

Capítulo 7

***K*-SVCR Regresión Ordinal. Aplicación a un Problema Financiero**

“Sometimes, the world of credit risk management seems to have more in common with religion than with science”

Dr. Risk.

Para comenzar

Tal como se apuntó en el Capítulo 1, el problema de regresión ordinal u ordenación posee características mixtas de clasificación y regresión por lo que constituye una clase de problemas a ser considerado de forma independiente. Sus características propias deberían ser tenidas en cuenta de forma explícita por las máquinas de aprendizaje que se enfrentan a él mediante la modificación cuanto menos de la función de coste a ser considerada. En el presente Capítulo el problema de ordenación es formulado y analizado desde el punto de vista de la nueva máquina *K*-SVCR. En este estudio se compara su definición con otros resultados existentes en la literatura. Finalmente la eficiencia de la nueva aproximación se muestra sobre un problema de riesgo crediticio extraído del mundo financiero.

7.1 Introducción

Los problemas de regresión ordinal surgen en numerosas ocasiones en actividades de la vida cotidiana: las tareas diarias se agrupan o clasifican siguiendo un cierto orden

de importancia; los resultados de una búsqueda de información en páginas web sería conveniente que se mostrasen agrupados en subcategorías de preferencias del usuario; las agrupaciones de clientes o proveedores se realizan en numerosas ocasiones atendiendo a un cierto orden de satisfacción y/o expectativas de resultado; una medida continua puede ser dividida siguiendo algún criterio intervalar. Por tanto, si bien se está trabajando con clases, un número finito de elementos, éstos poseen una cierta ordenación atendiendo a algún criterio de preferencia.

Releyendo el enunciado de *problema de regresión ordinal o de ordenación* como aquel definido en la página 22 del Capítulo 1 se aprecia que su especial caracterización se halla fundamentalmente en el espacio de salida, que está constituido por un número finito de clases que poseen una cierta ordenación, $y \in \mathcal{Y} = \{\theta_1, \dots, \theta_K\}$ con $\theta_K \succ_{\mathcal{Y}} \theta_{K-1} \succ_{\mathcal{Y}} \dots \succ_{\mathcal{Y}} \theta_1$. Usualmente se ha considerado que, o bien \mathcal{Y} es un conjunto finito no ordenado por lo que el problema a solucionar es un *problema de clasificación*, y la 0 – 1 función de coste resulta adecuada, o bien \mathcal{Y} es un espacio métrico, por ejemplo $\mathcal{Y} \subseteq \mathbb{R}$, por lo que se estaría considerando el *problema de estimación de una regresión*, cosa que obliga a considerar una función de coste que tenga en cuenta la estructura métrica, por ejemplo la norma 2 dentro del espacio de los números reales. En un problema de regresión ordinal el algoritmo de aprendizaje considera una ordenación, igualmente que en el caso de la regresión métrica, pero sobre el conjunto \mathcal{Y} finito, como en el caso de la clasificación. La existencia de esta escala ordinal sobre el conjunto de clases que permite realizar una comparación entre elementos de entrada en función de este orden provoca la necesidad de definir una función de coste que resulte más acertada para esta tipología.

El tratamiento de esta clase de problemas de regresión ordinal será realizado a partir de una modificación del algoritmo original de las máquinas de aprendizaje K-SVCR. Esta nueva técnica utiliza la máquina trclasificadora para establecer una ordenación ± 1 entre pares de elementos que pertenecen a clases diferentes, mientras que asigna etiqueta nula cuando los elementos comparados se hallan en una misma clase. En comparación con trabajos relacionados como [Herbrich et al., 1999b] y [Phillips, 1999], el algoritmo propuesto permite trabajar sobre un número menor de patrones de aprendizaje y realizar comparaciones tanto en sentido ascendente, $+1$, como descendente, -1 .

7.2 Formulación de Regresión Ordinal

Con la intención de distinguir la formulación del problema de ordenación de aquellas realizadas para el caso de clasificación y de regresión, se denominará *rango* a cada uno de los elementos del espacio de salida \mathcal{Y} , para así significar que existe una relación de orden entre estos componentes que permite hacer comparaciones.

Conjetura 7.1. *Sea*

$$\mathcal{T} = \{(\mathbf{x}_p, y_p)\}_{p=1}^{\ell} \subset \mathcal{X} \times \mathcal{Y} = \mathcal{X} \times \{r_1, \dots, r_K\} \sim P_{\mathcal{X}\mathcal{Y}}^{\ell}, \quad (7.1)$$

el i.i.d. conjunto de entrenamiento de una máquina de aprendizaje para un problema de ordenación, siendo ésta definida sobre los rangos del espacio salida como

$$r_K \succ_{\mathcal{Y}} r_{K-1} \succ_{\mathcal{Y}} \dots \succ_{\mathcal{Y}} r_1, \quad (7.2)$$

donde el símbolo $\succ_{\mathcal{Y}}$ podría traducirse por “es preferido a”.

Se considerará que cada función f del espacio de aproximación \mathcal{LM} induce un orden $\succ_{\mathcal{X}}$ sobre los elementos del espacio de entrada siguiendo la regla

$$\mathbf{x}_i \succ_{\mathcal{X}} \mathbf{x}_j \Leftrightarrow f(\mathbf{x}_i, \omega) \succ_{\mathcal{Y}} f(\mathbf{x}_j, \omega). \quad (7.3)$$

Siguiendo las notaciones, un método de aprendizaje para regresión ordinal será aquel capaz de hallar la función $f(\mathbf{x}, \omega) \in \mathcal{LM}$ induciendo una ordenación sobre el espacio de entrada \mathcal{X} que cometa el mínimo número de inversiones de orden sobre los pares de objetos $(\mathbf{x}_1, \mathbf{x}_2)$ ([Sobel, 1990]).

Definición 7.2. *El cálculo de probabilidad de realizar una inversión de orden se define a partir del funcional de riesgo ordinal como la esperanza estadística*

$$R^{pref}(\omega) = \mathbf{E}[L_{pref}(f(\mathbf{x}_1, \omega), f(\mathbf{x}_2, \omega), y_1, y_2)], \quad (7.4)$$

utilizando como medida una función de coste ordinal definida por

$$L_{pref}(\hat{y}_1, \hat{y}_2, y_1, y_2) = \begin{cases} 1 & \text{si } y_1 \succ_{\mathcal{Y}} y_2 \wedge \neg(\hat{y}_1 \succ_{\mathcal{Y}} \hat{y}_2) \\ 1 & \text{si } y_2 \succ_{\mathcal{Y}} y_1 \wedge \neg(\hat{y}_2 \succ_{\mathcal{Y}} \hat{y}_1) \\ 0 & \text{en otro caso} \end{cases}. \quad (7.5)$$

Aplicando el principio ERM, la función hallada será aquella que hace mínimo el funcional de riesgo empírico ordinal

$$R_{emp}^{pref}(\omega) = \frac{1}{\ell^2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} L_{pref}(f(\mathbf{x}_i, \omega), f(\mathbf{x}_j, \omega), y_i, y_j), \quad (7.6)$$

utilizado como una aproximación del funcional de riesgo desconocido (7.4).

