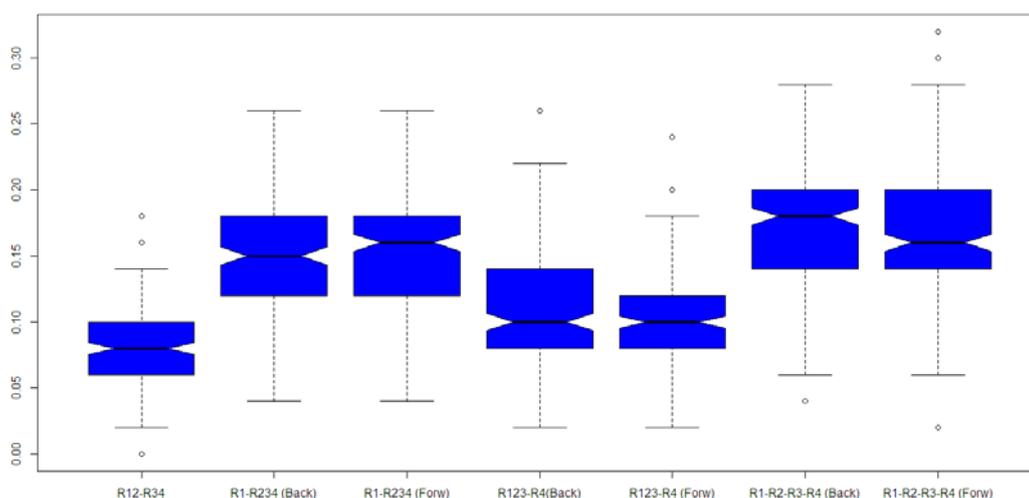


Figura 7.7. Errores estimados para el LDA.

En la tabla 7.17 se comparan los mejores errores aparentes y los mejores errores estimados para cada caso. Se comprueba que el cambio en el error de validación no es excesivo y en algunos casos se comporta mejor el error estimado lo que indica que el sesgo que aporta el error aparente no es muy grande.

	R12-R34	R1-R234	R123-R4	R1-R2-R3-R4
Errores aparentes	0.0837	0.1255	0.1339	0.1423
Errores estimados	0.0837	0.1467	0.1048	0.1687

Tabla 7.17. Comparativa errores aparentes y errores estimados.

En el siguiente capítulo se aplicarán a cada combinación de fallos los dos algoritmos de machine learning elegidos a las características que se han ido seleccionando en los diferentes apartados de este capítulo.

Para trabajar en R con las diferentes características seleccionadas arriba se fueron guardando en archivos de texto, ejemplo, "backR1R234.txt". En ninguno de los análisis de este capítulo han sido elegidas las características BandInf2, BandInf7 ni cumalante2 de lo que se desprende que no tienen apenas poder discriminante.

Capítulo 8

Aplicación de KNN y SVM

8.1. INTRODUCCIÓN

En este capítulo se van a describir y comentar los modelos y resultados obtenidos con las técnicas multivariantes de los k vecinos más próximos (KNN) y el Support Vector Machine (SVM). Los fundamentos teóricos de estas técnicas fueron expuestos en el capítulo 5 de este trabajo. Su aplicación en el contexto de la clasificación de los fallos en motores constituye una de las aportaciones fundamentales de este trabajo de fin de grado.

8.2. K-VECINOS MÁS PRÓXIMOS

8.2.1. Modelo y su error aparente

Utilizando la técnica de los k-vecinos más próximos con diferentes números de vecinos, distancia euclídea y leave-one-out como método de validación se han obtenido los siguientes resultados, que no son más que el porcentaje de error para cada vecino utilizando los datos con los que se construyó el modelo, es decir, el error aparente, que como se comentó en apartados anteriores suele dar un error más optimista pues se está evaluando el modelo con los mismos datos con los que ha sido construido.

Todos los vecinos son impares, así el algoritmo funcionaba mejor en R (ver código en el anexo), y su número ha sido elegido como en [FIL09].

Clasificación R12-R34

La tabla 8.1 muestra los errores aparentes para cada vecino cuando se aplicó el algoritmo a todas las características y a las que fueron seleccionadas con el método backward en el apartado 7.2.

R12-R34		
K	ALL	BACK
1	0.1213	0.1464
3	0.1422	0.1338
5	0.1255	0.1255
7	0.1255	0.1297
9	0.1338	0.1297
13	0.1506	0.1338
15	0.1380	0.1297
21	0.1506276	0.1422594
25	0.1464435	0.1380753
29	0.1631799	0.1338912

Tabla 8.1. Error aparente para el caso R12-R34.

La figura 8.1 no es más que la representación de la tabla anterior para facilitar el análisis, se puede observar el efecto que tiene el número de vecinos en el comportamiento del error aparente, se comprueba que el mejor error se obtiene aplicando el método con un solo vecino a todas las características obteniendo un error aparente igual a 0.1213. El error aparente tiende a aumentar conforme lo hace el número de vecinos.

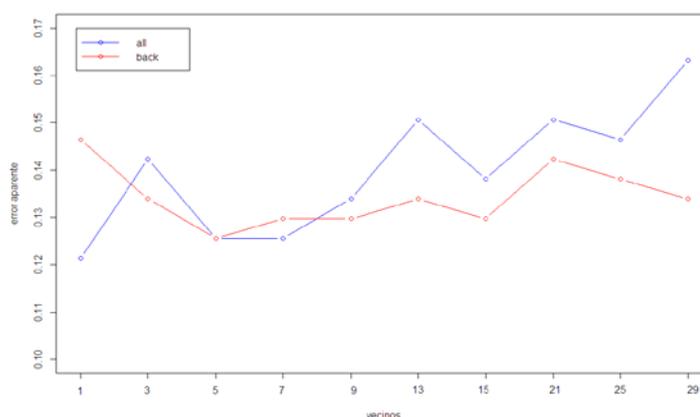


Figura 8.1. Efecto del n° de vecinos. R12-R34.

Clasificación R1-R234

La tabla 8.2 muestra los errores aparentes para cada vecino cuando se aplicó el algoritmo a todas las características, a las que fueron seleccionadas con el método Backward y a las seleccionadas tras el Forward, en el apartado 7.3.

R123-R4			
K	ALL	BACK	FORW
1	0.1046	0.1380	0.1380
3	0.1004	0.1129	0.1129
5	0.1129	0.1380	0.1380
7	0.1129	0.1297	0.1297
9	0.1213	0.1297	0.1297
13	0.1297	0.1338	0.1338
15	0.1380	0.1422	0.1422
21	0.1464	0.1422	0.1422
25	0.1799	0.1464	0.1464
29	0.2008	0.1631	0.1631

Tabla 8.2. Error aparente para el caso R123-R4.

La figura 8.2 muestra el efecto que tiene el número de vecinos en el error aparente, a partir del vecino tercero el añadir más vecinos hace que el error

aparente aumento. El mejor error, 0.1004, se ha obtenido con un número de vecinos igual a tres cuando se aplicó el algoritmo a todas las características.

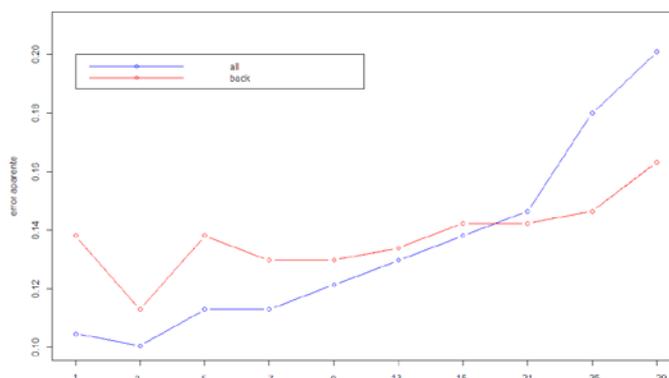


Figura 8.2. Efecto del nº de vecinos para el caso R1-R234.

Clasificación R123-R4

La tabla 8.3 muestra los errores aparentes para cada vecino cuando se aplicó el algoritmo a todas las características, a las que fueron seleccionadas con el método Backward y a las seleccionadas tras el Forward, en el apartado 7.4.

R123-R4			
K	ALL	BACK	FORW
1	0.1422	0.1422	0.1422
3	0.1129	0.1422	0.1422
5	0.1171	0.1422	0.1422
7	0.1087	0.1297	0.1297
9	0.1004	0.1338	0.1338
13	0.1046	0.1422	0.1422
15	0.1087	0.1380	0.1380
21	0.1213	0.1297	0.1297
25	0.1255	0.1422	0.1422
29	0.1213	0.1338	0.1338

Tabla 8.3. Error aparente para el caso R123-R4.

