



---

# TRABAJO PRÁCTICO NRO 2

---

MINERIA DE DATOS



FECHA DE ENTREGA: 15 DE NOVIEMBRE DE 2024

AUTOR: CADER FERNANDA

PROFESOR: REA OSCAR

Consignas: De acuerdo con las clases vistas:

1. ¿Qué se entiende por tarea y método en Minería de datos?
2. ¿Cuáles son las tareas más importantes que podemos encontrar en Minería de Datos?
3. ¿Cuál es problema de los modelos paramétricos? Ejemplifique.
4. Explique con sus palabras reglas de asociación
5. Como funciona las máquinas de vectores de soporte

Desarrollo

### 1. Tarea y Método en Minería de Datos

- **Tarea:** Una tarea en minería de datos se refiere al **tipo de problema que se intenta resolver**. Por ejemplo, clasificar productos, predecir ventas, o agrupar clientes en segmentos son ejemplos de tareas.
- **Método:** Un método, también llamado técnica o algoritmo, es la **herramienta específica que se utiliza para resolver una tarea**. Árboles de decisión, redes neuronales, y análisis de regresión son ejemplos de métodos.

### 2. Tareas Importantes en Minería de Datos

Las tareas en minería de datos se pueden clasificar en dos grandes categorías: **predictivas y descriptivas**.

**Tareas Predictivas:**

- **Clasificación:** Asignar elementos a clases predefinidas. Ejemplo: Clasificar correos electrónicos como spam o no spam.

- **Clasificación Suave:** Igual que la clasificación, pero proporciona una medida de certeza en la predicción. Ejemplo: Clasificar un tumor como benigno o maligno, incluyendo la probabilidad de que sea maligno.
- **Estimación de Probabilidad de Clasificación:** Determinar la probabilidad de que un elemento pertenezca a cada una de las clases posibles.
- **Categorización:** Asignar múltiples categorías a un elemento. Ejemplo: Categorizar un artículo de noticias en "política", "economía", y "deportes".
- **Preferencias o Priorización:** Establecer un orden de preferencia entre elementos. Ejemplo: Ordenar candidatos para un puesto de trabajo según su idoneidad.
- **Regresión:** Predecir un valor numérico. Ejemplo: Predecir el precio de una casa en función de sus características.

### **Tareas Descriptivas:**

- Agrupamiento (Clustering): Encontrar grupos de elementos similares sin tener clases predefinidas. Ejemplo: Segmentar clientes en grupos con hábitos de compra similares.
- Correlaciones y Factorizaciones: Identificar relaciones entre atributos. Ejemplo: Determinar qué atributos están correlacionados con la satisfacción del cliente.
- Reglas de Asociación: Descubrir patrones de co-ocurrencia entre elementos. Ejemplo: Identificar productos que se compran juntos con frecuencia en un supermercado.
- Dependencias Funcionales: Encontrar relaciones de causa y efecto entre atributos.
- Detección de Valores e Instancias Anómalas: Identificar valores o elementos que se desvían significativamente del patrón general. Ejemplo: Detectar transacciones fraudulentas en una tarjeta de crédito.

### 3. Problema de los Modelos Paramétricos

Los modelos paramétricos asumen que los datos se ajustan a una forma o distribución específica. El problema es que **si la forma asumida no se ajusta bien a los datos reales, el modelo no será preciso.**

**Ejemplo:** Si intentamos modelar la relación entre el número de habitaciones en una casa (room) y el porcentaje de población con bajo estatus social (lstat) usando una línea recta (regresión lineal), el modelo no capturará la verdadera relación, que es no lineal.

### 4. Reglas de Asociación

Las reglas de asociación buscan **relaciones entre elementos que ocurren juntos con frecuencia.** Se expresan en la forma "SI...ENTONCES...", donde "SI" representa una condición y "ENTONCES" representa una consecuencia.

**Ejemplo:** En un supermercado, una regla de asociación podría ser: "SI un cliente compra cerveza ENTONCES también compra papas fritas".

### 5. Funcionamiento de las Máquinas de Vectores de Soporte

Las máquinas de vectores de soporte (SVM) son un método de clasificación que busca encontrar el **hiperplano que mejor separa los datos en diferentes clases.** El hiperplano se elige de manera que **maximice el margen** entre las clases, lo que mejora la capacidad de generalización del modelo.

- Se utiliza una función **kernel** para transformar los datos a un espacio de mayor dimensionalidad donde la separación lineal sea posible.
- Se introduce el concepto de **margen blando** para permitir algunos errores de clasificación y evitar el sobreajuste.

- La elección de la función kernel adecuada es crucial para el éxito de las SVM en un problema específico.