Como puede observarse, las definiciones de los funcionales de riesgo ordinales están basadas en pares de objetos haciendo plausible la característica específica de comparación provocada por el orden en los rangos de salida. Utilizando las abreviaciones $\mathbf{x}^{(1)}$ y $\mathbf{x}^{(2)}$ para denotar el primer y el segundo objeto del par, puede definirse,

Definición 7.3. Sea \mathcal{T} el conjunto de entrenamiento original (7.1) para una máquina de aprendizaje en un problema de regresión ordinal. Si se establece la operación diferencia de rango, notada como \ominus , de la forma

$$y_1 \ominus y_2 = j - k \quad \text{si } y_1 = r_j, y_2 = r_k \in \mathcal{Y}, \quad (7.7)$$

entonces puede definirse el conjunto de entrenamiento modificado en pares $\tilde{\mathcal{T}}$ como la unión de dos subconjuntos disjuntos

$$\tilde{\mathcal{T}} = \mathcal{T}^{(\approx y)} \cup \mathcal{T}^{(> y, < y)}, \quad (7.8)$$

siendo

$$\begin{aligned} \mathcal{T}^{(\approx y)} &= \left\{ \left(\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)} \right), z_i \right\}_{i=1}^{\ell^{(1)}} \in \mathcal{X} \times \mathcal{X} \times \{0\} \\ \mathcal{T}^{(> y, < y)} &= \left\{ \left(\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)} \right), z_i \right\}_{i=1}^{\ell^{(2)}} \in \mathcal{X} \times \mathcal{X} \times \{-1, +1\} \end{aligned}, \quad (7.9)$$

con

$$z_i = \text{sign}(y_1 \ominus y_2), \quad (7.10)$$

y teniendo cada subconjunto un número de patrones igual a^1

$$\begin{aligned} \ell^{(1)} = \#\mathcal{T}^{(\approx y)} &= \sum_{i=1}^K \ell_i^2 \\ \ell^{(2)} = \#\mathcal{T}^{(> y, < y)} &= 2 \sum_{i=1}^{K-1} \sum_{j=i+1}^K \ell_i \ell_j \end{aligned}. \quad (7.11)$$

En los trabajos de [Herbrich et al., 1999b] y [Phillips, 1999] ya se utilizó una formulación parecida con la intención de aplicar máquinas de soporte vectorial a problemas de ordenación, pero en su caso, al basarse en la formulación estándar de la máquina de vectores soporte, sólo se tiene en consideración el subconjunto de entrenamiento $\mathcal{T}^{(> y, < y)}$ ya que las máquinas estándares sólo trabajan sobre dicotomías $+1/-1$. La cardinalidad de este subconjunto es mucho mayor que la del subconjunto descartado $\mathcal{T}^{(\approx y)}$ por lo que queda asegurada una débil pérdida de información durante el entrenamiento, pero esta realidad también provoca que el nuevo conjunto de entrenamiento sobre pares sea muy numeroso y, por tanto, sólo sea practicable trabajar sobre un subconjunto reducido de estos nuevos patrones. Así, considerando por ejemplo una condición usual como la de que todos los rangos tengan el mismo número de representantes, se obtiene el siguiente resultado

$$\begin{aligned} \ell^{(1)} = \#\mathcal{T}^{(\approx y)} &= \sum_{i=1}^K \ell_i^2 = K\tilde{\ell}^2 \\ \ell^{(2)} = \#\mathcal{T}^{(> y, < y)} &= 2 \sum_{i=1}^{K-1} \sum_{j=i+1}^K \ell_i \ell_j = (K-1)K\tilde{\ell}^2 \end{aligned}. \quad (7.12)$$

¹ Continuando con la notación de Capítulos anteriores, notaremos por ℓ_i el número de patrones que poseen como salida rango r_i .

La nueva aproximación aquí presentada del proceso de aprendizaje basado en las K-SVCRs, aunque parta de un conjunto de datos inicial más elevado, $\#\tilde{\mathcal{T}} = \ell^{(1)} + \ell^{(2)}$, que aquel trabajado en los artículos antes mencionados, acabará desarrollándose sobre un conjunto de patrones más reducido y significativo como más adelante se observará.

7.3 K-SVCR para Regresión Ordinal

En la línea de exposición de [Angulo and Català, 2000b] y [Herbrich et al., 1999b], utilizando las definiciones de la Sección anterior y la teoría de las máquinas K-SVCRs, se derivará un nuevo algoritmo para regresión ordinal.

Conjetura 7.4. *Se asumirá que para un conjunto de aproximación \mathcal{LM} , definido sobre funciones capaces de inducir un orden (7.3) sobre el espacio de entrada importado de la ordenación del espacio de salida \mathcal{Y} , existe un conjunto \mathcal{U} de funciones lineales sobre el espacio de características \mathcal{F} definidas desde el espacio origen \mathcal{X} hacia \mathbb{R} tal que para toda función $f \in \mathcal{LM}$ existe una función $U \in \mathcal{U}$ (y viceversa), denominada usualmente función utilidad, cumpliendo que*

$$f(\mathbf{x}, \omega) = r_i \iff U(\mathbf{x}) \in [\theta(r_{i-1}), \theta(r_i)] \quad (7.13)$$

donde

$$U(\mathbf{x}) = \langle \omega, \mathbf{x} \rangle_{\mathcal{F}} \quad (7.14)$$

con $\theta(r_0) = -\infty$ y $\theta(r_K) = \infty$.

Se demuestra con facilidad que la función utilidad, gracias a su linealidad dentro del espacio de características² \mathcal{F} , no incurre en error sobre el i -ésimo ejemplo del conjunto de entrenamiento modificado en pares $\tilde{\mathcal{T}}$ si, y sólo si, se cumplen la pareja de dobles implicaciones

$$\begin{aligned} z_i \cdot U(\mathbf{x}_i^{(1)}) > z_i \cdot U(\mathbf{x}_i^{(2)}) &\iff z_i \cdot U(\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}) > 0 \\ z_i \cdot U(\mathbf{x}_i^{(1)}) = z_i \cdot U(\mathbf{x}_i^{(2)}) &\iff z_i \cdot U(\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}) = 0 \end{aligned} \quad (7.15)$$

Las restricciones a cumplir por los pares de vectores d -dimensionales $(\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)})$ que poseen rangos diferentes, $z_i = \text{sign}(y_1 \ominus y_2) = \pm 1$, deberán permitir su separación sobre la recta de trabajo de la función utilidad, mientras que aquellos pares que son de igual rango deben restringirse a zonas acotadas que permitan definir los

² Obsérvese que la función utilidad es usualmente no lineal sobre el espacio de entrada original \mathbb{R}^d debido a la inserción realizada.

intervalos de funcionamiento de la función utilidad. De esta forma, pueden definirse las restricciones a cumplir por el QP problema a solucionar como

$$z_i \cdot \left\langle \omega, \mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)} \right\rangle_{\mathcal{F}} \geq 1 \quad i = 1, \dots, \ell^{(2)}, \quad (7.16)$$

$$-\delta \leq \left\langle \omega, \mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)} \right\rangle_{\mathcal{F}} \leq \delta \quad i = \ell^{(2)} + 1, \dots, \ell^{(2)} + \ell^{(1)}. \quad (7.17)$$

Para impedir la intersección entre los intervalos definidos para la función utilidad U se hace necesario restringir los valores del parámetro de insensitividad. Por ejemplo, serviría con $\delta < 0.5$ para asegurar que $|\theta(r_{i-1}), \theta(r_i)| < 1$ y evitar así esta posible intersección. Teniendo en cuenta esta consideración, ya puede realizarse la definición

Definición 7.5. *Se define el problema de optimización restringida asociado al método K-SVCR, para el caso general de regresión ordinal como: para $0 \leq \delta < 0.5$ elegido a priori,*

$$\arg \min R_{K-SVCR}^{Ord}(\omega, \xi, \varphi^{(*)}) = \frac{1}{2} \|\omega\|_{\mathcal{F}}^2 + C \sum_{i=1}^{\ell^{(2)}} \xi_i + D \sum_{i=\ell^{(2)}+1}^{\ell^{(2)}+\ell^{(1)}} (\varphi_i + \varphi_i^*), \quad (7.18)$$

sujeto a las restricciones (7.16)–(7.17).

