Apuntes Semana 4 - 11/03/2025 **Sebastián Vargas Mesén.**

I. Abstract

Se inicia con un repaso del algoritmo KNN (K-Nearest Neighbors), se analiza en profundidad la regresión lineal, tanto simple como múltiple, incluyendo su formulación matemática y el tratamiento de outliers mediante técnicas como el rango intercuartílico (IQR). Se examinan las funciones de pérdida y costo utilizadas para evaluar el rendimiento de modelos predictivos, centrándose en el error cuadrático. Se incluyen conceptos de derivadas, descenso del gradiente y derivadas parciales, mínimos locales y globales en funciones convexas y no-convexas.

II. Repaso K-Nearest Neighbors (KNN)

Es un algoritmo de clasificación y regresión que asigna a un nuevo dato la clase o valor predominante entre sus K vecinos más cercanos en el espacio de características. Utiliza la distancia L2 (Euclidiana) y requiere seleccionar el hiper parámetro K, que determina cuántos vecinos considerar.

Ventajas	Desventajas
Sencillo de implementar	No es rápido (computacionalmente costoso)
No requiere entrenamiento específico	Requiere tener los datos cargados en memoria
Intuitivo y fácil de entender	Poco eficiente con volúmenes de datos grandes
Eficaz para conjuntos de datos pequeños	Sensible a la dimensionalidad de los datos

III. Regresión lineal.

Es un algoritmo supervisado utilizado para predecir una variable continua basándose en la relación lineal con una o más variables independientes.

Desde una perspectiva estadística, la regresión lineal establece una relación entre una variable dependiente y un conjunto de variables independientes.

Buscamos determinar la importancia (weights) de cada feature en la predicción.

Implementación genérica:

$$f_{w,b}(x) = w \cdot x + b$$

Donde:

- x es el vector de características (features)
- w es el vector de pesos (weights)
- b es el término de sesgo (bias)
- $f_{w,b}(x)$ es la predicción resultante

El objetivo es encontrar los valores óptimos de w y b que minimicen el error entre las predicciones y los valores reales.

A. Tratamiento de Outliers

Los outliers son datos sobresalientes o atípicos que se desvían significativamente del resto de la distribución de datos. Estos valores pueden afectar negativamente la precisión de los modelos predictivos, especialmente en algoritmos sensibles a valores extremos.

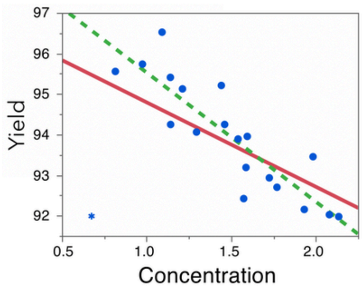


Figura 1. Alteración de predicción por outliers.

Técnicas de eliminación de outliers:

- Preparación de los datos.
- Selección de features.
- Rango intercuartílico (IQR).

El método del rango intercuartílico (IQR) es particularmente efectivo para la detección y eliminación de outliers.

1) Regresión lineal simple

La regresión lineal simple modela la relación entre una variable independiente (x) y una variable dependiente (y) usando una línea recta.

Formulación matemática: $y = mx + b$
Donde:

- y es la variable dependiente.
- x es la variable independiente
- m es la pendiente (tasa de cambio)
- b es el intercepto (valor de y con x = 0)

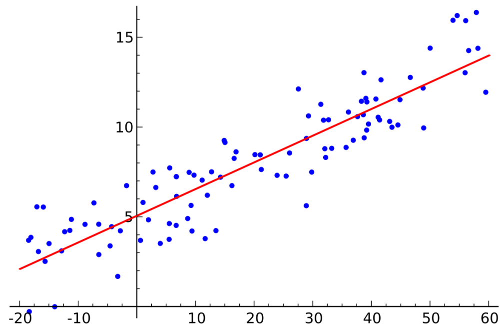


Figura 2. Ejemplo de recta de regresión lineal.

2) Regresión Lineal Múltiple

La regresión lineal múltiple extiende el concepto a múltiples variables independientes ($x_1, x_2, ..., x_n$).