En el gráfico de la figura 8.3 se comprueba que no existe ninguna relación entre el error aparente y el número de vecinos usados por el algoritmo para esta clasificación. El mejor error es 0,1004 que se da para los nueve vecinos más próximos aplicados a todas las características.

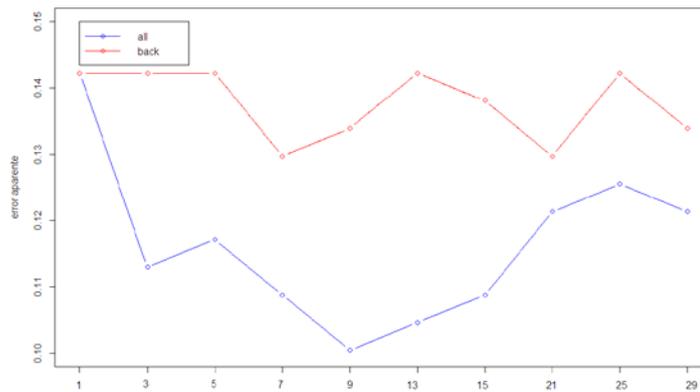


Figura 8.3. Efecto del n° de vecinos para el caso R123-R4.

Clasificación R1-R2-R3-R4

La tabla 8.4 muestra los errores aparentes para cada vecino cuando se aplicó el algoritmo a todas las características, a las que fueron seleccionadas con el método Backward y a las seleccionadas tras el Forward, en el apartado 7.5.

R1-R2-R3-R4			
K	ALL	BACK	FORW
1	0.2426	0.2803	0.2803
3	0.2259	0.2259	0.2175
5	0.2426	0.2301	0.2259
7	0.2468	0.2217	0.2259
9	0.2217	0.2217	0.2301
13	0.2384	0.2217	0.2301
15	0.2510	0.2050	0.20920
21	0.2970	0.2217	0.2301
25	0.3138	0.2384	0.2343
29	0.3472	0.2761	0.2594

Tabla 8.4. Error aparente para el caso R1-R2-R3-R4.

En la figura 8.4 se muestra el efecto del número de vecinos en el comportamiento del error aparente, el mejor error igual a 0,255 se da para los 15 vecinos más próximos aplicado a las características seleccionadas por el método backward.

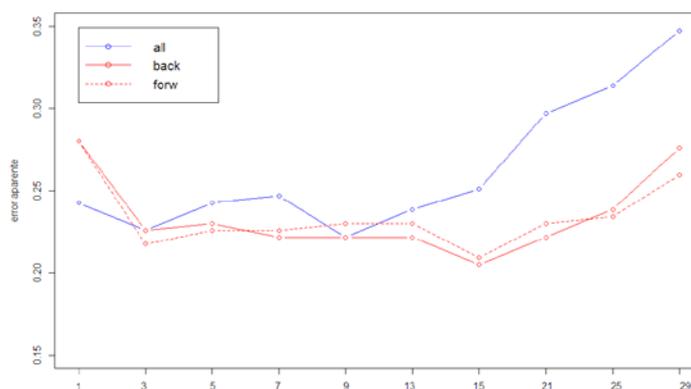


Figura 8.4. Efecto del nº de vecinos para el caso R1-R2-R3-R4.

8.1.2. Error estimado KNN

En este apartado se va a estudiar el comportamiento del algoritmo k vecinos más próximos (KNN) utilizando como medida de la bondad de ajuste el error estimado con el método de validación bootstrapping, que fue explicado en el apartado 3.3 de este trabajo. El bootstrapping fue programado en R donde se hicieron 200 iteraciones (ver código en anexo) para cada vecino dividiendo el conjunto de datos en dos partes, una de entrenamiento formada por el 80% de los datos, (139 muestras) y otra de test compuesta por el restante 20% (50 muestras) de manera aleatoria. A continuación se mostrarán los resultados obtenidos para cada clasificación.

Clasificación R12-R34

La tabla 8.5 muestra los errores estimados para cada vecino aplicando el algoritmo a todas las características y a las características seleccionadas tras el método backward, que se obtuvieron en el apartado 7.2 de este trabajo.

R12-R34		
K	ALL	BACK
1	0.1224	0.1573
3	0.1361	0.1367
5	0.1317	0.1357
7	0.1315	0.1291
9	0.1383	0.1327
13	0.1457	0.1315
15	0.1494	0.1322
21	0.1607	0.1357
25	0.1736	0.1366
29	0.1823	0.1383

Tabla 8.5. Error estimado para el caso R12-R34

La figura 8.5 representa el efecto en el error estimado del número de vecinos, el error tiene a aumentar conforme lo hace el número de vecinos.

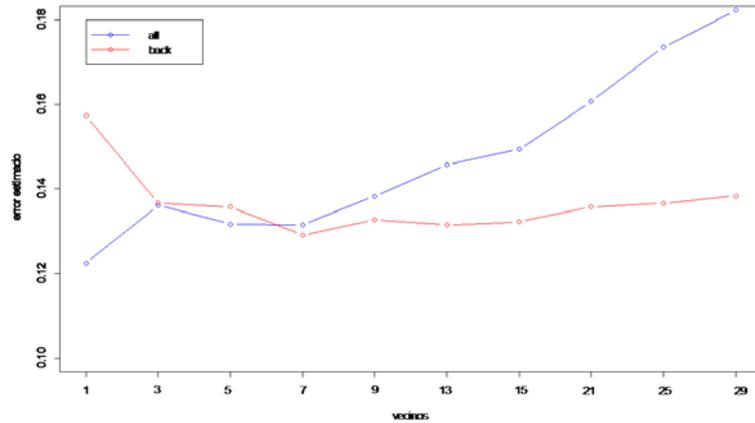


Figura 8.5. Efecto del n° de vecinos para el caso RR12-R34.

Las figuras 8.6 y 8.7 muestran la dispersión en sendos diagramas de cajas y bigotes de los errores estimados para cada vecino, con esto se pretende obtener una mejor interpretación de estos resultados, comparándolos entre sí y observando el efecto del número de vecinos en la dispersión de estos. Con la gráfica anterior y los diagramas se comprueba que el mejor error estimado se consigue aplicando el algoritmo a todas las características para un solo vecino que es igual a 0,1224, ya que su error es el menor de todos y la dispersión de sus errores se comporta mejor que la del mejor error estimado utilizando las características seleccionadas por el Backward, que es 0,1291 para un número de vecinos igual a siete

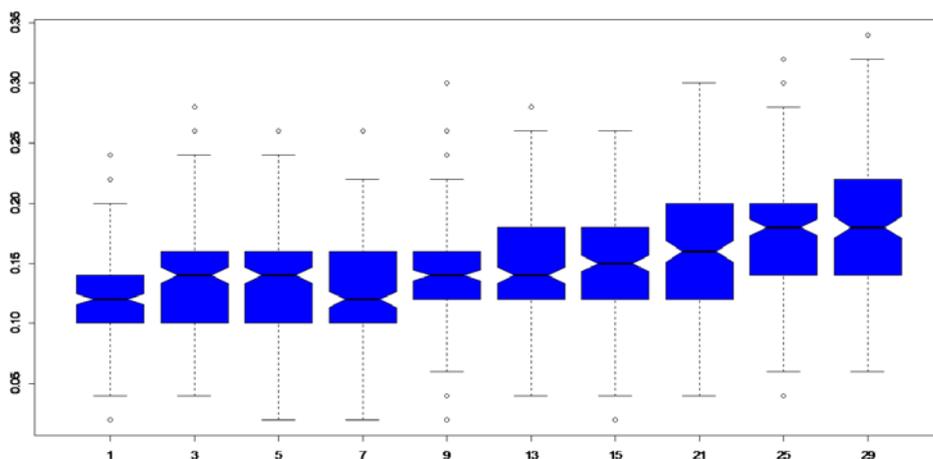


Figura 8.6. Dispersión de errores usando todas las características para el caso R12-R34.