Para solucionar el anterior QP problema se utiliza el lagrangiano

$$\begin{aligned} \arg \min L_p(\omega) = & \frac{1}{2} \|\omega\|_{\mathcal{F}}^2 - \\ & - \sum_{i=1}^{\ell^{(2)}} \alpha_i \left[z_i \cdot \left\langle \omega, \mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)} \right\rangle_{\mathcal{F}} - 1 \right] + \\ & + \sum_{i=\ell^{(2)}+1}^{\ell^{(2)}+\ell^{(1)}} \beta_i \left[\left\langle \omega, \mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)} \right\rangle_{\mathcal{F}} - \delta \right] - \\ & - \sum_{i=\ell^{(2)}+1}^{\ell^{(2)}+\ell^{(1)}} \beta_i^* \left[\left\langle \omega, \mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)} \right\rangle_{\mathcal{F}} + \delta \right] \end{aligned}, \quad (7.19)$$

sujeto a

$$\begin{aligned} \frac{\partial L_p}{\partial \alpha_i} = 0 \quad ; \quad \alpha_i \geq 0 \quad i = 1, \dots, \ell^{(2)} \\ \frac{\partial L_p}{\partial \beta_i^*} = 0 \quad ; \quad \beta_i^* \geq 0 \quad i = \ell^{(2)} + 1, \dots, \ell^{(2)} + \ell^{(1)} \end{aligned}, \quad (7.20)$$

que llevado al espacio dual y utilizando la notación

$$\gamma_i = \alpha_i \cdot z_i \quad i = 1, \dots, \ell^{(2)} \quad (7.21)$$

nos conduce a tener que solucionar el problema restringido equivalente

$$\begin{aligned} \arg \min L_p(\gamma) = & \\ & \frac{1}{2} \left\langle \left[\sum_{i=1}^{\ell^{(2)}} \gamma_i \cdot (\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}) - \sum_{i=\ell^{(2)+1}^{\ell^{(2)}+\ell^{(1)}} (\beta_i - \beta_i^*) \cdot (\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}) \right], \right. \\ & \left. \left[\sum_{j=1}^{\ell^{(2)}} \gamma_j \cdot (\mathbf{x}_j^{(1)} - \mathbf{x}_j^{(2)}) - \sum_{j=\ell^{(2)+1}^{\ell^{(2)}+\ell^{(1)}} (\beta_j - \beta_j^*) \cdot (\mathbf{x}_j^{(1)} - \mathbf{x}_j^{(2)}) \right] \right\rangle_{\mathcal{F}}, \quad (7.22) \\ & - \sum_{i=1}^{\ell^{(2)}} \frac{\gamma_i}{z_i} + \delta \sum_{i=\ell^{(2)+1}^{\ell^{(2)}+\ell^{(1)}} (\beta_i + \beta_i^*) \end{aligned}$$

con

$$\begin{aligned} \frac{\gamma_i}{z_i} &\geq 0 \quad i = 1, \dots, \ell^{(2)} \\ \beta_i, \beta_i^* &\geq 0 \quad i = \ell^{(2)} + 1, \dots, \ell^{(2)} + \ell^{(1)} \end{aligned} \quad (7.23)$$

Finalmente, realizando las substitutiones

$$\begin{aligned} \gamma_i &= \beta_i \quad i = \ell^{(2)} + 1, \dots, \ell^{(2)} + \ell^{(1)} \\ \gamma_i &= \beta_{i-\ell^{(1)}} \quad i = \ell^{(2)} + \ell^{(1)} + 1, \dots, \ell^{(2)} + 2\ell^{(1)} \end{aligned} \quad (7.24)$$

se llega al enunciado final del problema de optimización dual de Wolfe: para $0 \leq \delta < 0.5$ elegido *a priori*,

$$\arg \min W(\gamma) = \frac{1}{2} \gamma^\top \cdot \mathbf{H} \cdot \gamma + \mathbf{c}^\top \cdot \gamma, \quad (7.25)$$

con

$$\gamma^\top = (\gamma_1, \dots, \gamma_{\ell^{(2)}}, \gamma_{\ell^{(2)}+\ell^{(1)}}, \dots, \gamma_{\ell^{(2)}+2\ell^{(1)}}) \in \mathbb{R}^{\ell^{(2)}+\ell^{(1)}+\ell^{(1)}}, \quad (7.26)$$

$$\mathbf{c}^\top = \left(\frac{-1}{z_1}, \dots, \frac{-1}{z_{\ell^{(2)}}}, \delta, \dots, \delta \right) \in \mathbb{R}^{\ell^{(2)}+\ell^{(1)}+\ell^{(1)}}, \quad (7.27)$$

$$\mathbf{H} = \begin{pmatrix} (k_{i,j}) & -(k_{i,j}) & (k_{i,j}) \\ -(k_{i,j}) & (k_{i,j}) & -(k_{i,j}) \\ (k_{i,j}) & -(k_{i,j}) & (k_{i,j}) \end{pmatrix}, \quad (7.28)$$

siendo

$$(k_{i,j}) = \left\langle \mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}, \mathbf{x}_j^{(1)} - \mathbf{x}_j^{(2)} \right\rangle_{\mathcal{F}} \in \mathbb{R}^{\ell^{(a)}} \times \mathbb{R}^{\ell^{(b)}}, \quad (7.29)$$

$$\mathbf{H} = \mathbf{H}^\top \in \mathcal{M} \left(\mathbb{R}^{\ell^{(2)}+\ell^{(1)}+\ell^{(1)}}, \mathbb{R}^{\ell^{(2)}+\ell^{(1)}+\ell^{(1)}} \right), \quad (7.30)$$

sujeto a

$$\begin{aligned} \gamma_i \cdot z_i &\geq 0 \quad i = 1, \dots, \ell^{(2)} \\ \gamma_i &\geq 0 \quad i = \ell^{(2)} + 1, \dots, \ell^{(2)} + 2\ell^{(1)} \end{aligned} \quad (7.31)$$

En esta ocasión, la función solución f se halla a partir de la función utilidad U mediante su propiedad definitoria (7.13) y ésta se expresa como

$$U(\mathbf{x}) = \left\langle \sum_{i=1}^{\ell^{(2)}} \gamma_i \cdot (\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}) - \sum_{i=\ell^{(2)+1}^{\ell^{(2)}+\ell^{(1)}} (\gamma_i - \gamma_{i+\ell^{(1)}}) \cdot (\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}), \mathbf{x} \right\rangle_{\mathcal{F}} \quad (7.32)$$

que puede reescribirse en

$$U(\mathbf{x}) = \left\langle \sum_{i=1}^{SV} \nu_i \cdot (\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}), \mathbf{x} \right\rangle_{\mathcal{F}}, \quad (7.33)$$

si se utiliza la notación

$$\begin{aligned} \nu_i &= \gamma_i & i &= 1, \dots, \ell^{(2)} \\ \nu_i &= -\gamma_i + \gamma_{i+\ell^{(1)}} & i &= \ell^{(2)} + 1, \dots, \ell^{(2)} + \ell^{(1)} \end{aligned} \quad (7.34)$$

La intención de esta nueva formulación del problema de regresión ordinal es centrar el entrenamiento de la maquina de aprendizaje sobre pares de patrones que pertenecen al mismo rango, es decir aquellos que comportan salida 0, estableciendo que la amplitud de la función utilidad sobre cada rango dentro de una recta inscrita en el espacio \mathcal{F} , $U(\mathbf{x}) = \langle \omega, \mathbf{x} \rangle_{\mathcal{F}}$, es como máximo 1. En cambio, en el desarrollo comparado de [Herbrich et al., 1999b] y [Phillips, 1999] se tiene en consideración únicamente los pares de patrones que pertenecen a rangos diferentes, no pudiendo utilizar toda la información relativa a los pares de patrones de igual rango.

