

Capítulo 5

Métodos de Kernel y Máquinas de *vectores de soporte*

5.1. Introducción

Tal y apuntamos en el Apartado 1.3.2, las BOCs de nuestro estudio se corresponden con IgG presentes en el LCR de los pacientes. Estas IgG se sitúan de forma que cada posición de la BOC se corresponde con un tipo de proteína concreta.

Esta disposición de las IgG aporta sentido a la idea de que cada banda es independiente de las demás y hace que pierda peso la relación entre la acumulación de bandas en proteínas cercanas. Por tanto, podemos interpretar en nuestro análisis que cada banda es una característica. Así, podemos tomar la idea de Pontil y Veri en [19] y crear un vector de posiciones de cada una de las bandas y asignar a cada una de estas posiciones/características una dimensión de nuestro vector para estimar el lugar que ocupan dentro del grupo.

5.2. Máquinas de *vectores de soporte* (SVM)

En los últimos años el procesado de señal ha experimentado un fuerte avance gracias al uso de técnicas estadísticas avanzadas como los algoritmos basados en kernels. Este tipo de técnicas unidas a métodos de clasificación como las Máquinas de Vectores de Soporte (SVM), están aportando la suficiente solidez estadística para mejorar los métodos

de detección de señales en condiciones desfavorables.

En principio, los algoritmos basados en kernels, son versiones no lineales de un algoritmo lineal en el que la información es procesada previamente para transformarla a un espacio dimensional superior, en el que tan solo tenemos que ser capaces de realizar productos escalares entre los vectores de información.

Las SVM son un conjunto de métodos de aprendizaje automático usados en clasificación y regresión. La base de este conjunto de métodos es conseguir un hiperplano que separe de una forma más eficiente las muestras de un conjunto de datos. En nuestro caso nos referiremos exclusivamente a su uso para la clasificación.

5.3. Clasificación por vectores de soporte (SVC)

El objetivo de la SVC es conseguir un método de aprendizaje automático computacionalmente eficiente que permita la separación de hiperplanos en un espacio característico superior. Para ello, en todo momento se trabaja en un espacio vectorial denominado espacio característico y que viene definido por un producto vectorial de los patrones de entrada.

Para la generación de los hiperplanos se utilizan dos tipos de márgenes en función del grado de dificultad y de separación que tengan nuestras muestras:

- Hiperplanos de márgenes fijos (Hard Margin Clasifiers)
- Hiperplanos de márgenes variables (Soft Margin Clasifiers)

5.3.1. Separación de márgenes fijos

Este tipo de clasificación es la más simple que se puede tener en una SVM y tan sólo funciona para los datos que son separables linealmente en el espacio característico. La estrategia de este tipo de separación consiste en encontrar la distancia óptima entre los puntos del espacio característico que más cerca estén de la frontera que queremos hallar. Así pues, si observamos la Figura 5.1 podemos ver cómo esta distancia se corresponde a la que existe entre los círculos y las cruces más cercanos a sus respectivas línea de puntos. A éstos se les denomina Vectores de Soporte.

Si observamos detenidamente la Figura, vemos que la línea en la que se encuentran estos puntos puede expresarse matemáticamente como:

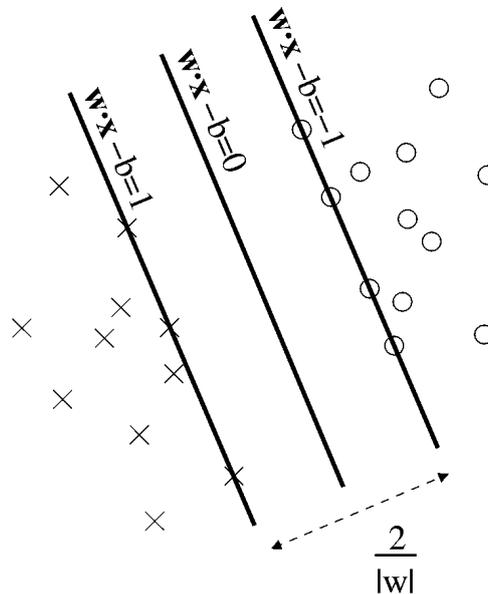


Figura 5.1: Espacio característico

$$\langle \mathbf{w}, \mathbf{x} \rangle - b = -1 \quad (5.1)$$

$$\langle \mathbf{w}, \mathbf{x} \rangle - b = 1 \quad (5.2)$$

Y por tanto, se puede demostrar que el problema de encontrar el separador óptimo para un conjunto de puntos con un margen $\gamma = \frac{1}{\|\mathbf{W}\|_2}$ viene de minimizar $\langle \mathbf{w} \cdot \mathbf{w} \rangle$ teniendo como restricción que $y_i (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) \geq 1$.