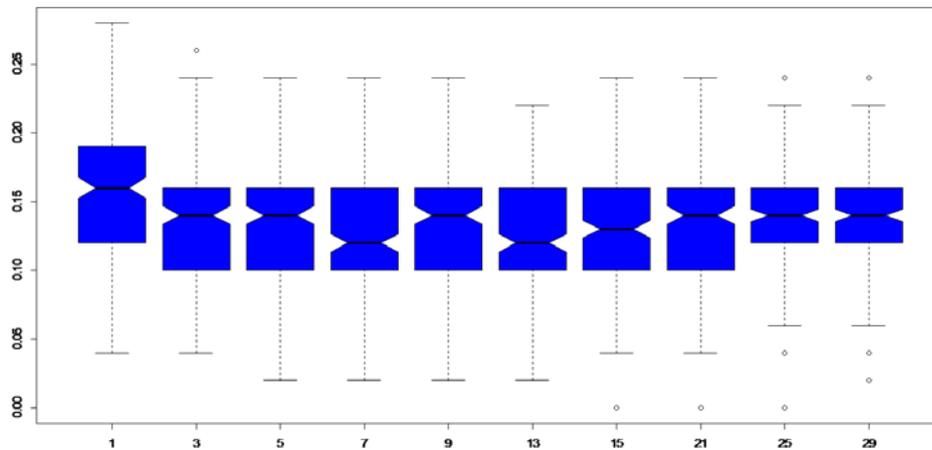


Figura 8.7. Dispersion de errores usando las características del Backward para el caso R12-R34.

Clasificación R1-R234

La tabla 8.6 muestra los errores estimados para cada vecino aplicando el algoritmo a todas las características, a las características seleccionadas tras el método Backward y a las características seleccionadas tras el método Forward que se obtuvieron en el apartado 7.3 de este trabajo.

R1-R234			
K	ALL	BACK	FORW
1	0.1118	0.1372	0.1400
3	0.1081	0.1288	0.1251
5	0.1151	0.1376	0.1340
7	0.1157	0.1366	0.1372
9	0.1226	0.1338	0.1313
13	0.1393	0.1399	0.1398
15	0.1467	0.1443	0.1451
21	0.1823	0.1574	0.1579
25	0.2076	0.1671	0.1662
29	0.2270	0.1750	0.1746

Tabla 8.6. Error estimado para el caso R1-R234.

La figura 8.8 muestra el efecto del número de vecinos en el error estimado, se comprueba que a partir del vecino nueve los errores comienzan a aumentar.

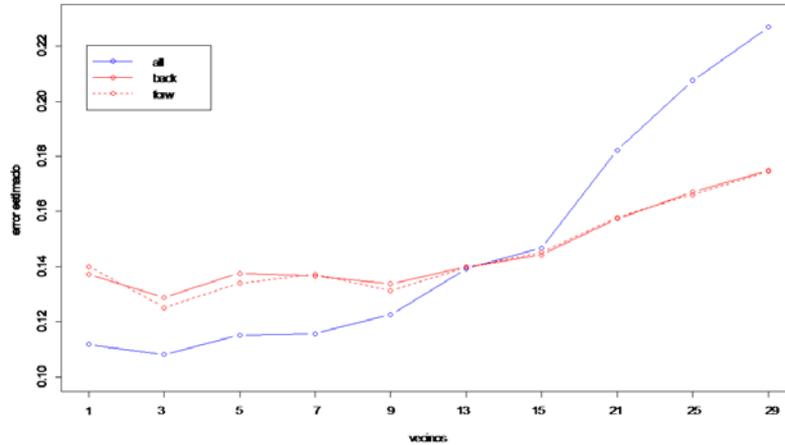


Figura 8.8. Efecto del n° de vecinos para el caso R1-R234.

Las figuras 8.9 y 8.10 muestran la dispersión de los errores, el menor error se comete cuando se aplica el algoritmo a todas las características y el número de vecinos es igual a tres, ya que es el menor error, igual a 0,1081, además se puede apreciar que su dispersión es mejor que para los mejores errores cuando el algoritmo trabaja sobre las características del Backward y del Forward, que también se dan para un número de vecinos igual a tres.

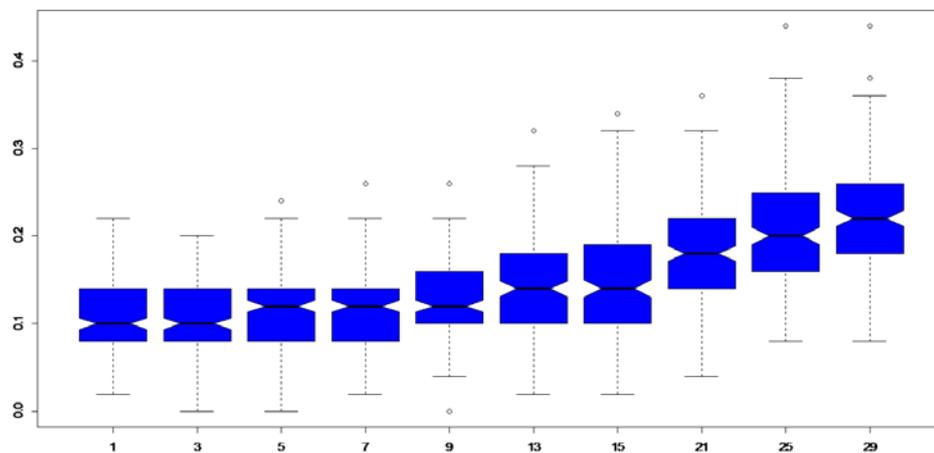


Figura 8.9. Dispersión de errores usando todas las características para el caso R1-R234.

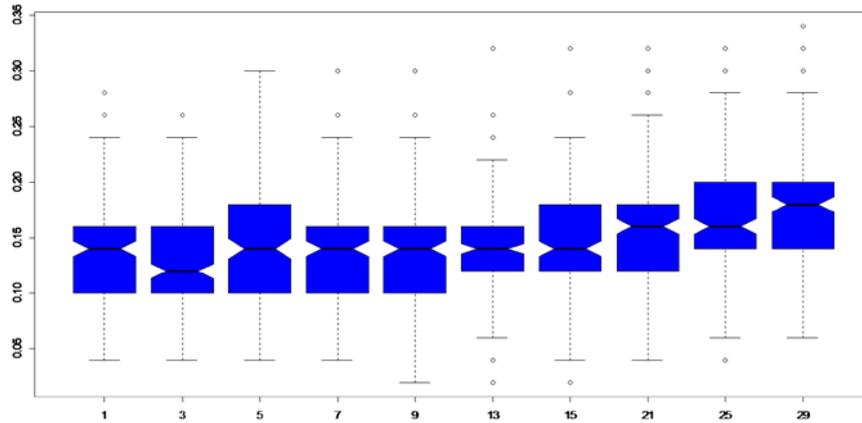


Figura 8.10. Dispersión de errores usando las características del Backward para el caso R1-R234.

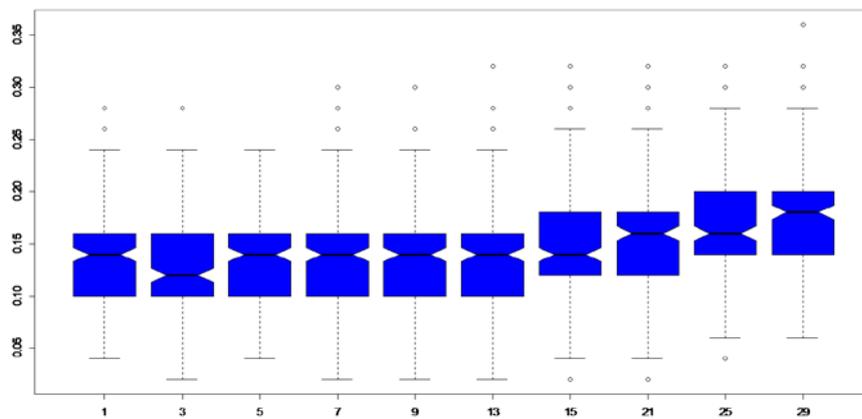


Figura 8.11. Dispersión de errores usando las características del Forward para el caso R1-R234

Clasificación R123-R4

La tabla 8.7 muestra los errores estimados para cada vecino aplicando el algoritmo a todas las características, a las características seleccionadas tras el método Backward y a las características obtenidas tras el método Forward que se obtuvieron en el apartado 7.4 de este trabajo.

R123-R4			
K	ALL	BACK	FORW
1	0.1394	0.1450	0.1469
3	0.1125	0.1422	0.1405
5	0.1156	0.1392	0.1383
7	0.1080	0.1366	0.1350
9	0.1052	0.1347	0.1327
13	0.1126	0.1347	0.1310
15	0.1154	0.1326	0.1298
21	0.1232	0.1348	0.1323
25	0.1325	0.1337	0.1310
29	0.1371	0.1323	0.1299

Tabla 8.7. Error estimado para el caso R123-R4.

La figura 8.12 muestra el efecto del número de vecinos en el error estimado, se puede apreciar que menos para los dos últimos vecinos el error estimado cuando se aplica el algoritmo a todas las características es siempre mejor que cuando se aplica a las variables obtenidas tras los métodos de selección.

