

UNIVERSITAT POLITÈCNICA DE CATALUNYA

**APRENDIZAJE CON MÁQUINAS
NÚCLEO EN ENTORNOS DE
MULTICLASIFICACIÓN**

Por

Cecilio Angulo Bahón

TESIS DOCTORAL

Presentada en el Departament d'Enginyeria de Sistemes, Automàtica i Informàtica
Industrial

de la Universitat Politècnica de Catalunya

de acuerdo con los requerimientos para acceder al
grado de Doctor en Ciencias. Especialidad Matemáticas

Vilanova i la Geltrú, Abril 2001

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Tesis Doctoral presentada de acuerdo con los
requerimientos para acceder al grado de
Doctor en Ciencias. Especialidad Matemáticas

APRENDIZAJE CON MÁQUINAS NÚCLEO EN ENTORNOS DE MULTICLASIFICACIÓN

CECILIO ANGULO BAHÓN

Dirigida por:

ANDREU CATALÀ MALLOFRÉ
Profesor Titular de Universidad
Director de la EUPVG
Dept. ESAII, U.P.C., Vilanova i la Geltrú

Aprobada por:

JOSEPH AGUILAR-MARTIN
Director de Investigación
Laboratoire d'Analyse et
d'Architecture de Systèmes.
Centre National de la Recherche
Scientifique, Toulouse

ANTONIO ARTÉS RODRÍGUEZ
Profesor Titular de Universidad
Dept. Tecnologías de las
Comunicaciones.
Universidad Carlos III de Madrid

APRENDIZAJE CON MÁQUINAS NÚCLEO EN ENTORNOS DE MULTICLASIFICACIÓN

Abril 2001

Resumen

La propiedad de generalización de una máquina de aprendizaje, es decir su capacidad para emitir una respuesta correcta ante una nueva entrada semejante a aquellas con las que ha sido entrenada, es la característica principal que se busca en los sistemas conexionistas supervisados y sirve de justificación en la elección de los principios inductivos y el tipo de estructuras de aprendizaje para elaborar el presente estudio.

La regularización o penalización es uno de estos principios que favorecen a nivel teórico la generalización, sobre el cual se ha desarrollado un método de cálculo directo de la matriz de regularización cuando se utiliza como estabilizador un operador diferencial de segundo grado, precisamente aquel que minimiza el grado de convexidad de la función solución, evitando así el proceso iterativo de cálculo de la matriz hessiana y fijando el tipo de núcleo a ser utilizado.

Los nexos de unión entre la regularización y el principio de minimización del riesgo estructural así como las excelentes características teóricas mostradas por este último principio trabajando, por definición, sobre conjuntos finitos de datos y expandiendo su solución sobre un número pequeño de núcleos, han llevado a desplazar el foco de trabajo de numerosos investigadores hacia las máquinas de soporte vectorial, su materialización procedimental. En este contexto, se ha desarrollado una máquina que permite extender de forma natural el comportamiento binario de estas máquinas núcleo de margen máximo sobre problemas de clasificación hacia una solución ternaria más acorde con la estructura geométrica de los datos, en especial en las situaciones habituales de espacios de salida que poseen más de dos clases. El uso de la nueva arquitectura, bautizada *K-SVCR*, en problemas de multclasificación resulta más adecuado que las reducciones estándares de problemas multiclase sobre máquinas biclasificadoras en estructuras en paralelo o arbóreas puesto que cada nodo de

dicotomía considera todo el espacio de entrenamiento y se fuerza al hiperplano de separación a considerar la estructura geométrica de los patrones de entrenamiento. En especial, se demuestra la robustez del nuevo método ante fallos en las predicciones de algunos de sus nodos de trabajo cuando se considera un tipo especial de combinación de estas respuestas.

La nueva arquitectura de multclasificación ha sido modificada con posterioridad para ser implementada sobre un problema de clasificación con características independientes, la ordenación o problema de aprendizaje de preferencias. Sus prestaciones son evaluadas sobre una aplicación financiera en la determinación de riesgos crediticios. Finalmente, una aplicación de categorización o discriminación de escenarios de depuración donde incide el efecto de la temporalidad sirve también como ejemplo de funcionamiento.