La resolución de este problema de optimización se aborda mediante el uso de multiplicadores de Lagrange [11]:

$$L_P = \frac{1}{2} \langle \mathbf{w} \cdot \mathbf{w} \rangle - \sum_{i=1}^l \alpha_i [y_i (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) - 1] \quad (5.3)$$

De la ecuación 5.3 nos interesa encontrar el \mathbf{w} y b que la minimizan y el α que la maximiza. Así pues diferenciando e igualando a cero:

$$\frac{\partial L_P}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^L \alpha_i y_i \mathbf{x}_i \quad (5.4)$$

$$\frac{\partial L_P}{\partial b} = 0 \sum_{i=1}^L \alpha_i y_i = 0 \quad (5.5)$$

Si sustituimos 5.4 y 5.5 en 5.3 obtenemos la formulación dual del problema que es independiente de α y que habrá que maximizar:

$$L_D \equiv \sum_{i=1}^L \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \iff \alpha_i \geq 0 \forall i, \sum_{i=1}^L \alpha_i y_i = 0 \quad (5.6)$$

$$L_D \equiv \sum_{i=1}^L \alpha_i - \frac{1}{2} \alpha^T \mathbf{H} \alpha \iff \alpha_i \geq 0 \forall i, \sum_{i=1}^L \alpha_i y_i = 0 \quad H_{ij} \equiv y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (5.7)$$

Esta nueva formulación tan solo requiere del producto vectorial de cada vector de entrada x_i para ser calculado.

Así pues, con estas modificaciones nuestro problema consiste ahora en maximizar L_D , por tanto:

$$\max_{\alpha} \left[\sum_{i=1}^L \alpha_i - \frac{1}{2} \alpha^T \mathbf{H} \alpha \right] \iff \alpha_i \geq 0 \forall i, \text{ y } \sum_{i=1}^L \alpha_i y_i = 0 \quad (5.8)$$

Este problema se identifica con un problema de optimización cuadrática convexa, del cual obtendremos α , y consecuentemente a través de la condición 5.4, obtendremos w .

5.4. Funcionamiento de una SVM

Como todo método de aprendizaje automático, las SVMs necesitan unos datos de entrenamiento que permitan determinar los vectores de soporte para establecer las regiones a las que asignar los grupos a clasificar.

En este entrenamiento se somete a los vectores de datos a las ecuaciones del apartado 5.3.1, y se obtienen la delimitación antes señalada. Tras este entrenamiento, se aplican los datos a la máquina de vectores que los clasifica en grupos según sus características.

Casos Entrenamiento	Número de Bandas	Convierte
M-1066	2	No
M-108	15	No
M-1215	13	No
M-1260	15	Sí
M-1296	10	Sí
M-157	5	No
M-176	9	Sí
M-183	8	No
M-287	8	No
M-429	12	No
M-45	10	Sí
M-571	8	Sí
M-653	3	No
M-88	5	Sí

Tabla 5.1: Casos utilizados para el entrenamiento de la SVM

5.5. Aplicación de SVM

Para aplicar esta técnica de procesado de señal a nuestros datos, se han utilizado los órdenes de Matlab[®] que nos permiten trabajar de una forma estándar y contrastada con SVM. Así pues, se han tomado los 27 casos y se ha creado un vector de posiciones por cada uno de ellos en el que se asignaba el valor lógico 1 a las posiciones que estaban ocupadas en las BOCs y el valor 0 en caso contrario.

Una vez creados los vectores se ha seleccionado aleatoriamente un grupo de 14 casos para entrenar a la máquina de vectores, que se muestran en la Tabla 5.1 y se han usado los 13 restantes para evaluar los resultados de la misma.

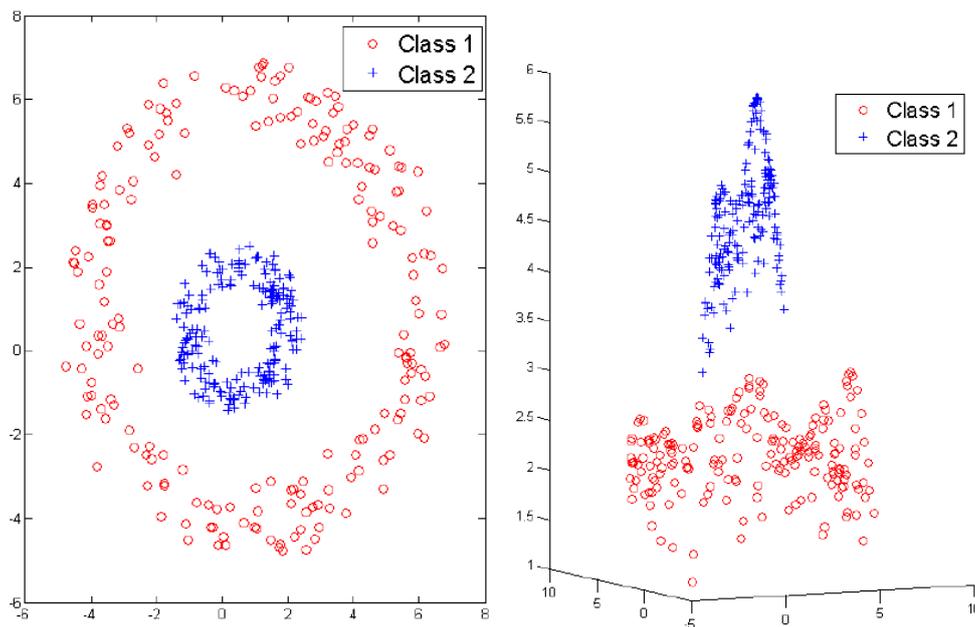
Como se aprecia en la Tabla 5.2, este análisis nos ha dado 9 aciertos en la clasificación, 1 falso positivo y 3 falsos negativos.