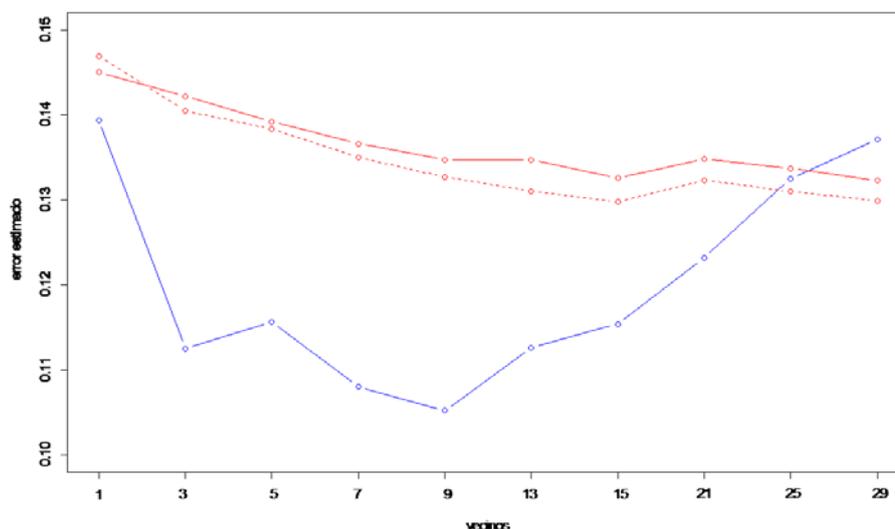


Figura 8.12. Efecto del número de vecinos para el caso R123-R4.

En este caso el mejor resultado se obtiene, sin duda, para los nueve vecinos más cercanos cuando se aplica el algoritmo a todas las características, igual a 0,1052. La figura 8.13 muestra la dispersión de los errores cuando se utilizan todas las características, se mantiene constante hasta que pasa el vecino nueve donde comienza a aumentar. En las otras dos figuras 8.14 y 8.15 la dispersión de los errores se mantiene bastante constante.

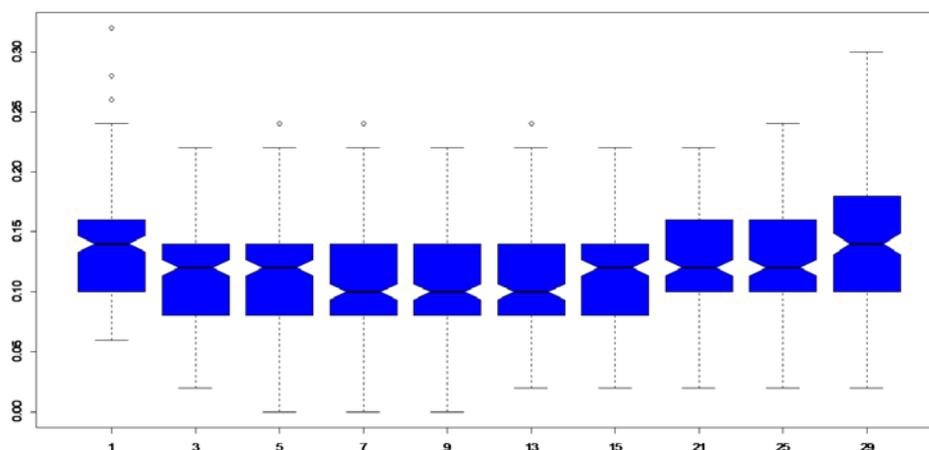


Figura 8.13.. Dispersión de errores usando todas las características para el caso R123-R4.

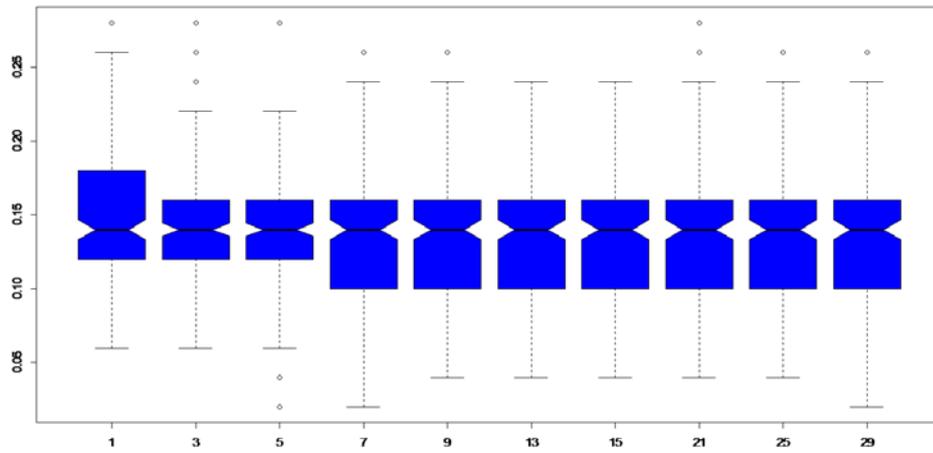


Figura 8.14. Dispersión de errores usando las características del Backward para el caso R123-R4

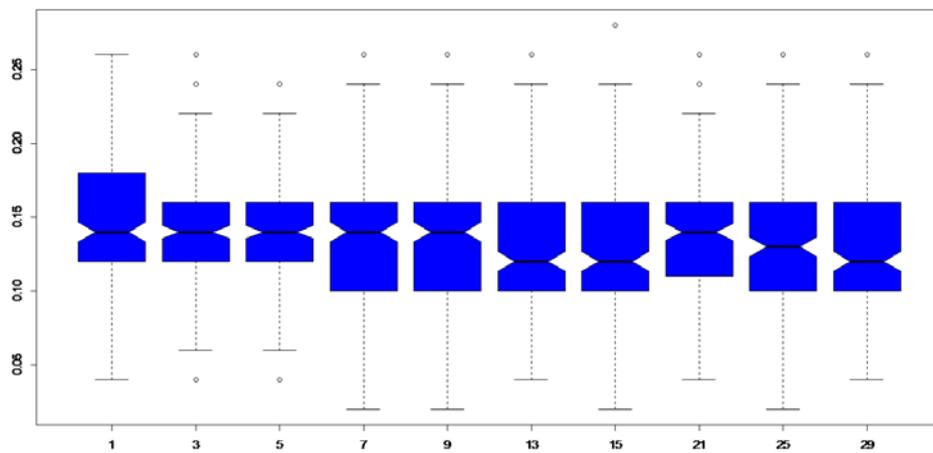


Figura 8.15. Dispersión de errores usando las características del Forward para el caso R123-R4

Clasificación R1-R2-R3-R4

La tabla 8.8 muestra los errores estimados para cada vecino aplicando el algoritmo a todas las características, a las características seleccionadas tras el método backward y a las características obtenidas tras el método Forward que se obtuvieron en el apartado 7.5 de este trabajo.

R1-R2-R3-R4			
K	ALL	BACK	FORW
1	0.2491	0.2796	0.2805
3	0.2464	0.2291	0.2338
5	0.2466	0.2340	0.2357
7	0.2524	0.2266	0.2303
9	0.2490	0.2258	0.2333
13	0.2761	0.2278	0.2302
15	0.2876	0.2281	0.2327
21	0.3352	0.2528	0.2522
25	0.3716	0.2719	0.2694
29	0.3922	0.2891	0.2832

Tabla 8.8. Error estimado para el caso R1-R2-R3-R4.

La figura 8.16 muestra el efecto del número de vecinos en el error estimado, se comprueba que en esta ocasión cuando aplicamos el algoritmo k vecinos más próximos a las características seleccionadas por los métodos de selección se consiguen los mejores resultados.

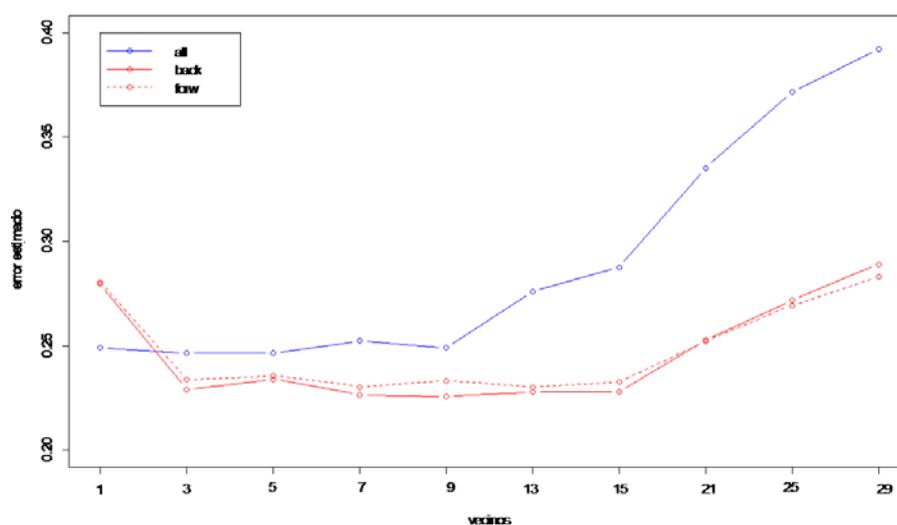


Figura 8.16. Efecto del nº de vecinos para el caso R1-R2-R3-R4.