Agradecimientos

Quisiera agradecer a los investigadores del Grup de Recerca en Enginyeria del Coneixement (GREC) el excelente ambiente de trabajo y la ayuda prestada desde mis primeros cursos de doctorado hasta la materialización final de esta investigación; a Quico Prats y Mónica Sánchez de MA2-UPC por su visión matemática de los problemas tan cercana a la mía; Núria Agell y el equipo de ESADE-URL y Juan Carlos Aguado de ESAII-UPC por mostrarme nuevos horizontes de investigación y sus conocimientos sobre aritmética intervalar; Haydemar Núñez por las fructíferas reuniones mantenidas para extraer conocimiento a partir de la aplicación de SVMs; Joseph Aguilar de LAAS-CNRS que muestra siempre interés en aquello que signifique descubrir en los métodos de aprendizaje y sabe aplicar su mirada crítica y siempre acertada, pero sobre todo por saber llevar adelante un proyecto como el Laboratorio Europeo Asociado (LEA-SICA) que facilita a los jóvenes investigadores un marco de intercambio científico único; al grupo de GREC en ESAII-UPC de Vilanova i la Geltrú, Andreu Català, Xavier Parra, Pere Ponsa y Joan Aranda, con los que he mantenido provechosas discusiones y se ha establecido unos lazos de especial amistad más allá del intercambio investigador. Un agradecimiento muy especial a Xavier Parra, quien me aportó esa primera visión de las redes neuronales desde una perspectiva externa a los perceptrones multicapa, cuyos comentarios e impresiones sobre la validación de resultados empíricos me han sido de una inestimable ayuda y cuya colaboración y actuación junto a mi director de tesis, Andreu Català, ha servido para crear un grupo de trabajo dentro de un ambiente de colaboración magnífico. Del laboratorio LEA-SICA, agradecer a Monique Polit de LP2A-Université de Perpignan su acogida durante el periodo que desarrollé mi investigación en su laboratorio dentro del marco de referencia de la lógica borrosa. Gracias también a Joan Colomer y el grupo de trabajo del Institut d'Informàtica i Aplicacions de la UdG y a Teresa Escobet de ESAII-UPC por su inestimable colaboración en el estudio y generación de datos de los procesos de depuración.

A la dirección y a mis ex-compañeros del IES Terra Roja de Santa Coloma de Gramenet agradecerles las facilidades que me ofrecieron para que pudiera desarrollar mi labor de investigación complementaria a mi labor educativa, de igual manera que a mis colegas de ESAII-UPC en Manresa, Carles Torrens, Teresa Escobet, Toni Escobet y Jordi Solà, agradecerles

su calidez humana en el trabajo, en especial a Carles, que junto a Jaume Figueras en Terrassa han despertado en mí la conciencia del desarrollo técnico de aquello que la base de mi formación matemática permite teorizar.

Agradecer a los profesores de ESAII-UPC en Terrassa y a la Administración del departamento su disposición en todas aquellas formalidades ajenas a la investigación pero cuya resolución, no siempre evidente, hacen más sencillo el día a día, en especial, agradecer a Joan Vert y a Josep Contreras su eficiencia y amabilidad.

Un obligado agradecimiento a mis ex-camaradas de facultad pero aún compañeros de viaje, con los cuales he podido compartir, pues muchos lo saben de primera mano, las inquietudes y estímulos de realizar una tesis doctoral. En especial, quisiera recordar a Mario, que se quedó en el camino, porque Dios es egoísta y se lleva siempre lo mejor.

El agradecimiento a mis padres, Modesto y Eladia, y a mis hermanos, Jordi y Montse, su cariño y su paciencia de todos estos años, sin su ayuda y reconocimiento no hubiera sido de ninguna forma posible este trabajo. Finalmente, un muchas gracias enorme a mi mujer, Mari, a quien he tenido a mi lado desde el inicio de esta singladura y a mi hija Maria, su sonrisa cuando vuelvo de noche a casa es el mejor pago que pueda recibir al trabajo realizado.

