**CLASIFICACION 3 – SVM**

**Resumen de la Sección 9: Support Vector Machines**

**1. Introducción a las Support Vector Machines (SVM)**

**Support Vector Machines (SVM)** son métodos de aprendizaje supervisado utilizados principalmente para problemas de clasificación y, en menor medida, para regresión. Estos métodos buscan encontrar un **hiperplano óptimo** que divida los datos en clases distintas con el mayor margen posible.

**2. Clasificador de Margen Máximo (Maximal Margin Classifier)**

**2.1 Definición**

El **clasificador de margen máximo** es el modelo más simple dentro de las SVM. Funciona para problemas donde las clases son linealmente separables y busca maximizar la distancia (margen) entre los datos de las dos clases y el hiperplano separador.

* Un hiperplano en ppp-dimensiones se define como: β0+β1X1+β2X2+…+βpXp=0
* donde:
  + X1,X2,…,Xp: Predictores.
  + β0,β1,…,βp​: Coeficientes.

**2.2 Construcción**

1. Encuentra el hiperplano que maximiza el margen (distancia más pequeña entre el hiperplano y las observaciones más cercanas de cada clase).
2. Resuelve el problema de optimización

A black text with black text

Description automatically generated with medium confidence ​

sujeto a: A group of black letters

Description automatically generated

**3. Clasificador de Vectores de Soporte**

Cuando las clases **no son linealmente separables**, el clasificador de margen máximo no funciona bien. En estos casos, se usa el **Clasificador de Vectores de Soporte (SVC)**.

**3.1 Función de Costo**

Se introduce un término de penalización C para manejar violaciones al margen:

A number and a plus and six

Description automatically generated with medium confidence

sujeto a:



* ξi​: **Variables de holgura** (Slacks) que permiten violaciones al margen.
* C: Parámetro que controla la tolerancia a violaciones (mayor C implica menor tolerancia).

**4. Máquinas de Vectores de Soporte (SVM)**

Cuando las fronteras de decisión no son lineales, las SVM introducen un **kernel** para transformar los datos a un espacio de mayor dimensión, donde puedan separarse linealmente.

**4.1 Kernel**

Un kernel es una función que calcula un producto interno en un espacio de mayor dimensión, sin necesidad de calcular explícitamente las transformaciones. Los kernels comunes incluyen:

* **Lineal:** K(X,Z)=XtZ
* **Polinómico:** K(X,Z)=(1+XtZ)
* **RBF (Radial Basis Function):** K(X,Z)=exp(−γ∥X−Z∥2)

**4.2 Problema de Optimización**

En el espacio transformado, el problema de optimización se convierte en:

A black and white math symbols

Description automatically generated

sujeto a:

A black and white image of a mathematical equation

Description automatically generated

**Función de Transformación ϕ(x)\phi(x)ϕ(x)**

En situaciones donde los datos no son linealmente separables en su espacio original (espacio predictor), se utiliza una función de transformación ϕ(x) para proyectar los datos a un espacio de mayor dimensión, llamado **espacio de características** o **espacio transformado**. En este espacio, los datos tienen mayor probabilidad de ser linealmente separables.

**Transformación Implícita**

* + 1. **Directa:** La transformación explícita mapea x a un espacio de características H: ϕ(x):Rp→H

Por ejemplo:

* + x=[x1,x2]→ϕ(x)=[x1^2,x2^2,x1x2]
    1. **Implícita (mediante un kernel):** En lugar de calcular explícitamente ϕ(x)), las SVM utilizan funciones de kernel K(x,z), que computan directamente el producto interno en el espacio transformado: K(x,z)=ϕ(x)Tϕ(z)

Esto evita tener que trabajar directamente con ϕ(x), lo cual es computacionalmente costoso en espacios de alta dimensionalidad.

**El Teorema de Mercer**

El **Teorema de Mercer** proporciona una condición matemática que asegura que una función K(x,z) puede actuar como un kernel válido para SVM.