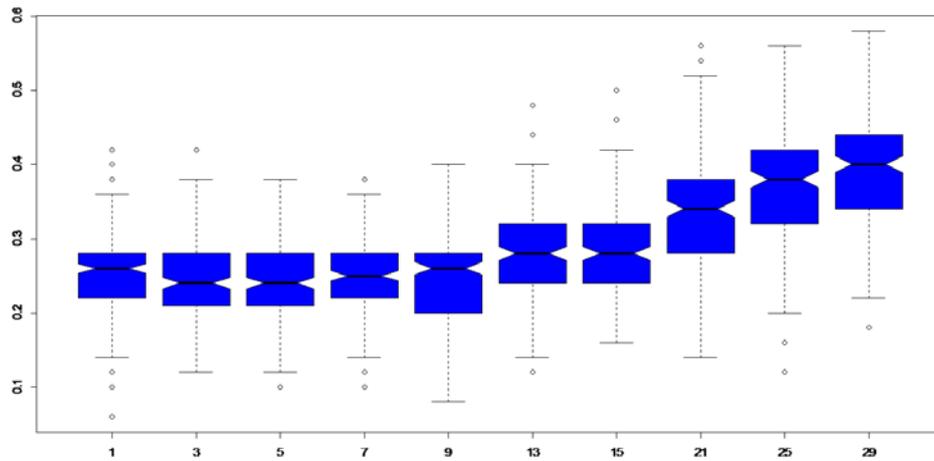


Figura 8.17. Dispersión de errores usando todas las características para el caso R123-R4

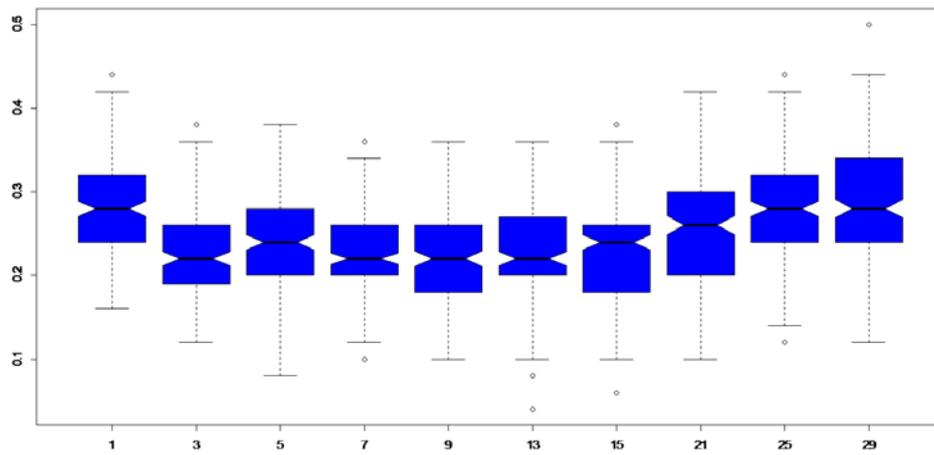


Figura 8.18. Dispersión de errores usando las características del Backward para el caso R123-R4

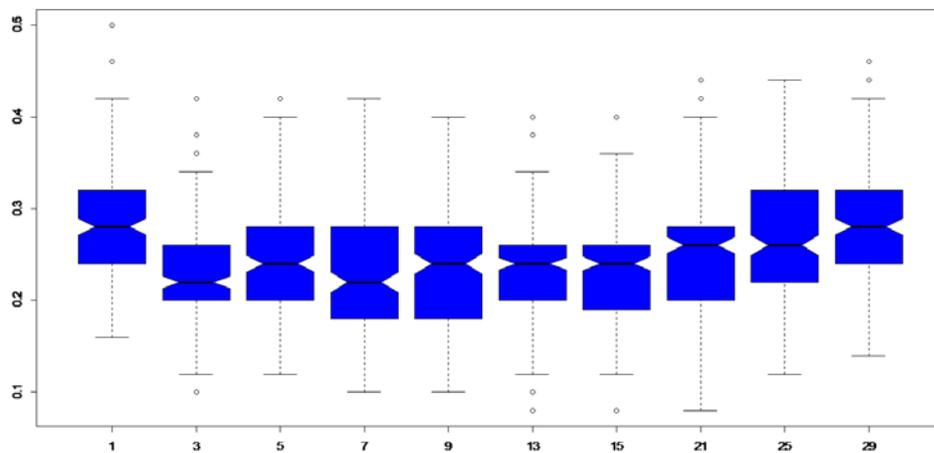


Figura 8.19. Dispersión de errores usando las características del Forward para el caso R123-R4

La tabla 8.9 muestra una comparación entre los mejores errores aparentes y los mejores errores estimados para cada clasificación. Podemos ver como el cambio en el error de clasificación no es excesivo al utilizar la técnica del bootstrap para corregir el sesgo del error aparente. Esto indica que el sesgo inicial del error aparente no es grande. Como este dato no se conoce a priori siempre se deben utilizar las técnicas de corrección del sesgo.

	R12-R34	R1-R234	R123-R4	R1-R2-R3-R4
Errores aparentes	0.1213	0.1004	0.1004	0.205
Errores estimados	0.1224	0.1081	0.1052	0.225

Tabla 8.9. Comparación de los mejores errores.

8.3. SUPPORT VECTOR MACHINE (SVM)

8.3.1. Error estimado SVM

Se utilizará el comando TUNE de R [KAR06] para intentar averiguar cuáles son los parámetros gamma y coste con los que mejor se comporta el algoritmo SVM, estos dos parámetros fueron explicados en el capítulo 5. Para conocer la bondad del ajuste se ha utilizado el error estimado obtenido mediante el método de validación bootstrapping. Se calculará para cada combinación de fallos el error estimado usando las variables obtenidas después de aplicar el método discriminante en el capítulo 7 y para una rejilla de gammas formada por (0.125, 0.25, 0.5, 1, 1.25) y un coste que fue variando entre los siguientes puntos (2, 4, 6, 8, 10).

Clasificación R12-R34

La figura 8.20 muestra la gráfica obtenida con el comando TUNE, en ella se puede apreciar de forma visual como se comporta el algoritmo en términos de error para las diferentes combinaciones de los parámetros gamma y coste, siendo las zonas más oscuras las que mejor resultado obtienen.

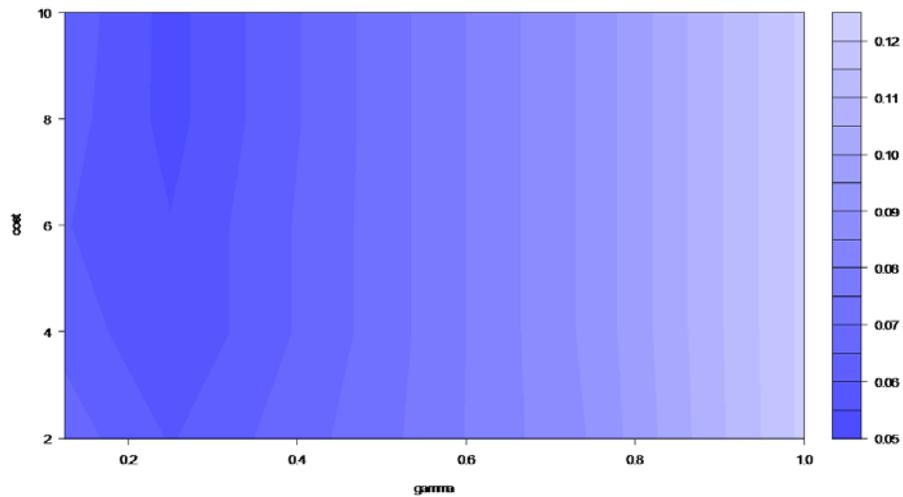


Figura 8.20. Backward R12-R34

En la tabla 8.10 se muestra cual fue el mejor error estimado, la dispersión y cuáles fueron los parámetros con los se obtuvo ese error.

	Gamma	Coste	Error estimado	Dispersión
Backward	0,25	8	0,0532	0,02267

Tabla 8.10. Mejor error estimado para el caso R12-R34.