Lista de Abreviaciones

1-v-1 SVMC	Nodo de dicotomía uno contra el resto usando SVMC
1-v-r SVMC	Nodo de dicotomía uno contra uno usando SVMC
1,1-v-r SVMC	Nodo de dicotomía usando K -SVCR
ANN	Red neuronal artificial
BPM	Máquina del punto de Bayes
DAGSVM	Grafo Acíclico Dirigido con nodos SVMC
DDAG	Grafo Acíclico Dirigido de Decisión
ECOC	Técnica de códigos de salida de corrección de error
ECOC K -SVCR	Nodo de dicotomía K -SVCR con tratamiento ECOC de las salidas
ERM	Principio inductivo de minimización del riesgo empírico
\mathcal{F}	Espacio de características
k -RLP	Problema de programación lineal robusta para multclasificación
K -SVCR	Máquina de soporte vectorial trclasificadora orientada a multclasificación
KSVMC	Máquina de soporte vectorial para multclasificación considerando todas las clases a la vez
LM	Máquina general de aprendizaje
\mathcal{LM}	Espacio de aproximación generado por la arquitectura de las LMs
M-SVM	Máquina de aprendizaje para multclasificación híbrida de k -RLP con SVM
MLP	Máquina de aprendizaje Perceptrón multicapa
MSE	Error mínimo-cuadrático
PGAE	Problema general de aprendizaje a partir de ejemplos
QP	Problema de programación cuadrática
$R(\omega)$	Funcional de riesgo teórico
$R_{emp}(\omega)$	Funcional de riesgo empírico
$R_{emp}^{L_2}(\omega)$	Funcional de riesgo empírico estándar
$R_{emp}^{reg}(\omega)$	Funcional de riesgo empírico regularizado

$R^{pref}(\omega)$	Funcional de riesgo ordinal
$R_{emp}^{pref}(\omega)$	Funcional de riesgo empírico ordinal
$R_{K-SVCR}^{pref}(\omega, \xi, \varphi^{(*)})$	Funcional de riesgo ordinal asociado a una $K - SVCR$
$R^{reg}(\omega)$	Funcional de riesgo regularizado
$R_{srm}(\omega)$	Funcional de riesgo estructural
$R_{K-SVCR}(\omega, \xi, \varphi^{(*)})$	Funcional de riesgo general asociado a una $K - SVCR$
$R_{KSVMC}(\omega, \xi)$	Funcional de riesgo asociado a una $KSVMC$
$R_{M-SVM}(\omega, \xi)$	Funcional de riesgo asociado a una $KSVMC$ combinación de RLP con $SVMC$
$R_{SVMC}(\omega, \xi)$	Funcional de riesgo asociado a una $SVMC$
$R_{SVMR}(\omega, \varphi^{(*)})$	Funcional de riesgo asociado a una $SVMR$
$R_{Unif}(\omega, \xi)$	Funcional de riesgo asociado a una $KSVMC$ resultado de convergencia uniforme
RBF	Función de base radial
RBFNN	Red neuronal artificial de función de base radial
RKHS	Espacio de Hilbert generado por núcleo
RLP	Problema de programación lineal robusta
RN	Red de regularización
RVM	Máquina de vectores relevantes
SLT	Teoría del aprendizaje estadístico
SRM	Principio inductivo de minimización del riesgo estructural
SVDT	Árbol de decisión basado en vectores soporte
SVM	Máquina de soporte vectorial
SVMC	Máquina de soporte vectorial para clasificación
SVMR	Máquina de soporte vectorial para regresión
\mathcal{T}	Conjunto de entrenamiento
\mathcal{X}	Espacio de entrada
\mathcal{Y}	Espacio de salida

Índice General

I	Algoritmos	15
1	Introducción	17
1.1	Problema de Aprendizaje a partir de Ejemplos	18
1.1.1	Planteamiento del Problema General	18
1.1.2	Casos Particulares del Problema General	20
1.2	Algoritmo de Aprendizaje	23
1.2.1	Definición de Algoritmo de Aprendizaje	23
1.2.2	Principios Inductivos	25
1.2.3	Métodos de Aprendizaje	32
1.3	En Resumen	33
2	Cálculo de Matrices de Regularización	35
2.1	Introducción	36
2.2	Regularización	38
2.2.1	Regularización para la Estabilidad Numérica	39
2.2.2	Regularización de Tikhonov para la Unicidad	40

2.3	RBFFN Regularizadas	40
2.3.1	Regularización de Orden Cero desde una Visión Bayesiana (Ridge Regression)	42
2.3.2	Regularización de Segundo Orden desde una Visión Bayesiana (Regresión no lineal)	42
2.4	Problema Variacional de la Regularización	43
2.5	Similaridades entre las Visiones Bayesiana y de Tikhonov	44
2.5.1	Equivalencia entre las Visiones Bayesiana y Variacional para la Regularización de Orden Cero	46
2.5.2	Equivalencia entre las Visiones Bayesiana y Variacional para la Regularización de Segundo Orden	46
2.5.3	Simplicidad Computacional del Nuevo Método	47
2.6	En Resumen	48
3	Redes de Regularización y Máquinas de Soporte Vectorial	51
3.1	Máquinas de Soporte Vectorial	52
3.1.1	SVM para Clasificación	52
3.1.2	SVM para Regresión	54
3.2	Hacia un Marco Unificado para SVMs y RNs	55
3.2.1	Conexión vía Operadores de Regularización	56
3.2.2	Espacios de Hilbert definidos por Núcleos	57
3.2.3	Conexión vía el Principio Inductivo SRM	58
3.2.4	Implicaciones Bayesianas	59
3.3	En Resumen	60