El conjunto de restricciones (7.17) dentro del QP problema original (7.18) garantiza la inclusión de cada rango dentro de una bola acotada, mientras que el conjunto de restricciones (7.16) sólo es útil para realizar la distribución de estos intervalos sobre una recta en el espacio de características \mathcal{F} . Por tanto, en el caso más extremo, considerando únicamente un par de patrones por cada par de rangos diferentes se dispondría de la suficiente información para realizar la distribución sobre la recta. La nueva formulación es capaz de trabajar con un número de restricciones que se puede reducir a $2\ell^{(1)} + K(K-1)$ desigualdades³. Si por ejemplo se considera, tal como se realizó al final de la Sección 7.2, que todos los rangos vienen representados por un mismo número de representantes, ℓ , se demuestra de forma directa que la

³ Yendo algo más allá, a nivel práctico puede observarse que todas las restricciones en (7.17) cumpliendo $\mathbf{x}_i^{(1)} = \mathbf{x}_i^{(2)}$ pueden expresarse de forma única como

$$-\delta \leq \langle \omega, 0 \rangle_{\mathcal{F}} \leq \delta \quad i = \ell^{(2)} + 1, \dots, \ell^{(2)} + \ell^{(1)}, \quad (7.35)$$

por lo que estas ℓ restricciones pueden ser tomadas por 1 sola.

nueva formulación trabaja sobre $2K\tilde{\ell}^2 + K(K-1)$ desigualdades, mientras que aquella definida en [Herbrich et al., 1999b] o [Phillips, 1999] trabaja sobre $K(K-1)\tilde{\ell}^2$ desigualdades. Esto significaría para el ejemplo seguido desde el Capítulo 4 de un problema de multclasificación con $\tilde{\ell} = 1000$ patrones por clase y $K = 10$ clases que la nueva propuesta utiliza $20 \cdot 10^6 + 90$ desigualdades⁴ mientras que aquella que se ha tomado por referencia utiliza $90 \cdot 10^6$. Puesto que el número de representantes por clase es moderadamente grande, tal como ya se anunció el número de restricciones resulta de una magnitud muy elevada, de lo que se observa que este tipo de formulación está especialmente dedicada a problemas con pocos datos de entrenamiento. En cualquier caso, el índice de reducción en el número de restricciones es de aproximadamente $2/K$ respecto a la formulación que se ha tomado como base de referencia sobre máquinas SVMC estándares.

El único punto crítico de la presente formulación es aquel que hace referencia a la estimación de los valores frontera $\theta(r_i)$. Es necesario asegurar que los pares de patrones de rangos diferentes tomados en consideración durante el entrenamiento no posean ‘ruido’ o distorsión alguna, o lo que es equivalente, si se hubiera tenido en consideración la posibilidad de introducir variables artificiales para permitir errores en los valores fronteras, éstos deberían estar asociados a una $\xi_i = 0$, hecho que en ningún caso puede afirmarse *a priori* de modo general. Por otra parte, si estas variables artificiales hubieran sido consideradas deberían haber sido acotadas para asegurar la no intersección entre los intervalos creados por la función utilidad⁵ por lo que se hubieran añadido un conjunto de restricciones a la definición del problema de regresión ordinal (7.18) que alterarían significativamente la bondad del resultado estándar sobre vectores soporte. Una vez asegurada esta condición aún a costa de no utilizar variables ‘slack’, la estimación del rango quedará definida fijando los extremos $\theta(r_i)$ de los intervalos definitorios. Con la intención de maximizar el margen, estas cotas vendrán dadas por la formulación

$$\begin{aligned} \theta(r_k) &= \frac{U(\mathbf{x}_1) + U(\mathbf{x}_2)}{2} = \frac{\langle \omega, \mathbf{x}_1 \rangle_{\mathcal{F}} + \langle \omega, \mathbf{x}_2 \rangle_{\mathcal{F}}}{2} \\ &= \frac{\left\langle \sum_{i=1}^{SV} \nu_i \cdot \left(\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)} \right), \mathbf{x}_1 \right\rangle_{\mathcal{F}} + \left\langle \sum_{i=1}^{SV} \nu_i \cdot \left(\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)} \right), \mathbf{x}_2 \right\rangle_{\mathcal{F}}}{2}, \end{aligned} \quad (7.36)$$

siendo los vectores \mathbf{x}_1 y \mathbf{x}_2 utilizados aquellos tal que

$$(\mathbf{x}_1, \mathbf{x}_2) = \arg \min_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{X}\mathcal{X}(k)} [U(\mathbf{x}_i) - U(\mathbf{x}_j)], \quad (7.37)$$

con

$$\mathcal{X}\mathcal{X}(k) = \left\{ \left(\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)} \right) \mid y_i^{(1)} = r_k \wedge y_i^{(2)} = r_{k+1} \right\}. \quad (7.38)$$

⁴ Haciendo uso de la observación anterior, esta cantidad se reduciría en la práctica en $10^4 - 1$ restricciones.

⁵ Por ejemplo podía haberse impuesto *a priori* que $\delta \leq 0.25$ y además restringir las variables artificiales a $\xi_i \leq 0.125$.

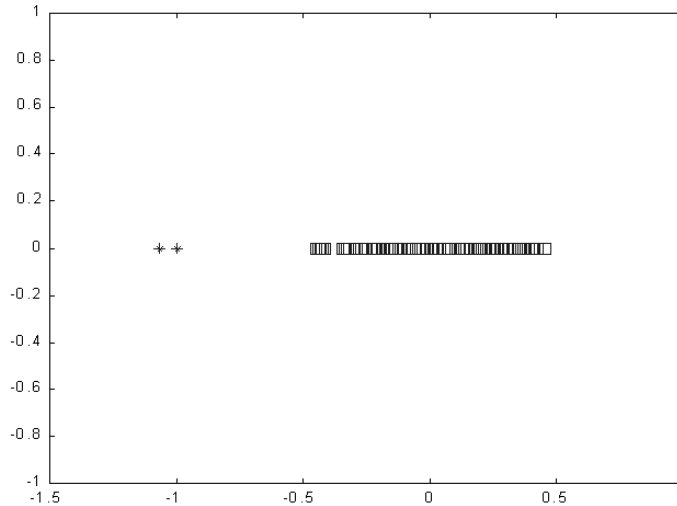


Figura 7.1: Cumplimiento de las restricciones del QP problema por parte de los patrones diferencia $\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}$.

Acabado el entrenamiento de la máquina de aprendizaje y una vez definidos los valores frontera de los rangos, los nuevos elementos a clasificar se etiquetarán utilizando la expresión (7.13)⁶. En el caso no lineal que ocupa el estudio debe tenerse en cuenta la observación realizada en [Herbrich et al., 2000] de realizar el producto interno en el espacio \mathcal{F} entre un elemento diferencia $\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}$ y un vector de test \mathbf{x} como una resta entre núcleos más que el núcleo de una diferencia $\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}$ por un vector \mathbf{x}

$$U(\mathbf{x}) = \langle \omega, \mathbf{x} \rangle_{\mathcal{F}} = \sum_{i=1}^{SV} \nu_i \cdot \mathcal{K}(\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}, \mathbf{x}), \quad (7.39)$$

con

$$\mathcal{K}(\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}, \mathbf{x}) = k(\mathbf{x}_i^{(1)}, \mathbf{x}) - k(\mathbf{x}_i^{(2)}, \mathbf{x}), \quad (7.40)$$

para que así puedan mantenerse las propiedades de antisimetría y transitividad en la ordenación inducida del espacio de entrada \mathcal{X} .

En las gráficas adjuntas se muestra el funcionamiento de este tipo de aprendizaje de preferencias. En la Figura 7.1 se aprecia como se cumplen las restricciones del QP problema asociado sobre los patrones diferencias, mientras que en la Figura 7.2 puede observarse como los nuevos vectores de entrada son ordenados por la función utilidad dentro de intervalos acotados de la recta real.

⁶ Aunque esta es la forma de actuación más común en la literatura, resulta algo forzado evaluar la función utilidad sobre elementos del estilo \mathbf{x}_i cuando todo el proceso de aprendizaje se ha realizado sobre pares de patrones expresados en la forma $\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}$. Una nueva posibilidad a estudiar de evaluación de los patrones de test sería aquella de comparación con los patrones que han sido utilizados para el aprendizaje.

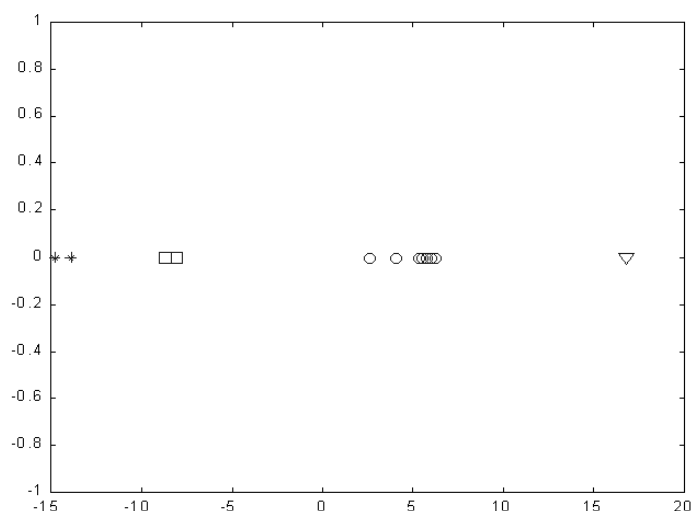


Figura 7.2: Ordenación de las entradas \mathbf{x} dentro de la recta real mediante la función utilidad $U(\mathbf{x})$.