Formulación matemática:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \varepsilon$

Donde:

- Y es la variable dependiente
- $X_1, X_2, ..., X_n$ son las variables independientes
- $\beta_0, \beta_1, \beta_2, ..., \beta_n$ son los coeficientes del modelo
- ε es el término de error

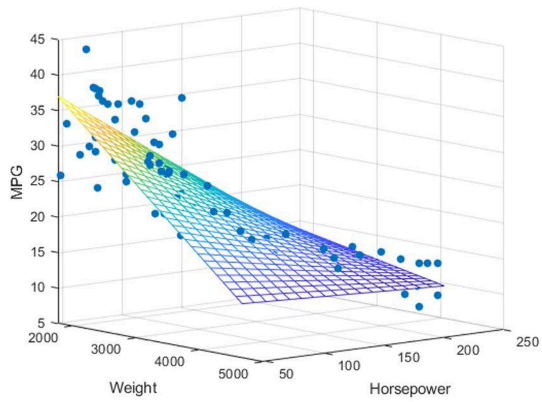


Figura 3. Ejemplo de regresión lineal múltiple.

IV. Optimización de modelos

A. Funciones de Pérdida y Costo

Las funciones de pérdida (loss functions) miden la discrepancia entre las predicciones del modelo y los valores reales. Para cada ejemplo individual, la función de pérdida cuantifica cuán lejos está la predicción del valor verdadero.

En regresión lineal, una función de pérdida común es el error cuadrático:

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

Donde:

- $f_{w,b}(x_i)$ es la predicción para el ejemplo i
- y_i es el valor real para el ejemplo i

Esta medida penaliza más los errores grandes que los pequeños, debido a la operación de elevación al cuadrado.

La función de costo (cost function) es el promedio de las funciones de pérdida para todos los ejemplos en el conjunto de datos:

L = $\frac{1}{n}\Sigma(f_{w,b}(x_i) - y_i)^2$, i=1,...,n

Donde n es el número total de ejemplos en el conjunto de datos.

B. Minimización del Error

El objetivo principal en el entrenamiento de modelos es encontrar los valores de parámetros (w y b) que minimicen la función de costo.
Interpretación:

- L grande → malas predicciones
- L pequeño → buenas predicciones

La minimización se logra generalmente mediante algoritmos de optimización que ajustan iterativamente los parámetros en la dirección que reduce el error. Uno de los algoritmos más utilizados es el descenso del gradiente.

V. CÁLCULO PARA OPTIMIZACIÓN

A. Fundamentos de Derivadas

Las derivadas son herramientas fundamentales del cálculo que miden la tasa de cambio de una función. En el contexto de la optimización de modelos, las derivadas indican la dirección y magnitud del ajuste necesario para los parámetros.

Propiedades importantes:

- La derivada otorga información sobre la pendiente y dirección
- En un punto mínimo, la derivada es igual a cero
- El signo de la derivada indica si la función está creciendo (positiva) o decreciendo (negativa)

B. Descenso del Gradiente

El descenso del gradiente es un algoritmo iterativo de optimización para encontrar el mínimo de una función. Funciona actualizando repetidamente los parámetros en la dirección opuesta al gradiente de la función de costo.
Proceso básico:

1. Inicializar los parámetros con valores aleatorios
2. Calcular el gradiente de la función de costo respecto a los parámetros
3. Actualizar los parámetros en dirección opuesta al gradiente

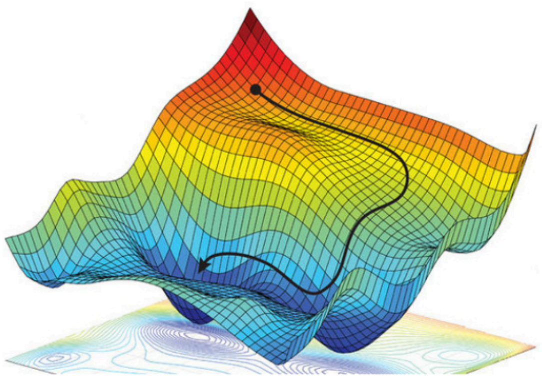
4. Repetir los pasos 2 y 3 hasta la convergencia

Fórmula de actualización:

$w = w - \alpha \nabla L(w)$

Donde:

- w son los parámetros
- α es la tasa de aprendizaje (learning rate)
- $\nabla L(w)$ es el gradiente de la función de costo



C. Derivadas Parciales

Las derivadas parciales son extensiones del concepto de derivada a funciones de múltiples variables. En estas funciones, podemos calcular la tasa de cambio con respecto a una variable manteniendo las demás constantes.

Notación: $\frac{\delta f}{\delta x}$ representa la derivada parcial de f con respecto a x.

Las derivadas parciales permiten calcular cómo cambiar cada parámetro individualmente para reducir el error.