**Definición**

Una función K(x,z) es un kernel válido si satisface las siguientes propiedades:

1. **Simetría:** K(x,z)=K(z,x) ∀ x,z ∈ R^p
2. **Matriz semidefinida positiva:** Para cualquier conjunto de puntos {x1,x2,…,xn}, la matriz K con elementos K(xi,xj) debe ser semidefinida positiva. Es decir, para cualquier vector α, se cumple: A black text on a white background

   Description automatically generated

**Implicación Práctica**

El Teorema de Mercer asegura que las funciones K(x,z) que cumplen estas condiciones equivalen a un producto interno en algún espacio transformado H. Esto permite que las SVM trabajen con kernels sin conocer explícitamente ϕ(x).

**5. SVM para Problemas Multiclase**

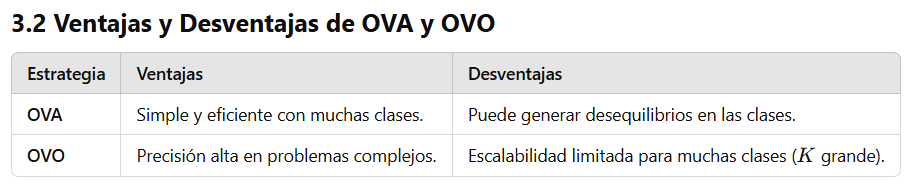
Para problemas con más de dos clases, se pueden usar estrategias como:

1. **One-Versus-One (OVO):** Construir un clasificador para cada par de clases.
2. **One-Versus-All (OVA):** Construir un clasificador para cada clase frente a las demás.

Las SVM tradicionales están diseñadas para problemas binarios (Y∈{−1,+1}). Para problemas con más de dos clases (Y∈{1,2,…,K} ), se utilizan estrategias para combinar múltiples clasificadores binarios.

**Estrategias Comunes**

1. **One-Versus-All (OVA):**
   * Se construye un clasificador binario para cada clase k, donde las observaciones de Y=k son la clase positiva (+1), y todas las demás clases son negativas (−1).
   * Para una nueva observación, se predice la clase con la mayor distancia al hiperplano (es decir, mayor ∣βTx+β0∣​∣).
2. **One-Versus-One (OVO):**
   * Se construye un clasificador binario para cada par de clases (Ck y Cl​).
   * Esto resulta en K(K−1)/2​ clasificadores.
   * La predicción se realiza usando un esquema de votación: cada clasificador binario vota por una clase, y la clase con más votos es la predicción final.



**Ejemplo Práctico**

Supongamos un problema con tres clases: Y ∈ {1,2,3} .

1. **OVA:** Se construyen tres clasificadores binarios:
   * Clasificador 1: Clase Y=1 versus Y≠1.
   * Clasificador 2: Clase Y=2 versus Y≠2.
   * Clasificador 3: Clase Y=3 versus Y≠3.

Si un punto nuevo tiene predicciones f1=1.2 , f2=0.8 , f3=0.6, se asigna a la clase 1, porque f1​ es el más alto.

1. **OVO:** Se construyen tres clasificadores:
   * Clasificador 1: Clase Y=1 versus Y=2.
   * Clasificador 2: Clase Y=1 versus Y=3.
   * Clasificador 3: Clase Y=2 versus Y=3.

Cada clasificador vota por una clase, y se asigna la clase con más votos.

**6. Relación con la Regresión Logística**

SVM y regresión logística son similares en que ambos buscan maximizar el margen, pero SVM es más robusta a valores atípicos debido al uso de funciones de pérdida basadas en margen.

**7. Resumen de Conceptos Clave y Fórmulas**

1. **Ecuación del hiperplano:** β0+β1X1+β2X2+…+βpXp=0
2. **Margen:** Margen=2∥β​∥
3. **Problema de optimización SVM (kernel):** 