7.4 Problema Financiero de Riesgo Crediticio

El resultado que se persigue en esta aplicación es la creación de un sistema clasificador que permita realizar una ordenación sobre el riesgo crediticio de diferentes empresas en función de una serie de ratios económicas. Por riesgo crediticio — del inglés, *credit risk* — se entiende la evaluación que otorga un grupo de expertos, en general una tercera empresa como Fitch, Moody's, Standard & Poor's, etc, intentando reflejar la capacidad de la empresa para solventar sus deudas y su capacidad de pago. Esta calificación se establece mediante etiquetas que van desde 'AAA' para aquellas que obtienen una medida de solvencia sobresaliente, etiquetas de niveles intermedios como por ejemplo 'BBB+' o 'BB' si la confianza en la devolución es baja, hasta 'B' en el caso de que el índice de riesgo de no devolución sea muy alto⁷. La asignación de este índice de riesgo no es una cuestión baladí puesto que tanto las entidades financieras en la concesión de sus posibles créditos como los inversores en el manejo de su portafolio lo tienen muy presente en la toma de sus decisiones.

En una primera aproximación al problema deben determinarse las variables del espacio de entrada que resulten representativas para mostrar el estado financiero y de negocio de las empresas por lo que se hace necesaria la presencia de un experto que ayude en esta decisión⁸, aunque, como era de esperar, ya existe un conjunto amplio

⁷ Existen calificaciones de nivel inferior pero la dispersión de las variables económicas de estas empresas desaconsejan su utilización en esta clase de valoraciones.

⁸ Agradecer desde aquí la colaboración y tiempo dedicado en la elección de estas variables a Carmen Ansotegui, Xari Rovira y Nuria Agell, profesoras y doctoras de diferentes departamentos de la Escuela Superior de Administración y Dirección de Empresas (ESADE) - Universitat Ramon Llull.

de indicadores estándares que facilitan la selección⁹. Para esta aplicación se decidió tomar como entradas los 5 indicadores siguientes: la cobertura de intereses el año anterior y las ratios de endeudamiento, de rendimiento sobre activo, de deuda a corto sobre largo y de activo circulante sobre total. Por otra parte, se han separado las empresas por sectores de producción, en total se han considerado 7 sectores, debido a la especificidad que muestran éstas respecto a su área de negocio para así, además de crear una máquina de aprendizaje sobre la globalidad de los datos, desarrollar para cada sector una máquina de aprendizaje específica.

Por lo que se refiere al conjunto de entrenamiento, los patrones han sido recopilados de diferentes bases de datos y las salidas corresponden a las calificaciones de riesgo crediticio asignadas por la empresa Standard & Poor's para el año 1998¹⁰. Sólo puede disponerse de un conjunto de empresas reducido de las cuáles se tengan todos los datos de entrada necesarios y su r ating crediticio. Aunque en principio esto resulta un contratiempo para poder realizar un adecuado entrenamiento, es bien conocida la gran capacidad de las SVMs para generalizar sobre un n mero peque o de datos en entornos de gran dimensionalidad, por lo que resultan una elecci n del todo adecuada. Adem s, la m quina K-SVCR para regresi n ordinal presentada en la Secci n anterior se ve a obligada a trabajar sobre un n mero de representantes peque o si se quer a evitar la explosi n en la dimensi n del conjunto de entrenamiento modificado sobre pares. De esta forma, se tomar n como patrones de aprendizaje aquellas empresas de las que se dispone de todos los datos y como patrones de test las empresas de las que carecemos de la informaci n de salida, su riesgo crediticio, pero cuyo evaluaci n aproximada no resulta demasiado complicada en una primera valoraci n para un experto. Esta decisi n ha sido tomada teniendo en cuenta el n mero tan reducido de datos disponibles, hecho que motiva no reservar en un principio datos reales para la fase de test¹¹.

Los rangos de salida tomados han sido 4, elegidos de forma que se intenta minimizar la cantidad de informaci n perdida en esta reducci n del n mero de etiquetas, que se corresponder an con las calificaciones ordenadas de mayor a menor seguridad en el retorno del cr dito {'AAA', 'AA'}, 'A', 'BBB' y {'BB', 'B'}, con sus respectivas modificaciones en la etiqueta mediante los signos + y -. El n mero de datos disponibles para cada relaci n sector-r ating se encuentra representado en la Tabla 7.1.

Como puede observarse existe una gran dispersi n en el n mero de patrones de

⁹ Al respecto, destacar que incluso una SVMC puede servirnos como selector de  temes representativos, tal como se recoge en el estudio [Fan and Palaniswami, 2000] sobre selecci n de predictores de bancarrota.

¹⁰Debe agradecerse a Montserrat Caellas de la Escuela Superior de Administraci n y Direcci n de Empresas (ESADE) - Universitat Ramon Llull el tiempo y esfuerzo dedicados en la elaboraci n de la base de datos que ha permitido realizar el presente an lisis.

¹¹Durante la elaboraci n del estudio continu  el proceso de b squeda de la informaci n no disponible y en alg n caso se obtuvo recompensa. Sin embargo, estos datos se mantuvieron ignorados durante el proceso de entrenamiento de las m quinas para conseguir que el test fuera completamente 'a ciegas'.

	Test	Entrenamiento				Total
		1	2	3	4	
1	9	18	19	18	1	65
2	6	11	23	21	7	68
3	10	3	10	8	6	37
4	7	7	17	5	0	36
6	1	1	11	11	2	26
7	4	3	26	14	2	49
8	2	2	2	7	1	14
Total	39	45	108	84	19	295
	39	256				

Tabla 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.

entrenamiento y del conjunto de test, por lo que la eficacia en la ordenación variará en función del rango de asignación del patrón mostrado. Aún más, existen casos como el rango 4 en el sector 4 que no posee representante alguno por lo que no podrá mostrar su existencia ante la máquina de aprendizaje y en otras 3 oportunidades los patrones son unrepresentativos, como en el rango 1 del sector 6 por ejemplo, sobre los cuales no será posible construir pares de patrones dentro de la misma clase, por lo que sólo será factible aprender la disimilaridad entre rangos. Por último, señalar que los conjuntos de entrenamiento originales fueron normalizados de forma que los datos tuvieran una distribución normal de media nula y varianza unitaria, $\mathcal{N}(0, 1)$.

7.4.1 Uso de máquinas K -SVCR en multclasificación.

Con el fin de poder contrastar los resultados que se obtengan con la máquina que aprende preferencias, definida en exclusividad para problemas de ordenación, se ha desarrollado en paralelo una arquitectura basada en máquinas K -SVCR y elemento combinación tal como se desarrolló en el Capítulo 6. Puesto que los datos de entrada no poseen en principio error ni ruido alguno se considerará que los parámetros de penalización o regularización de la máquina de aprendizaje tienen un valor muy elevado, $C = D \approx \infty$. De esta forma sólo resta por determinar qué tipo de núcleo utilizar, cuál será el parámetro asociado elegido y que grado de insensitividad δ se tendrá en consideración. Al no disponer de las respuestas válidas del conjunto de test no es posible aplicar técnicas de tipo de validación cruzada o similares, mientras que al ser el conjunto de entrenamiento muy reducido se estaría perdiendo una información muy valiosa si se reservara una parte del conjunto de entrenamiento como conjunto de validación tal como es práctica habitual entre los investigadores para sintonizar los parámetros. Por todo ello, se ha optado por realizar un doble entrenamiento considerando núcleos de tipo polinomial que generan espacios de dimensión VC finita y

núcleos funciones gaussianas que hacen infinita la dimensión VC del espacio no lineal donde se inserta el conjunto original.