4	Clasificación Multiclase con Máquinas de Soporte Vectorial	61
4.1	Definiendo la Clasificación Multiclase	62
4.2	Arquitectura de Descomposición y Reconstrucción	64
4.2.1	Arquitecturas Estándares	64
4.2.2	Otras arquitecturas en paralelo	69
4.2.3	Arquitecturas en árbol	73
4.3	Esquemas “Todas las Clases a la Vez”	75
4.3.1	Generalización de una SVMC	75
4.3.2	K-SVMC sobre Resultado de Convergencia Uniforme	76
4.3.3	Combinando RLP con SVMC	78
4.4	Estudio de la Robustez	79
4.5	En Resumen	81
5	Máquina de Aprendizaje K-SVCR	85
5.1	Motivación e Interpretación	85
5.1.1	Idea Motivadora	86
5.2	Máquina de Aprendizaje K -SVCR	88
5.2.1	Caso No Lineal Separable	88
5.2.2	Caso No Lineal No Separable	92
5.3	Experimentación	93
5.3.1	Problemas Artificiales	94
5.3.2	Factor de insensitividad δ	94

5.3.3	Comparativa con la SVMR	95
5.3.4	Problemas Artificiales No Linealmente Separables	99
5.4	En Resumen	102
6	Máquinas de Aprendizaje K-SVCR en Multclasificación	107
6.1	Esquema de Descomposición para Multclasificación	108
6.2	Método de Reconstrucción	111
6.2.1	Interpretación de las Predicciones	111
6.2.2	Elemento de Combinación y Análisis de Robustez	113
6.3	Comparativa con la Metodología ECOC	115
6.3.1	Robustez de la multclasificación {Biclase, ECOC}	117
6.3.2	Máquina ECOC K -SVCR	120
6.3.3	Comparativa ECOC K -SVCR <i>versus</i> {Biclase, ECOC}	121
6.4	Experimentación	125
6.5	En Resumen	127
II	Aplicaciones	129
7	K-SVCR Regresión Ordinal. Aplicación a un Problema Financiero	131
7.1	Introducción	131
7.2	Formulación de Regresión Ordinal	132
7.3	K -SVCR para Regresión Ordinal	135
7.4	Problema Financiero de Riesgo Crediticio	141

7.4.1	Uso de máquinas K -SVCR en multclasificación.	143
7.4.2	Uso de máquinas K -SVCR en regresión ordinal.	147
7.4.3	Resultados sobre test	151
7.5	En Resumen	153
8	Identificación de Escenarios en un Proceso de Depuración	155
8.1	Introducción	155
8.1.1	La EDAR y el Simulador	156
8.1.2	Los Escenarios	157
8.2	Uso de Máquinas K -SVCR	158
8.3	En Resumen	163
	Conclusiones	165
	Bibliografía	170

Índice de Figuras

1.1	Modelo general de un proceso de aprendizaje a partir de ejemplos. . .	19
1.2	Procesos de inferencia: inducción-deducción y transducción.	25
1.3	Visualización del error de aproximación, de estimación y de generalización.	27
1.4	Visualización del error de aproximación, de estimación y de generalización.	31
3.1	El margen es la distancia perpendicular entre el hiperplano separador y el hiperplano que pasa sobre los puntos más cercanos, los vectores soporte.	53
4.1	Modelo general de una arquitectura de aprendizaje en multclasificación siguiendo un esquema de descomposición en paralelo.	63
5.1	Conjuntos de entrenamiento linealmente separables.	94
5.2	Conjuntos de entrenamiento no linealmente separables.	95
5.3	Resultados para diferentes niveles de insensitividad del entrenamiento sobre el conjunto \mathcal{T}_1 . Las cantidades que acompañan cada subfigura representan el nivel de insensitividad, el tiempo de entrenamiento y el número de vectores soporte utilizados.	96
5.4	Resultados del entrenamiento sobre el conjunto \mathcal{T}_2 utilizando funciones núcleo polinomiales de grado $n = 3$	97