Clasificación R1-R234

La figura 8.21 y 8.22 muestra la gráfica obtenida con el comando TUNE para la clasificación R1-R234, en ella se puede apreciar de forma visual como se comporta el algoritmo en términos de error para las diferentes combinaciones de los parámetros gamma y coste.

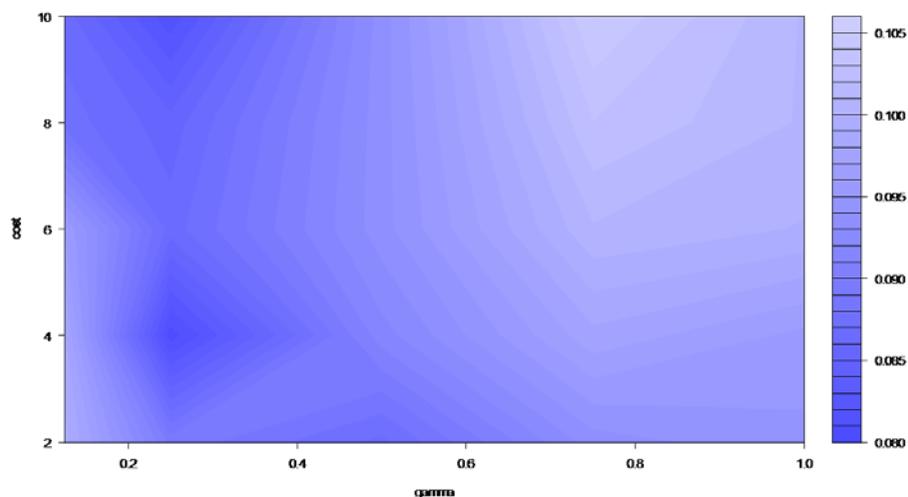


Figura 8.21. Backward R1-R234

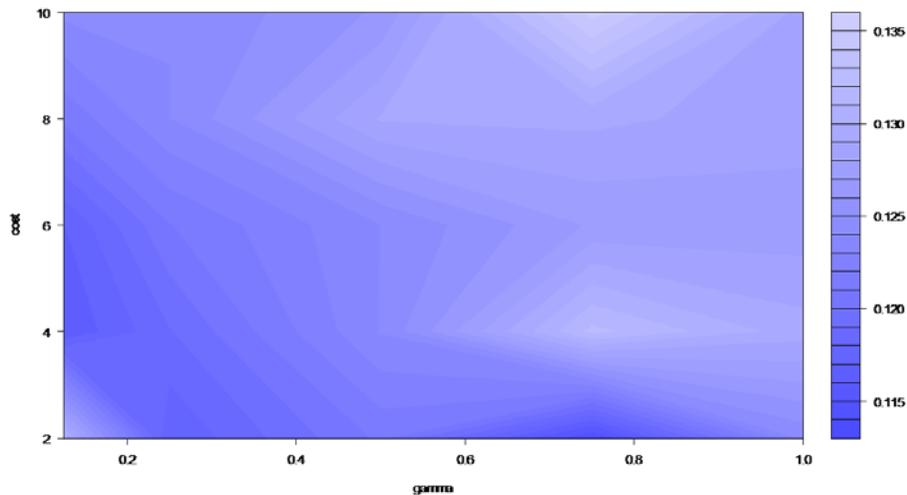


Figura 8.22. Forward R1-R234

En la tabla 8.11 se muestra cual fue el mejor error estimado, la dispersión y cuáles fueron los parámetros con los se obtuvo ese error.

	Gamma	Coste	Error estimado	Dispersión
Backward	0,25	4	0,0805	0,01763
Forward	0,75	2	0,1139	0,04553

Tabla 8.11. Mejor error estimado para el caso R1-R234.

Clasificación R123-R4

La figura 8.23 y 8.24 muestra la gráfica obtenida con el comando TUNE para la clasificación R123-R4, en ella se puede apreciar de forma visual cómo se comporta el algoritmo en términos de error para las diferentes combinaciones de los parámetros gamma y coste.

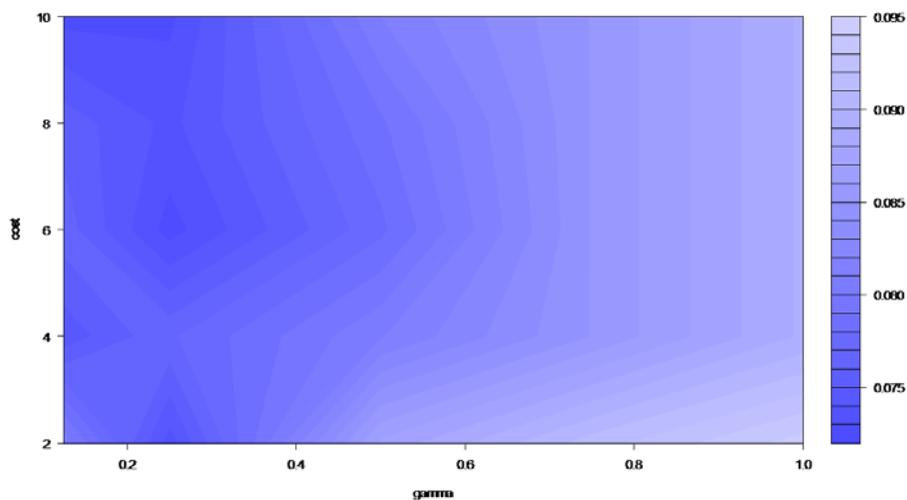


Figura 8.23. Backward R123-4

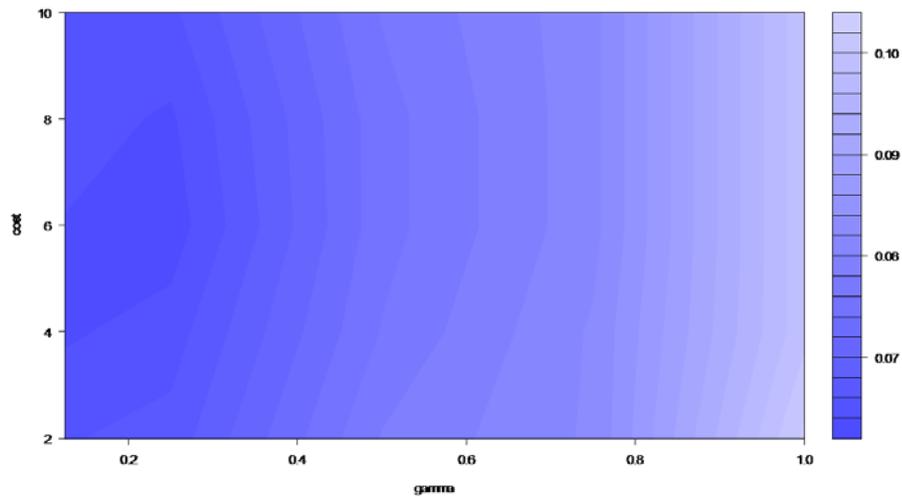


Figura 8.24. Forward R123-4

En la tabla 8.12 se muestra cual fue el mejor error estimado, la dispersión y cuáles fueron los parámetros con los se obtuvo ese error.

	Gamma	Coste	Error estimado	Dispersión
Backward	0,25	6	0,0725	0,0205
Forward	0,25	6	0,0628	0,0212

Tabla 8.12. Mejor error estimado para el caso R123-R4.

Clasificación R1-R2-R3-R4

La figura 8.25 y 8.26 muestra la gráfica obtenida con el comando TUNE para la clasificación R123-R4, en ella se puede apreciar de forma visual cómo se comporta el algoritmo en términos de error para las diferentes combinaciones de los parámetros gamma y coste.