En cuanto a los parámetros de cada tipo de núcleo y el factor de insensitividad, se ha determinado entrenar toda una serie de máquinas con diferentes elecciones de estos parámetros para al final poder considerar como resultado adecuado aquel que posea una mayor proporción de votos global. Para el caso de las funciones gaussianas, se han entrenado máquinas para 20 posibles valores del parámetro de varianza de forma que $\sigma \in [0.1, 1]$ y para cada valor de varianza 10 máquinas con parámetro de insensitividad $\delta \in [0.09, 0.90]$, lo que conforma un total de 200 máquinas. Si bien para valores elevados de δ y pequeños de σ la generalización será escasa, mientras que en condiciones inversas seguramente la máquina dará como posible salida válida el conjunto completo de clases debido a la gran indefinición de los núcleos clasificadores, se considera que como promedio la clase que resulte adecuada será aquella que más votaciones obtendrá sobre el global de todas las máquinas entrenadas. Los resultados del test obtenidos pueden observarse en la Tabla 7.2.

SECTOR 1	1	2	3	4
Wellpoin	00.47%	65.26%	06.10%	28.17%
Baker Hu	02.16%	38.96%	21.21%	37.66%
Meditroni	32.51%	00.49%	05.91%	61.08%
Loreal	62.81%	00.50%	04.52%	32.16%
Dixons	22.13%	22.98%	14.04%	40.85%
Bayer	64.04%	00.98%	04.93%	30.05%
Basf	6.80%	00.49%	54.85%	37.86%
Winn-Dix	09.81%	19.62%	20.75%	49.81%
United H	23.02%	26.80%	14.43%	35.74%

SECTOR 2	1	2	3	4
Genuine	15.14%	48.56%	20.62%	15.67%
Ford Mot	19.62%	39.02%	16.42%	24.95%
Meredith	25.07%	36.75%	21.65%	16.52%
McGraw-H	72.61%	13.69%	06.64%	07.05%
Interpub	17.33%	33.87%	33.33%	15.47%
Ikkon Off	09.61%	67.62%	12.81%	09.96%

SECTOR 3	1	2	3	4
Adobe Sy	21.95%	39.91%	19.73%	18.40%
Andrew	20.10%	22.13%	42.24%	15.52%
Gateway	21.95%	39.91%	19.73%	18.40%
Cabletro	15.50%	17.38%	54.27%	12.80%
Foster W	15.11%	59.49%	13.18%	12.22%
Shared M	17.87%	51.40%	16.48%	14.25%
Scientific	19.61%	35.56%	29.74%	15.09%

SAP	13.10%	15.17%	61.38%	10.34%
Olivetti	13.95%	62.93%	12.24%	10.88%
Cable	48.78%	19.24%	16.26%	15.72%

SECTOR 4	1	2	3	4
Ctl & Sw	09.05%	81.90%	09.05%	00.00%
Entergy	06.82%	86.36%	06.82%	00.00%
Pinnacle	28.51%	37.56%	33.93%	00.00%
Telecom	21.39%	57.23%	21.39%	00.00%
Siemens	21.73%	56.55%	21.73%	00.00%
Hoechst	33.93%	40.31%	25.77%	00.00%
Electrab	27.59%	44.81%	27.59%	00.00%

SECTOR 6	1	2	3	4
Sigma Al	05.76%	68.35%	19.78%	06.12%

SECTOR 7	1	2	3	4
Gen Elec	06.71%	57.93%	28.66%	06.71%
Nacco In	11.46%	37.26%	42.99%	08.28%
Volvo	04.98%	72.79%	17.24%	04.98%
Fiat	04.05%	76.92%	14.98%	04.05%

SECTOR 8	1	2	3	4
Schlumbe	14.79%	14.50%	56.21%	14.50%
Royal Du	15.04%	14.45%	56.05%	14.45%

Tabla 7.2: Resultados sobre los conjuntos de test separados por sectores utilizando máquinas K -SVCR en multi-clasificación con núcleos gaussianos.

De igual forma se ha actuado para el caso de núcleos polinomiales, considerando idéntica variación de insensitividad y dejando variar el grado del polinomio en el intervalo $n \in [1, 6]$, lo que genera tras entrenar 60 máquinas los resultados que se presentan en la Tabla 7.3.

SECTOR 1	1	2	3	4
Wellpoin	23.88%	41.79%	20.90%	13.43%
Baker Hu	09.21%	47.37%	43.42%	00.00%
Meditroni	46.15%	03.08%	29.23%	21.54%
Loreal	60.61%	34.85%	03.03%	01.52%
Dixons	19.72%	40.85%	16.90%	22.54%
Bayer	63.51%	17.57%	12.16%	06.76%
Basf	45.95%	09.91%	28.83%	15.32%

Winn-Dix	12.82%	42.31%	29.49%	15.38%
United H	42.67%	48.00%	00.00%	09.33%

SECTOR 2	1	2	3	4
Genuine	13.43%	79.10%	07.46%	00.00%
Ford Mot	12.16%	58.11%	06.76%	22.97%
Meredith	55.56%	33.33%	04.17%	06.94%
McGraw-H	61.54%	30.77%	00.00%	07.69%
Interpub	20.22%	34.83%	21.35%	23.60%
Ikkon Off	05.88%	64.71%	29.41%	00.00%

SECTOR 3	1	2	3	4
Adobe Sy	00.00%	00.00%	18.03%	81.97%
Andrew	00.00%	00.00%	100.0%	00.00%
Gateway	09.86%	56.34%	05.63%	28.17%
Cabletro	01.56%	25.00%	65.62%	07.81%
Foster W	00.00%	96.72%	01.64%	01.64%
Shared M	14.52%	64.52%	00.00%	20.97%
Scientific	02.60%	46.75%	48.05%	02.60%
SAP	09.59%	20.55%	69.86%	00.00%
Olivetti	00.00%	98.36%	01.64%	00.00%
Cable	55.42%	10.84%	24.10%	09.64%

SECTOR 4	1	2	3	4
Ctl & Sw	04.55%	87.88%	07.58%	00.00%
Entergy	05.88%	85.29%	08.82%	00.00%
Pinnacle	30.47%	35.16%	34.38%	00.00%
Telecom	15.48%	69.05%	15.48%	00.00%
Siemens	12.50%	48.75%	38.75%	00.00%
Hoechst	68.18%	19.70%	12.12%	00.00%
Electrab	44.59%	45.95%	09.46%	00.00%

SECTOR 6	1	2	3	4
Sigma Al	06.06%	51.52%	16.67%	25.76%

SECTOR 7	1	2	3	4
Gen Elec	05.45%	47.27%	34.55%	12.73%
Nacco In	15.29%	57.65%	16.47%	10.59%
Volvo	00.00%	82.00%	18.00%	00.00%
Fiat	00.00%	80.00%	20.00%	00.00%

SECTOR 8	1	2	3	4
Schlumbe	01.39%	30.56%	30.56%	37.50%

Royal Du	29.73%	10.81%	39.19%	20.27%
----------	---------------	--------	---------------	--------

Tabla 7.3: Resultados sobre los conjuntos de test separados por sectores utilizando máquinas K -SVCR en multi-clasificación con núcleos polinomiales.

Una comparación de los resultados de la Tabla 7.2 y la Tabla 7.3 muestra que en un 71.8% de las empresas las máquinas clasificadoras sobre ambos tipos de núcleo se muestran completamente de acuerdo sobre un único rango como salida posible por lo que puede afirmarse que el riesgo crediticio de este subconjunto de empresas ha sido aprendido por la arquitectura de clasificación. Con respecto a las divergencias, el restante 28.2% de las empresas han sido etiquetadas en clases diferentes en función del núcleo utilizado aunque habría que hacer notar que de este porcentaje más de un 45% corresponde a discrepancias existentes sobre patrones test de empresas del sector de producción 1. Finalmente, si se extiende el rango de salida en los casos de existencia de porcentajes similares para dos etiquetas, éstos representan alrededor de un 10% de empresas de las que podría decirse que su clasificación no está bien definida.