5.5	Resultados del entrenamiento sobre el conjunto \mathcal{T}_2 utilizando funciones núcleo gaussianas de varianza 0.5.	98
5.6	Resultados del entrenamiento sobre el conjunto \mathcal{T}_3 utilizando funciones núcleo polinomiales de grado $n = 3$	99
5.7	Resultados del entrenamiento sobre el conjunto \mathcal{T}_5 utilizando núcleos gaussianos para diferentes elecciones de etiqueta de clase.	101
5.8	Resultados del entrenamiento sobre el conjunto \mathcal{T}_5 utilizando núcleos polinomiales para diferentes elecciones de etiqueta de clase.	103
5.9	Resultados del entrenamiento sobre el conjunto \mathcal{T}_6 utilizando núcleos gaussianos para diferentes elecciones de etiqueta de clase.	104
6.1	Relación entre el número de clases implicadas en la clasificación y el porcentaje de nodos de dicotomía que admiten error en su clasificación sin que el resultado final correcto se vea alterado utilizando máquinas K -SVCR.	116
6.2	Factor de robustez de las diferentes arquitecturas empleadas en la comparativa en función del número de clases K	122
7.1	Cumplimiento de las restricciones del QP problema por parte de los patrones diferencia $\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}$	140
7.2	Ordenación de las entradas \mathbf{x} dentro de la recta real mediante la función utilidad $U(\mathbf{x})$	141
7.3	Ordenaciones generadas para los patrones de los diferentes sectores utilizando núcleos gaussianos con diferentes valores del parámetro σ . Cada columna corresponde a las salidas obtenidas para los patrones de cada rango.	150
7.4	Resultados de la ordenación sobre dos sectores de producción utilizando una arquitectura de ordenación con máquinas K -SVCR.	152
8.1	Gráficas con datos originales utilizados para entrenar la arquitectura de multclasificación con máquinas K -SVCR.	160

- 8.2 Gráficas con datos en forma incremental utilizados para entrenar la arquitectura de multclasificación con máquinas K -SVCR. 161
- 8.3 Resultados sobre diferentes escenarios de depuración de una EDAR utilizando una arquitectura de multclasificación con máquinas K -SVCR.162

Índice de Tablas

1.1	Sumario de características de los principios inductivos descritos. . . .	34
4.1	Comparativa sobre complejidad de arquitecturas multiclase.	83
4.2	Comparativa sobre características de arquitecturas multiclase.	83
5.1	Resultados sobre el conjunto de entrenamiento \mathcal{T}_1	100
5.2	Resultados sobre el conjunto de entrenamiento \mathcal{T}_5 utilizando núcleos gaussianos.	100
5.3	Resultados sobre el conjunto de entrenamiento \mathcal{T}_5 utilizando núcleos polinomiales.	102
5.4	Resultados sobre el conjunto de entrenamiento \mathcal{T}_6 utilizando núcleos gaussianos.	102
6.1	Comparativa sobre la complejidad del nuevo método de descomposición respecto a las descomposiciones estándares.	109
6.2	Comparativa sobre la complejidad del nuevo método de descomposición respecto a las descomposiciones basadas en la teoría ECOC.	109
6.3	Número de nodos de dicotomía y factor de robustez para el caso $3 \leq K \leq 7$	118
6.4	Número de nodos de dicotomía y factor de robustez para el caso $8 \leq K \leq 11$	119

6.5	Número de nodos de dicotomía y factor de robustez para el caso $K = 12$	120
6.6	Relación de resultados sobre igual número de nodos de dicotomía, L	123
6.7	Relación de resultados sobre igual distancia entre códigos, d	123
6.8	Prestaciones en porcentaje de error sobre el conjunto de test de la nueva máquina de aprendizaje respecto a combinaciones estándares de máquinas SVM y máquinas SVM multiclase.	126
7.1	Datos del problema de riesgo crediticio. Las columnas representan el rango asignado y las filas resumen el número de patrones para los diferentes sectores.	143
7.2	Resultados sobre los conjuntos de test separados por sectores utilizando máquinas K -SVCR en multclasificación con núcleos gaussianos.	145
7.3	Resultados sobre los conjuntos de test separados por sectores utilizando máquinas K -SVCR en multclasificación con núcleos polinomiales.	147
7.4	Resultados sobre los conjuntos de test separados por sectores utilizando máquinas K -SVCR en regresión ordinal con núcleos gaussianos.	149
7.5	Porcentaje de error cometido y número de vectores soporte necesarios sobre el conjunto de patrones total en conjuntos de test del 25% de patrones de entrenamiento.	152

Parte I

Algoritmos