Caso	Número de Bandas	Convierte	Clasificación Automática	Acierto
M-1222	3	No	No	Sí
M-1279	17	No	No	Sí
M-128	9	No	Sí	No
M-140	7	No	No	Sí
M-333	13	Sí	No	No
M-382	11	Sí	No	No
M-570	7	No	No	Sí
M-584	7	No	No	Sí
M-593	6	Sí	Sí	Sí
M-622	9	Sí	No	No
M-659	6	No	No	Sí
M-70	4	No	No	Sí
M-813	4	Sí	Sí	Sí

Tabla 5.2: Resultados de la máquina de vectores con función de kernel lineal.

5.6. Funciones de kernel

A la hora de trabajar con las SVM, un recurso muy utilizado es el uso de funciones de *kernel* para mejorar la separación de los vectores a clasificar y así obtener un mejor resultado. La idea que subyace dentro de esta técnica es la de convertir los datos a un espacio vectorial de orden superior en donde sean fácilmente separables por un hiperplano, tal y como se puede apreciar en la Figura 5.2. Este recurso se denomina *truco del kernel*.



(a) Imagen de un conjunto de datos que no se pueden separar linealmente. (b) Imagen del conjunto de datos tras aplicarles la función de kernel gaussiana.

Figura 5.2: Conjunto de datos en su espacio original y en un espacio superior tras la aplicación de la función de kernel gaussiana. Se puede apreciar como la función de kernel ha añadido una dimensión más a los datos y los ha convertido en fácilmente separables.

5.6.1. Definición de kernel

Una función de kernel no es más que una función matemática que nos permite establecer una correspondencia entre nuestros datos en su espacio matemático de origen y un espacio matemático de orden superior.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle. \quad (5.9)$$

5.6.2. Tipos básicos de Kernel

Existen muchos tipos de funciones de kernel, pero en nuestro caso sólo hemos usado los tipos lineal, gaussiano y polinómico.

Kernels lineales

Este tipo de kernels consisten en la aplicación de un producto escalar de los vectores de entrada, tal y como se puede ver en la Ecuación 5.10. Este tipo de kernels es en realidad el que hemos usado en las secciones anteriores.

$$k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j = \mathbf{x}_i^T \mathbf{x}_j \quad (5.10)$$

Kernels polinómicos

Los Kernels polinómicos se definen como:

$$k(\mathbf{x}_i, \mathbf{x}_j) = p(k_1(\mathbf{x}_i, \mathbf{x}_j)) \quad (5.11)$$

o en su forma más común como:

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle + R)^d \quad (5.12)$$

El resultado de aplicar este tipo de kernel a nuestra función se puede apreciar en la Tabla 5.3. Como se puede observar en la columna de *Clasificación*, la máquina no es capaz de distinguir los casos positivos.

Kernels Gaussianos

Este tipo de kernels son los más usados en la actualidad y se definen como:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (5.13)$$

Donde el parámetro σ controla la flexibilidad del kernel.

Caso	Número de Bandas	Convierte	Clasificación Automática
M-1222	3	No	No
M-1279	17	No	No
M-128	9	No	No
M-140	7	No	No
M-333	13	Sí	No
M-382	11	Sí	No
M-570	7	No	No
M-584	7	No	No
M-593	6	Sí	No
M-622	9	Sí	No
M-659	6	No	No
M-70	4	No	No
M-813	4	Sí	No

Tabla 5.3: Resultados de la máquina de vectores con función de kernel polinómica.

Caso	Número de Bandas	Convierte	Clasificación Automática
M-1222	3	No	No
M-1279	17	No	No
M-128	9	No	No
M-140	7	No	No
M-333	13	Sí	No
M-382	11	Sí	No
M-570	7	No	No
M-584	7	No	No
M-593	6	Sí	No
M-622	9	Sí	No
M-659	6	No	No
M-70	4	No	No
M-813	4	Sí	No

Tabla 5.4: Resultados de la máquina de vectores con función de kernel gaussiana.

A pesar de lo extendido de su uso, podemos apreciar en la Tabla 5.4 que este tipo de función de kernel tampoco identifica bien los vectores y no consigue diferenciar entre los dos grupos.