Además de entrenar separadamente para cada sector se ha considerado la posibilidad de realizar un entrenamiento conjunto, si bien esta opción era menos fiable en opinión de los expertos. De nuevo se optó por utilizar núcleos gaussianos y polinomiales y un conjunto de parámetros idéntico a aquel utilizado cuando los sectores se analizaron individualmente. De las 4 predicciones obtenidas en total — sectores individuales y en global y núcleos polinomiales y gaussianos —, en un 28.21% de los casos existió unanimidad, mientras que un porcentaje de 38.46% empresas obtuvo tres predicciones iguales y una discordante, lo que puede interpretarse como que el riesgo de crédito de 2 de cada 3 empresas ha sido aprendido. En un 25.64% de casos existió igualdad de resultado entre las máquinas entrenadas sobre sectores individuales, pero las dos máquinas globales, aquellas menos fiables según los expertos, predijeron salidas diferentes, mientras que el 7.69% restante mostró unas predicciones totalmente dispares.

Finalmente, destacar que aquellas empresas de las que se obtuvo posteriormente su rating fueron bien clasificadas. ‘Wellpoin’, ‘Baker Hu’ y ‘Ford Mot’ obtuvieron en las 4 predicciones un nivel 2 como así se comprobó que era correcto, mientras que ‘BASF’ obtuvo su rango exacto, 1, en 3 de las 4 predicciones.

7.4.2 Uso de máquinas K -SVCR en regresión ordinal.

Las máquinas K -SVCR en regresión ordinal serán ahora utilizadas sobre la aplicación financiera. En este caso un estudio de todos los sectores en conjunto, además de

desaconsejable desde el punto de vista experto resulta computacionalmente muy caro puesto que el conjunto de entrenamiento modificado en pares tendrá una cardinalidad muy elevada, por lo que se realizará una ordenación de las empresas ordenadas por sectores. Una vez realizado el entrenamiento se obtienen las divisiones intervalares de la recta real, algunas de las cuales pueden apreciarse en las gráficas de la Figura 7.3 para diferentes valores de σ .

En cuanto a los parámetros de cada tipo de núcleo y el factor de insensitividad, se ha determinado entrenar, al igual que en el caso de la Sección anterior, toda una serie de máquinas con diferentes elecciones de estos parámetros. Para el caso de las funciones gaussianas, se han entrenado máquinas para 20 posibles valores del parámetro de varianza de forma que $\sigma \in [0.1, 1]$ y para cada valor de varianza 5 máquinas con parámetro de insensitividad $\delta \in [0.09, 0.45]$ ¹², lo que conforma un total de 100 máquinas. Los resultados del test obtenidos pueden observarse en la Tabla 7.4.

SECTOR 1	1	2	3	4
Wellpoin	15.56%	84.44%	00.00%	00.00%
Baker Hu	42.22%	57.78%	00.00%	00.00%
Meditroni	20.00%	80.00%	00.00%	00.00%
Loreal	66.67%	33.33%	00.00%	00.00%
Dixons	31.11%	68.89%	00.00%	00.00%
Bayer	46.67%	53.33%	00.00%	00.00%
Basf	02.22%	82.22%	15.56%	00.00%
Winn-Dix	02.22%	97.78%	00.00%	00.00%
United H	40.00%	60.00%	00.00%	00.00%

SECTOR 2	1	2	3	4
Genuine	13.33%	46.67%	40.00%	00.00%
Ford Mot	00.00%	06.67%	93.30%	00.00%
Meredith	86.67%	13.33%	00.00%	00.00%
McGraw-H	100.0%	00.00%	00.00%	00.00%
Interpub	00.00%	46.67%	53.33%	00.00%
Ikkon Off	100.0%	00.00%	00.00%	00.00%

SECTOR 3	1	2	3	4
Adobe Sy	06.00%	28.00%	66.00%	00.00%
Andrew	04.00%	12.00%	80.00%	04.00%
Gateway	06.00%	28.00%	66.00%	00.00%
Cabletro	02.00%	10.00%	84.00%	04.00%
Foster W	28.00%	44.00%	28.00%	00.00%
Shared M	24.00%	50.00%	26.00%	00.00%

¹²Recuérdese que el parámetro de insensitividad debe ser menor que 0.50 para asegurar la no intersección entre intervalos.

Scientific	04.00%	20.00%	76.00%	00.00%
SAP	02.00%	06.00%	88.00%	04.00%
Olivetti	22.00%	40.00%	34.00%	04.00%
Cable	40.00%	36.00%	24.00%	00.00%

SECTOR 4	1	2	3	4
Ctl & Sw	46.78%	52.05%	01.17%	00.00%
Entergy	01.17%	56.14%	42.70%	00.00%
Pinnacle	00.00%	38.01%	61.99%	00.00%
Telecom	01.17%	97.66%	01.17%	00.00%
Siemens	01.17%	97.66%	01.17%	00.00%
Hoechst	39.77%	59.06%	01.17%	00.00%
Electrab	03.51%	95.32%	01.17%	00.00%

SECTOR 6	1	2	3	4
Sigma Al	02.34%	62.57%	34.50%	00.58%

SECTOR 7	1	2	3	4
Gen Elec	00.00%	94.00%	06.00%	00.00%
Nacco In	00.00%	38.00%	62.00%	00.00%
Volvo	00.00%	96.00%	04.00%	00.00%
Fiat	00.00%	98.00%	02.00%	00.00%

SECTOR 8	1	2	3	4
Schlumbe	01.17%	02.92%	95.32%	00.58%
Royal Du	01.17%	05.85%	92.40%	00.58%

Tabla 7.4: Resultados sobre los conjuntos de test separados por sectores utilizando máquinas K -SVCR en regresión ordinal con núcleos gaussianos.

Sobre los resultados obtenidos con las máquinas K -SVCR en multclasificación y en regresión ordinal usando núcleos gaussianos pueden realizarse las siguientes observaciones.

- El análisis de porcentajes en las respuestas resulta en la emisión de una predicción sobre los rangos que en ambos casos son muy similares, por lo que puede entenderse que ambas técnicas permiten un nivel de aprendizaje similar¹³
- En el caso de la regresión ordinal los porcentajes de test son significativos ya que el aprendizaje se ha realizado también sobre la preferencia de orden, por

¹³El presente razonamiento puede extenderse también al uso de núcleos polinomiales.

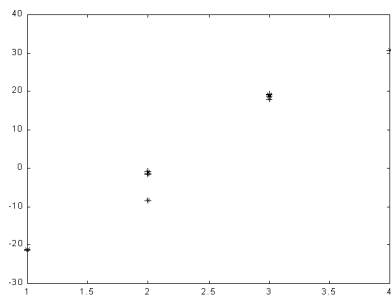
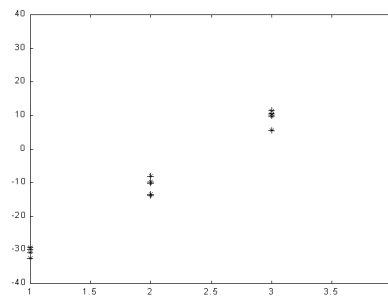
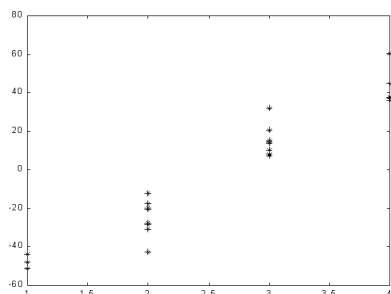
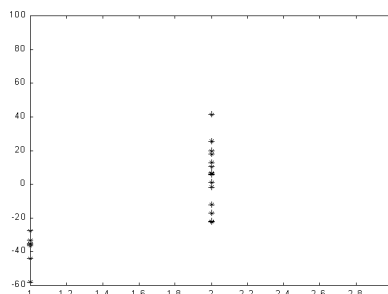
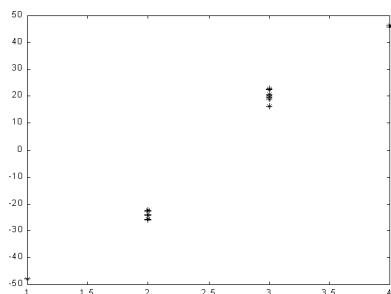
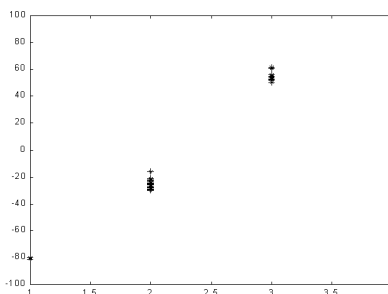
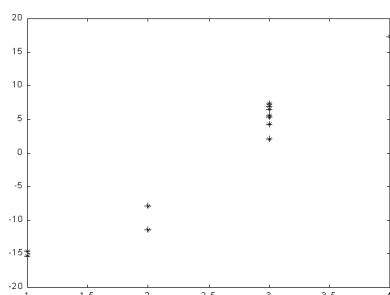
(a) Sector 1, $\sigma = 0.175$ (b) Sector 2, $\sigma = 0.125$ (c) Sector 3, $\sigma = 0.200$ (d) Sector 4, $\sigma = 0.350$ (e) Sector 6, $\sigma = 0.150$ (f) Sector 7, $\sigma = 0.150$ (g) Sector 8, $\sigma = 0.475$

