**2.4.1. Máquinas de Vectores de Soportes**

Las Máquinas de Vectores de Soporte (SVM) abordan problemas de clasificación lineal, mediante la búsqueda de la superficie óptima de separación de dos clases. La superficie óptima, maximiza la distancia entre las dos clases, utilizando los datos más cercanos (vectores de soporte) para construir dicha superficie, como se describe en la Figura 2.3 (Vapnik, 1995). Cuando el problema es no lineal se realiza una proyección de los datos de entrada hacia un espacio llamado “espacio de características”, generalmente de alta dimensión. Siempre existe un espacio, por más alta que sea su dimensión, en el que es posible encontrar un hiperplano lineal que separa dos clases, como lo muestra la Figura 2.3

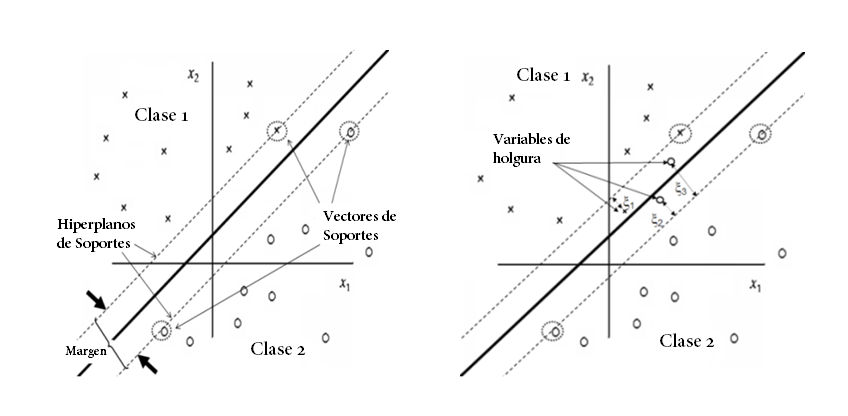


Figura 2.3 Hiperplano lineal de separación de dos clases: () Margen máximo () Margen blando. Los datos destacados con círculos son los vectores de soporte.

Los parámetros del hiperplano lineal (w, b) se obtienen mediante un algoritmo de optimización que encuentra la mayor distancia (margen) respecto de los vectores de soporte (Schölkopf et al. 2000). Se definen los hiperplanos de apoyo que se muestran en la Figura 2.3 (a). Por su parte, la Figura 2.3 () muestra la situación de margen blando, o sea cuando se permite que algunos puntos crucen los hiperplanos de soporte. Las variables de holgura ξ, son términos que indican en qué medida un punto se encuentra en el lado equivocado de su respectivo hiperplano de soporte. Para resolver problemas no lineales, se utiliza el “truco del kernel” (Schölkopf et al., 2000). La función de transformación ϕ, llamada kernel, proyecta el espacio de entrada en el espacio de características, como se puede observar en la Figura 2.4.

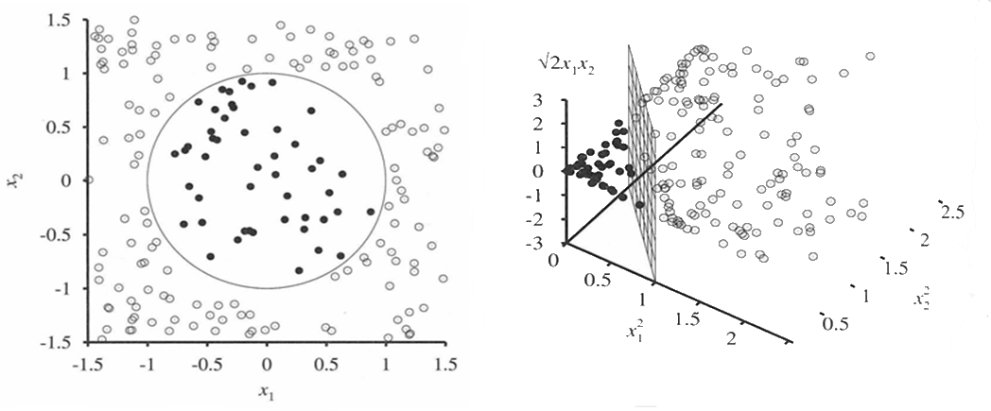


Figura 2.4 Efecto de la asignación del espacio de entrada en un espacio de características dimensional superior, donde es posible un plano de separación lineal.

Algunas de las funciones kernel permiten calcular el producto interno entre dos variables originales independientes  en el espacio de entrada como si fuera en el espacio de características, por lo que no es necesario evaluar el producto punto en el espacio de transformación. La *ecuación 2.1* muestra la expresión del kernel función Base Radial (Radial Base Function, RBF), que es la función kernel más utilizada en SVM.

**(2.1)**

Donde σ es el ancho de la función de base Radial (Gausseana) y es un hiperparámetro que es necesario ajustar al problema. El problema de optimización se expresa entonces como se muestra en la *ecuación 2.2.*

(2.2)

Donde c es un hiperparámetro que determina el equilibrio entre la complejidad del modelo y los puntos que quedan en el lado equivocado del hiperplano de decisión, expresados ​​por la variable de holgura ξ.