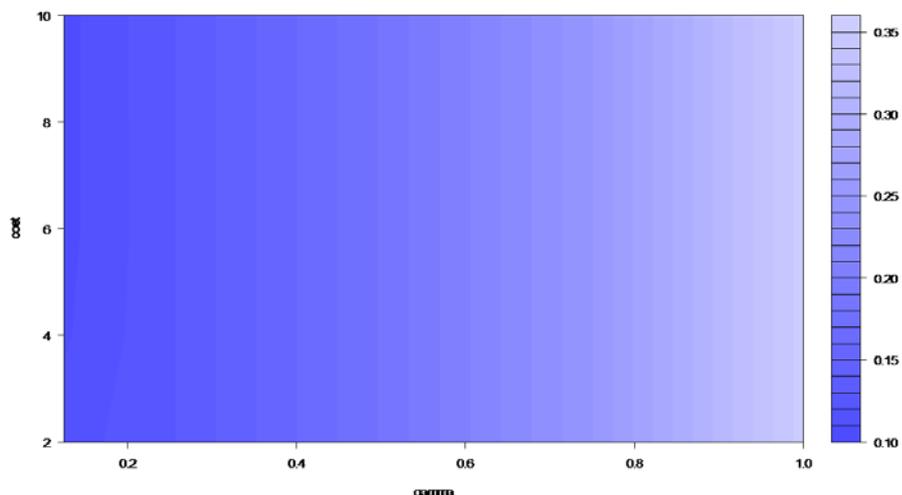


Figura 8.25. Backward R1-R2-R3-R4

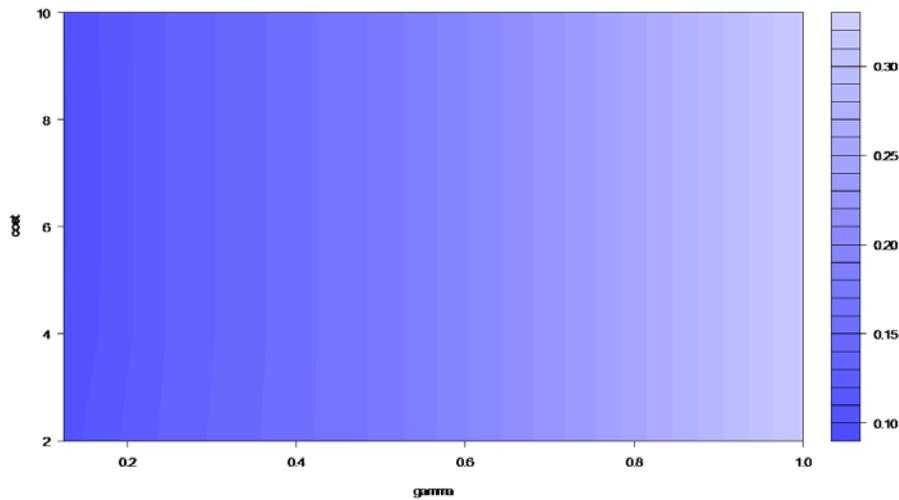


Figura 8.26. Forward R1-R2-R3-R4

En la tabla 8.13 se muestra cual fue el mejor error estimado, la dispersión y cuáles fueron los parámetros con los se obtuvo ese error.

	Gamma	Coste	Error estimado	Dispersión
Backward	0,125	6	0,1065	0,022
Forward	0,125	6	0,0987	0,0259

Tabla 8.13. Mejor error estimado para el caso R1-R2-R3-R4.

8.3.2. Modelo y error aparente SVM

En este apartado se va a utilizar como método de bondad del ajuste el error aparente. Se calculará con el programa R el error aparente para las mismas variables que en el apartado anterior, usando los mejores parámetros de gamma y coste que se obtuvieron.

Clasificación R12-R34

En la tabla 8.14 se muestra el error aparente y el número de vectores de soporte cuando se entrenó al algoritmo con todos los datos.

	Gamma	Coste	Error aparente	Nº vectores soporte
Backward	0,25	8	0	99

Tabla 8.14. Error aparente R12-R34.

Clasificación R1-R234

En la tabla 8.15 se muestra el error aparente y el número de vectores de soporte cuando se entrenó al algoritmo con todos los datos.

	Gamma	Coste	Error aparente	Nº vectores soporte
Backward	0,25	4	0,0209	87
Forward	0,75	2	0,0251	108

Tabla 8.15. Error aparente para el caso R1-R234.

Clasificación R123-R4

En la tabla 8.16 se muestra el error aparente y el número de vectores de soporte cuando se entrenó al algoritmo con todos los datos.

	Gamma	Coste	Error aparente	Nº vectores soporte
Backward	0,25	6	0	69
Forward	0,25	6	0	69

Tabla 8.16. Error aparente para el caso R123-R4.

Clasificación R1-R2-R3-R4

En la tabla 8.17 se muestra el error aparente y el número de vectores de soporte cuando se entrenó al algoritmo con todos los datos.

	Gamma	Coste	Error aparente	Nº vectores soporte
Backward	0,125	6	0	136
Forward	0,125	6	0	136

Tabla 8.17. Error aparente para el caso R1-R2-R3-R4.

La tabla 8.18 es una comparación entre los errores aparentes y los errores estimados de los clasificadores que mejor se han comportado. Se observa claramente que el error aparente es en principio más optimista corrigiendo el error estimado el sesgo de este, pasando de un error del 0% a uno del 9,8% para el caso más extremo.

	R12-R34	R1-R234	R123-R4	R1-R2-R3-R4
Errores estimado	0.0532	0,0805	0,0628	0,0987
Errores aparentes	0	0,0209	0	0

Tabla 8.18. Comparación de errores aparentes y estimados para los mejores casos considerados.

Capítulo 9

Conclusiones

9.1. CONCLUSIONES

En este trabajo se han aplicado dos técnicas estadísticas de reciente aparición a un problema de alto interés práctico en ingeniería como es el de la clasificación de fallos en motores de inducción. La importancia de este problema es obvia debido al elevado número de situaciones en los que estos motores son utilizados y la reducción en los costes que supone conseguir un método adecuado para el diagnóstico precoz de fallos sin necesidad de detener el motor para revisarlo.

Además en este trabajo se compara el comportamiento de dichos métodos de reciente aparición (KNN y SVM) con el método más habitual que se está utilizando en la actualidad, el análisis discriminante lineal. Este método es sencillo e intuitivo y suele dar unos resultados favorables por lo que los nuevos métodos más complejos deben suponer una mejora sustancial si se pretende utilizarlos en la práctica. Este trabajo por tanto da también una respuesta a esta cuestión comparando la capacidad de clasificación de todos los métodos utilizando los procedimientos estadísticos adecuados, basados en remuestreo, que permiten reducir el sesgo que tienen indicadores como el error aparente que erróneamente son a veces utilizados en la práctica.

Con estos métodos se ha intentado clasificar entre diferentes combinaciones de fallos con interés práctico, aquellas en las que se tiene un motor de importancia y se considera que el mínimo fallo debe ser inspeccionado y atendido se ha denotado como R1-R234, aquellas en las que el motor tiene menos relevancia y el interés reside en detectar fallos más avanzados R12-R34, las situaciones en las que solo es relevante un fallo grave en el motor, R123-R4 y las más complejas en las que se quiere saber además de que nivel es el fallo que tiene el motor, diferenciando entre los cuatro tipos posible de fallos R1-R2-R3-R4.

La tabla 9.1 contiene los errores estimados que se han obtenido por remuestreo (bootstrapping) para cada procedimiento de clasificación.

	R12-R34	R1-R234	R123-R4	R1-R2-R3-R4
LDA	0,0837	0,1467	0,1048	0,1687
KNN	0,1291	0,1081	0,1052	0,225
SVM	0,0532	0,0805	0,0628	0,0987

Tabla 9.1. Resumen errores estimados

Podemos comprobar que los k vecinos más próximos no superan excepto en una de las cuatro situaciones consideradas al LDA por lo que no se considera razonable recomendar su uso en este caso práctico. Por otra parte los SVM, que son una técnica mucho más compleja que el LDA sí que obtienen

resultados mejores que este para las cuatro situaciones mencionadas, con reducciones del error de al menos un 54%, llegando hasta el 63% en el mejor de los casos (R12-R34). Sin duda estas son mejoras lo suficientemente importantes como para plantearse el uso de los SVM en la valoración de los fallos de los motores de inducción.

Valladolid, 1 de septiembre de 2014

Fdo.: Gordo Martín, Daniel

Anexos

A.1. INTRODUCCIÓN

En este apartado incluiré los códigos utilizados para la realización del trabajo, solo incluiré una muestra de los mismos ya que varían únicamente en el nombre del archivo del cual se extraían los datos y del vector donde se guardaban los resultados.

En cada archivo de texto se fueron guardando las características apropiadas, por ejemplo, en el archivo "BACKR1R234.txt" se guardaron aquellas características que se obtuvieron en el capítulo 7 mediante la selección backward con el fin de discriminar entre el estado R1 y los demás.

He utilizado los paquetes de R: **class**, para el KNN, **e1071** para el SVM y **MASS** para el discriminante lineal.