Figura 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.

lo que pudiera pensarse que es posible además de determinar el *r ating* general, indicar la tendencia de esa empresa o su situaci n dentro del *r ating* pudiendo realizarse comparaciones de situaci n financiera entre dos empresas de igual *r ating* crediticio. As , ‘Hoechst’ y ‘Electrab’ han sido consideradas como rango 2, pero los porcentajes de pertenencia parecen indicar que la valoraci n de la primera es superior a la segunda pues posee un mayor  ndice de pertenencia al rango 1. Como bien puede intuirse, el uso de t cnicas de l gica difusa o de an lisis intervalar podr an acabar de determinar la calificaci n final de la empresa.

- La regresi n ordinal es menos dependiente de los par metros relacionados con el aprendizaje de las m quinas como puede deducirse al observar que los porcentajes obtenidos poseen una menor desviaci n que en el caso de multclasificaci n. Para ilustrarlo ha podido comprobarse que la varianza de porcentajes en el caso de ordenaci n con n cleos gaussianos es de un 46.81% de promedio mientras que en el caso de multclasificaci n con id nticos n cleos resulta en 72.14% si el an lisis es realizado con todos los sectores en conjunto o puede alcanzar el 79.09% en el caso de an lisis sectorial diferenciado.
- Destacar finalmente que el estudio en ordenaci n emite unos datos que pueden entenderse como m s congruentes con el tipo de problema analizado. Las variaciones en los porcentajes entre rangos siguen una l nea con s lo una zona de m ximo, mientras que en los resultados de multclasificaci n pueden existir dos rangos con niveles de porcentaje similares aunque no sean rangos continuos. Por ejemplo los datos obtenidos sobre el patr n ‘Meredith’ en multclasificaci n con todos los sectores arrojan un empate entre los rangos 2 y 4 o los datos sobre ‘Bayer’ analizando en multclasificaci n sectorial arrojan picos de m ximo en los rangos 1 y 4. Este tipo de inestabilidades no se producen en ning n caso cuando se utiliza la formulaci n en ordenaci n aunque el tipo de n cleos utilizado es el mismo.

7.4.3 Resultados sobre test

Por motivos de completitud, se ha realizado tambi n una evaluaci n del aprendizaje de la nueva m quina en ordenaci n reservando parte de la informaci n como conjunto de test, aunque ello suponga perder informaci n cr tica. Se han realizado 50 experimentaciones diferentes extrayendo de forma aleatoria y sin considerar el rango al que pertenecen un 75% de los patrones del conjunto de aprendizaje original para elaborar el conjunto de aprendizaje modificado en pares y reservando el 25% para realizar la evaluaci n. El conjunto modificado \tilde{T} se construy  de nuevo considerando todas las combinaciones posibles entre elementos de igual rango y s lo una diferenciaci n de entre todas las posibles entre dos rangos distintos cualesquiera. Tras seleccionar el tipo de n cleo, una funci n gaussiana, y fijar una varianza σ e insensitividad δ

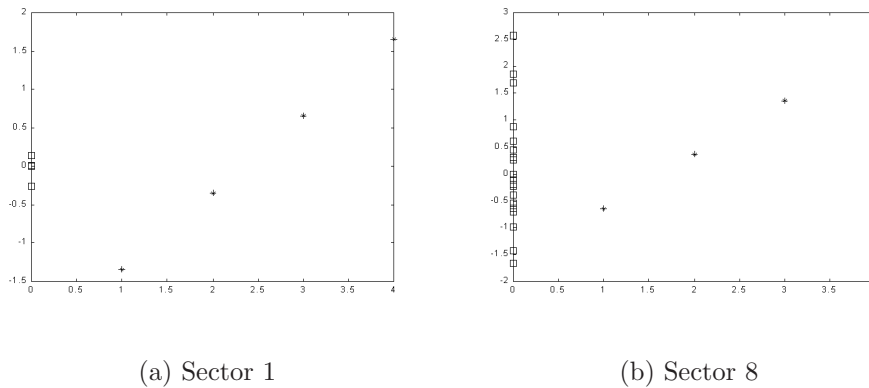


Figura 7.4: Resultados de la ordenación sobre dos sectores de producción utilizando una arquitectura de ordenación con máquinas K -SVCR.

sector	%error	SV	# patrones
sector 1	58.33%	141.04	559
sector 2	61.67%	139.60	601
sector 3	70.33%	49.30	121
sector 4	52.86%	61.76	195
sector 6	50.67%	39.60	127
sector 7	45.00%	77.56	507
sector 8	46.00%	19.96	37

Tabla 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.

única para todas las máquinas se procedió al entrenamiento. En la Figura 7.4 pueden observarse gráficamente algunos de los resultados del entrenamiento. Las marcas cuadradas situadas con valor de eje de abscisas 0 corresponden a las salidas de los patrones de test junto a los patrones sin rango conocido.

En la Tabla 7.5 se recogen los porcentajes de error sobre el conjunto de test. Tal como era de esperar el porcentaje de error cometido es muy elevado debido a la escasez de información disponible en la fase de entrenamiento, aunque en la mayoría de casos el nivel de error siempre está muy por debajo del 75% que se obtendría al trabajar sobre 4 clases en caso de no existir entrenamiento alguno.

7.5 En Resumen

Dentro de los problemas calificados de clasificación o categorización, existe un subtipo especial correspondiente a aquellos en los que las etiquetas están ordenadas, induciendo asimismo un orden de preferencia en el espacio de entrada. Este tipo de problemas debería consignarse como una clase especial pues recoge propiedades también de regresión, por lo que han sido denominados desde el Capítulo de Introducción como problemas de regresión ordinal u ordenación. Las máquinas que traten estos problemas deberían tener en consideración este orden de preferencia. La nueva máquina K -SVCR permite, mediante el uso de una función de coste ordinal (7.4) que recoge esta preferencia y una modificación del conjunto de entrenamiento en forma de pares de elementos (7.8), el diseño de una función utilidad que introduce los patrones de entrenamiento dentro de intervalos acotados de una recta que habita en el espacio de características.

Si bien el nuevo método tiene en cuenta la especial existencia de preferencia dentro del espacio de entrada debido a la ordenación del espacio de salida y reduce la dimensión del conjunto de entrenamiento respecto a otros estudios y aproximaciones, la principal desventaja del nuevo método se sigue centrando en el significativo incremento de la dimensionalidad del nuevo conjunto de patrones en forma de pares respecto al original, por lo que se hace particularmente atractivo en problemas que contengan un número reducido de datos. Estas condiciones las cumple el problema de riesgo crediticio analizado como aplicación y cuyos resultados sirven para mostrar la capacidad generalizadora del nuevo método, no necesitado de recurrir a la ingente cantidad de información utilizada usualmente por los expertos.