La solución al problema de optimización permite obtener los pesos de la superficie de decisión. La solución se encuentra aplicando el procedimiento de optimización estándar para un problema con restricciones de desigualdad sujeto a las condiciones de Karush-Kuhn-Tucker para el problema dual.

En las tareas de clasificación con kernel RBF, el diseño de las SVM consiste en encontrar los parámetros de la superficie de decisión (w,b) sujetos a los valores de dos hiperparámetros que es necesario ajustar: el parámetro RBF σ de la función kernel y el parámetro de penalización c del margen blando.

Estos parámetros se ajustan mediante un método llamado búsqueda de rejilla. En este método los valores de c y σ varían en potencias de 2 (c=2n  y σ=2m) siendo que n y m toman valores en rangos acotados (por ejemplo de [ -15:10] con un paso de uno). Esto permite hacer un barrido de los diferentes valores de c y σ para buscar el par de valores que optimizar el rendimiento del clasificador.

## OTRA DESCRIPCION

## Herramienta de Clasificación: Máquina de Vectores de Soporte

La herramienta a utilizar para llevar a cabo la clasificación lineal serán las Máquinas de Vectores de Soporte (SVM), las cuales buscan la superficie óptima de separación de dos clases (Vapnik,1995). La superficie óptima, maximiza la distancia entre las dos clases, utilizando los datos más cercanos (vectores de soporte) para construir dicha superficie. Cuando el problema es no lineal se realiza una proyección de los datos de entrada hacia un espacio llamado “espacio de características”, generalmente de alta dimensión. Siempre existe un espacio, por más alta que sea su dimensión, en el que es posible encontrar un hiperplano lineal que separa dos clases, como lo muestra la figura 1.

Los parámetros del hiperplano lineal (w, b) se obtienen mediante un algoritmo de optimización que encuentra la mayor distancia (margen) respecto de los vectores de soporte (Schölkopf et al. 2000). Las variables de holgura ξ, son términos que indican en qué medida un punto se encuentra en el lado equivocado de su respectivo hiperplano de soporte. Para resolver problemas no lineales, se utiliza el “truco del kernel” (Schölkopf et al., 2000). La función de transformación ϕ, llamada kernel, proyecta el espacio de entrada en el espacio de características, como se muestra en la figura 1.

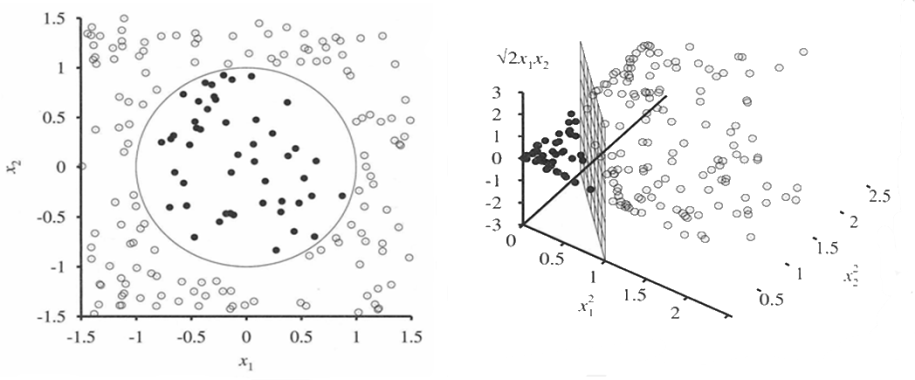


Figura. 1. Efecto de la asignación del espacio de entrada en un espacio de características dimensional superior, donde es posible un plano de separación lineal.

Uno de los hiperparámetros a utilizar es sigma (σ), que representa el ancho de la función de transformacion kernel (Base Radial). Al igual que sigma, otro hiperparámetro a utilizar es c, que determina el equilibrio entre la complejidad del modelo y los puntos que quedan en el lado equivocado del hiperplano de decisión, expresados ​​por la variable de holgura ξ. Es necesario ajustar estos parámetros al problema.

En las tareas de clasificación con kernel, el diseño de las SVM consiste en encontrar los parámetros de la superficie de decisión (w,b) sujetos a los valores de dos hiperparámetros que es necesario ajustar: el parámetro σ de la función kernel y el parámetro de penalización c.

Estos parámetros se ajustan mediante un método llamado búsqueda de rejilla. En este método los valores de c y σ varían en potencias de 2 (c=2n y σ=2m) siendo que n y m toman valores en rangos. Esto permite hacer un barrido de los diferentes valores de c y σ para buscar el par de valores que optimizar el rendimiento del clasificador. En nuestro caso, para seguir con la tónica de trabajos anteriores, utilizaremos los valores de n para c [-5:3:10], es decir, valores entre -5 y 10 de 3 en 3. Y para sigma utilizaremos [-15:3:9].