A.2. CÓDIGOS

A.2.1. Cálculo del error aparente del KNN en la clasificación R12-R34 a las variables obtenidas tras el Backward.

```
library("class");
train<-read.table("backR12R34.txt",header=TRUE);
cl<-factor(c(rep("r1r2",150),rep("r3r4",89)));
vecino<-c(1,3,5,7,9,13,15,21,25,29);
all<-c(rep(0,length(vecino)));

for (i in 1:length(vecino))
{
  knn<-knn.cv(train, cl,k = vecino[i],l=0, prob = FALSE,use.all = TRUE);
  for (j in 1:length(knn))
  {
    if (knn[j]==cl[j])
    {
      all[i]=all[i]+1;
    }
  }
  all[i]=(all[i]/length(knn))*100;
}
```

A.2.2. Cálculo de los errores estimados del análisis discriminante lineal mediante el método de validación bootstrapping para la clasificación R12-R34

```
#R12-R34 BACKWARD
Library("MASS")
train<-read.table("BACKR12R34.txt",header=TRUE);
cl<-factor(c(rep("r1r2",150),rep("r3r4",89)));
error<-c(rep(0,200));
tasa<-c(0);
boot<-c(0);

for (k in 1:200)
{
  testindex <- sample(1:nrow(train),50,replace=FALSE)
  testset <- train[testindex,]
  trainset <- train[-testindex,]
  grouptest<-cl[testindex]
  grouptrain<-cl[-testindex]
  lda<-lda(trainset,grouptrain)
  res<-predict(lda,testset)

  for (i in 1:length(grouptest))
  {
    if(res[[1]][[i]]==grouptest[i])
    {
      error[k]=error[k]+1;
    }
  }
  error[k]=1-(error[k]/length(grouptest));
}

for (k in 1:length(error))
{
  tasa=tasa+error[k];
}
tasa=tasa/length(error);
for (i in 1:200)
{
  boot=boot+0.632*error[i]+0.368*tasa;
}
boot=boot/200;
```

A.2.4. Dibujar los diagramas gamma-coste

```

train<-read.table("BACKR1R2R3R4.txt",header=TRUE);
cl<-factor(c(rep("r1",110),rep("r2",40),rep("r3",42),rep("r4",47)));
obj1<-tune(svm,train,cl,ranges = list(gamma = c(0.125,0.25,0.5,0.75,1,1.5)
, cost = c(2,4,6,8,10)))
tunecontrol = tune.control(sampling = "boot")
plot(obj1)
summary(obj1)
obj3 <- tune(svm,train, cl,ranges = list(gamma = c(0.125,0.25,0.5,0.75,1,1.5)
, cost = c(2,4,6,8,10)))
plot(obj3)
summary(obj3)

```

A.2.5. Calculo del error estimado del KNN con método de validación bootstrapping y dibujado de los diversos gráficos para la clasificación R12-R34.

```

library(e1071)
library(class)
#Para el fichero donde se recogen todas las variables
train<-read.table("ALL.txt",header=TRUE);
cl<-factor(c(rep("r1r2",150),rep("r3r4",89)));
vecino<-c(1,3,5,7,9,13,15,21,25,29);
error<-matrix(0,nrow=200,ncol=length(cont));
cont<-c(rep(0,length(vecino)));
final<-c(rep(0,length(cont)));
boot<-c(rep(0,length(vecino)));

for (k in 1:200)
{
    cont<-c(rep(0,length(vecino)));
    testindex <- sample(1:nrow(train),50,replace=FALSE)
    testset <- train[testindex,]
    trainset <- train[-testindex,]
    grouptest<-cl[testindex]
    grouptrain<-cl[-testindex]

    for (i in 1:length(vecino))
    {

```

```

        resKNN<-
        knn(trainset,testset,grouptrain,k=vecino[i],prob=FALSE);

        for (j in 1:length(resKNN))
        {
            if (resKNN[j]==groupptest[j])
            {
                cont[i]=cont[i]+1;
            }
        }
        cont[i]=1-(cont[i]/length(resKNN));
        error[k,]=cont;
    }
}

for (k in 1:length(cont))
{
    final[k]<-sum(error[,k])/200;
}
for(k in 1:length(vecino))
{
    for (i in 1:200)
    {
        boot[k]=boot[k]+0.632*error[i,k]+0.368*final[k];
    }
    boot[k]=boot[k]/200;
}

```

#Para el fichero donde se recogen las variables obtenidas del análisis backward.

```

train<-read.table("BACKR12R34.txt",header=TRUE);
error<-matrix(0,nrow=200,ncol=length(cont));
cont<-c(rep(0,length(vecino)));
final<-c(rep(0,length(vecino)));
boot1<-c(rep(0,length(vecino)));

for (k in 1:200)
{
    cont<-c(rep(0,length(vecino)));
    testindex <- sample(1:nrow(train),50,replace=FALSE)
    testset <- train[testindex,]
    trainset <- train[-testindex,]
}

```

```

groupstest<-cl[testindex]
grouptrain<-cl[-testindex]
for (i in 1:length(vecino))
{
  resKNN<-knn(trainset,testset,grouptrain,k=vecino[i],
  prob=FALSE);

  for (j in 1:length(resKNN))
  {
    if (resKNN[j]==groupstest[j])
    {
      cont[i]=cont[i]+1;
    }
  }
  cont[i]=1-(cont[i]/length(resKNN));
  error[k,]=cont;
}

for (k in 1:length(cont))
{
  final[k]<-sum(error[,k])/200;
}
for(k in 1:length(vecino))
{
  for (i in 1:200)
  {
    boot1[k]=boot1[k]+0.632*error[i,k]+0.368*final[k];
  }
  boot1[k]=boot1[k]/200;
}

t(rbind(vecino,boot,boot1))
boxplot(error[,1],error[,2],error[,3],error[,4],error[,5],error[,6],error[,7],error[,8],
+error[,9],error[,10],names=c("1","3","5","7","9","13","15","21","25","29"),
+col="blue",notch=TRUE,horizontal=TRUE)

plot(boot,type="b",col="blue",xlab="vecinos",ylab="erroraparente",
ylim=c(0,0.2))
neighbor=c("1","3","5","7","9","13","15","21","25","29")
axis(1,at=1:10,lab=neighbor)
lines(boot1,type="b", col="red")
legend(1,0.2,c("all","back"),col=c("blue","red"),lty=c(1,1),pch=21)

```


Bibliografía

BIBLIOGRAFÍA

- [ACH07] Achmad, W., y Bo-Suk Y. (2007). "Support vector machine in machine condition monitoring and fault diagnosis". *Mechanical Systems and Signal Processing*, 21(2007),2560–2574.
- [DUQ05] Duque, O., y Pérez, M. (2005). "Motores de inducción-Técnicas de mantenimiento predictivo". Madrid-Valladolid: @becedario.
- [EST14] Esteban, P.(2014). "Aportaciones al mantenimiento predictivo de motores de inducción mediante una metodología de diagnóstico basada en el uso combinado de técnicas estadísticas y Redes Neuronales Artificiales", tesis doctoral, Escuela de ingenierías industriales de Valladolid.
- [FIL09] Filzmoser, P., y Varmuza, K. (2009). "Introduction to multivariate analysis in chemometrics". Nueva York: CRC Press.
- [FRA08] Fraile, J.(2008). "Máquinas eléctricas 6ª ed". Madrid: McGraw-Hill.
- [HER07] Hernández, J., Ramírez, M., y Ferri, C.(2007). "Introducción a la minería de datos". Madrid: Pearson.
- [JAR09] Jaroslaw, K., y Stanislaw, O.(2009). "Support vector machine for fault diagnosis of the broken rotor bars of squirrel-cage induction motor". *Neuronal Comput & Applic*, 557-564.
- [KAR06] Karatzoglou, A;et al.(2006). "Support vector machines in R". *Journal of statistics software*, 5(9),1-5.
- [ONDO6] Onder, O., Boutleux, E., y Clerc, G. (2006). "A Method to Detect Broken Bars in Induction Machine Using Pattern Recognition Techniques". *IEEE Transactions on industry applications*,42(4),916-923.
- [PAJ10] Pajares, G. (2010). "Aprendizaje automatico". Madrid: Ra-Ma.
- [PAR10] Paradis, E. "R para principiantes". Universidad Montpellier.
- [PAY02] Payne, B., Ball, A., y Fenshou, G. (2002). "Detection and diagnosis of induction motor faults using statistical measures". *International Journal of COMADEM*, 5(2), 5-19.

- [PEÑ02] Peña, D. (2002). “Análisis de datos multivariante”. Madrid: McGrawHill.
- [PER13] Pérez, C. (2013). “Análisis multivariante de datos. Aplicaciones con IBM SPSS, SAS y StatGraphics”. Madrid: Garceta.
- [ROD13] Rodríguez, B. (2013). “Detección de fallos en motores de inducción utilizando histogramas”, proyecto fin de carrera, Escuela de Ingenierías Industriales de Valladolid.
- [SIE06] Sierra, B. (2006). “Aprendizaje automático: conceptos básicos y avanzados”. Madrid: Pearson Prentice Hall.

Páginas web

- [a] <http://cran.r-project.org/>
- [b] <http://www.coursera.org/>

