

Apuntes # 1 - Clase del martes 20 de febrero

Richard Araya Padilla, c. 2020214235

Al inicio de la clase se realizó el Quiz #1 con las siguientes preguntas:

1. Si u y v son dos vectores colineales con magnitudes 5 y 6 respectivamente
Desarrolle: ¿cuál es el resultado del producto punto entre u y v ? 40pts
R/ 30
2. Si se sabe que el producto punto del vector u con él mismo es 25
Desarrolle ¿cuál es la magnitud L2? 40pts R/ 5
3. De acuerdo al pipeline de Machine Learning visto en clase, mencione la
etapa donde se crean nuevos features a partir de features existentes.
5pts R/ Feature engineering.
4. Describa la diferencia entre un parámetro y un hiper parámetro 15 pts.
R/ el hiper parámetro se configura por el usuario, el parámetro es aprendido por el modelo.

1. Algoritmos fundamentales

1.1. K-near

Uno de los algoritmos de **clasificación** más básicos.

1.1.1. Ventajas

- Es simple.
- Robusto a los datos. No le afecta el ruido.
- No importan los datos outliers. (Datos mal clasificados o mal ingresados).

1.1.2. Desventajas

- Alto costo computacional.
- Una sola instancia contra todos.
- Ineficiente, respuesta lenta.
- Necesita mucha memoria.
- Complejidad de $O(n) * d$.
- Cambia mucho de comportamiento dependiendo del k elegido.

1.2. Regresión Lineal

Es una técnica de modelado estadístico que se utiliza para describir una variable de respuesta continua como una función de una o varias variables predictoras.

Dado un conjunto de tamaño N de la forma $\{x_i, y_i\}$ con x_i un vector d -dimensional y $y_i \in \mathbb{R}$. Se formula la ecuación de la regresión lineal de la siguiente manera:

$$wx + b = \hat{y}$$

Con w la pendiente y b el bias.

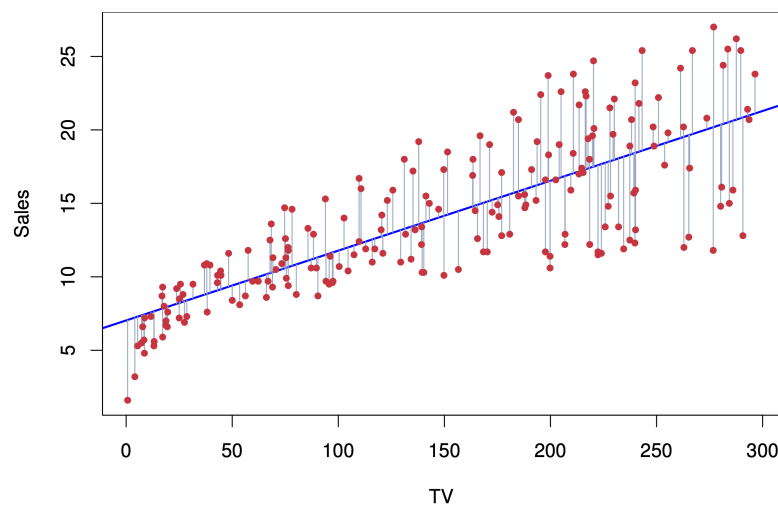


Figura 1: Gráfica de la función de regresión lineal sobre un conjunto de datos.

Este modelo se tiene que ajustar según los datos para lograr una mejor predicción. Para ello se buscan los valores óptimos de w y b . Es importante notar que el que el modelo se optimice, no significa que sea perfecto, lo que se busca reducir el error.

Para ello se busca minimizar la siguiente expresión Median Squared Error (MSE), en español Error Cuadrático Medio.

$$L = \frac{1}{N} \sum f_{w,b} (x_i - y_i)^2$$

En resumen:

Loss function: $f_{w,b} (x_i - y_i)^2$

Cost function: $\frac{1}{N} \sum f_{w,b} (x_i - y_i)^2$. Promedio de Lost Function.

¿Por qué se quiere minimizar?

Si L es grande, significa que los datos no se ajustan al modelo. Si L es pequeño, puede significar que hay un overfit o datos muy buenos.

Para hallar los valores de los parámetros hay que tomar en cuenta si la función es convexa o no convexa. Las funciones convexas son derivables, por lo que podemos utilizar la derivación como técnica para encontrar el mínimo de la MSE.

Una función conexa tiene solamente un mínimo global, mientras que una función no convexa tiene varios mínimos locales.

1.2.1. Derivadas

Razón de cambio de una función con respecto a su variable independiente.

1.2.2. Derivadas parciales

Es la generalización de las derivadas para más de una variable independiente. Se toma una variable como la independiente y las otras como constantes.