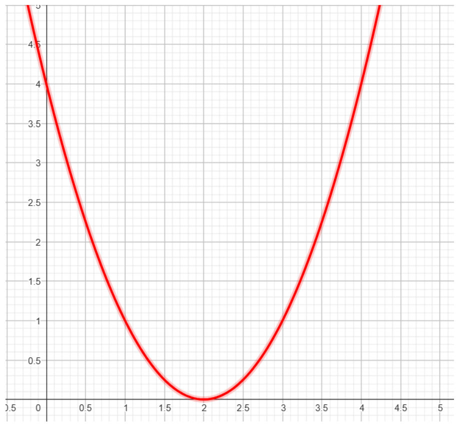
D. Mínimos Locales vs. Globales

Un desafío importante en la optimización es distinguir entre mínimos locales y globales:

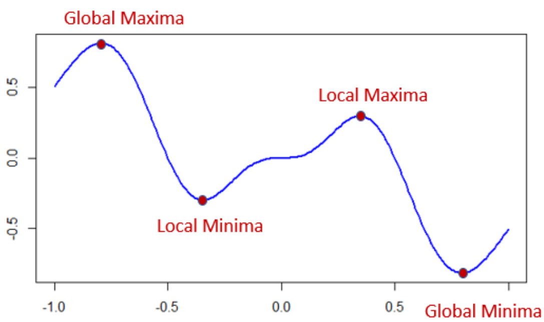
- Mínimo local: un punto donde la función tiene un valor menor que en todos los puntos cercanos, pero no necesariamente en todos los puntos del dominio
- Mínimo global: el punto con el valor más bajo de la función en todo su dominio

Funciones convexas vs. no-convexas:

- Funciones convexas: tienen un único mínimo, que es tanto local como global



- Funciones no-convexas: pueden tener múltiples mínimos locales, lo que complica la optimización



En funciones convexas, como la función de costo de la regresión lineal, el descenso del gradiente garantiza encontrar el mínimo global. Sin embargo, en funciones no-convexas, como las que aparecen en redes neuronales profundas, puede quedar atrapado en mínimos locales.

VI. RESPUESTAS A PREGUNTAS DEL QUIZ

1. Si u y v son dos vectores colineales con magnitudes 5 y 6 respectivamente, ¿cuál es el resultado del producto punto entre u y v?

Respuesta: Cuando dos vectores son colineales, forman un ángulo de 0° entre ellos. El producto punto entre dos vectores se define como:

$u \cdot v = |u| \cdot |v| \cdot \cos(\theta)$

donde θ es el ángulo entre los vectores.

Dado que son colineales ($\theta = 0°$) y $\cos(0°) = 1$:

$u \cdot v = |u| \cdot |v| \cdot \cos(0°) = 5 \cdot 6 \cdot 1 = 30$

Por lo tanto, el resultado del producto punto es 30.

2. Se sabe que el producto punto del vector u consigo mismo es 25, ¿cuál es su magnitud?

Respuesta: El producto punto de un vector consigo mismo está relacionado con su magnitud de la siguiente manera:

$u \cdot u = |u|^2$

Se nos da que $u \cdot u = 25$, por lo tanto:

$|u|^2 = 25$
$|u| = \sqrt{25} = 5$

La magnitud del vector u es 5.

3. Descripción de Supervised, Unsupervised y Reinforcement Learning.

Respuesta:

Supervised Learning (Aprendizaje Supervisado): Es un tipo de aprendizaje automático donde el algoritmo aprende a partir de un conjunto de datos etiquetados. Se proporciona al algoritmo ejemplos de entradas junto con sus salidas correspondientes (etiquetas), y el objetivo es aprender un mapeo general que pueda luego predecir la salida para nuevas entradas.

Unsupervised Learning (Aprendizaje No Supervisado): En este tipo de aprendizaje, el algoritmo trabaja con datos no etiquetados y debe encontrar patrones o estructuras por sí mismo. No hay respuestas correctas predefinidas y el sistema debe descubrir la organización intrínseca de los datos.

Reinforcement Learning (Aprendizaje por Refuerzo):. Un agente recibe retroalimentación en forma de recompensas o penalizaciones, y su objetivo es aprender una estrategia (política) que maximice la recompensa acumulada a largo plazo. No hay etiquetas explícitas como en el aprendizaje supervisado, sino señales de recompensa que pueden estar retrasadas en el tiempo respecto a las acciones que las causaron.

4. Descripción de la técnica de Grid Search.

Respuesta: Grid Search es una técnica de optimización de hiper parámetros que consiste en una búsqueda exhaustiva a través de un subconjunto especificado manualmente del espacio de hiper parámetros de un algoritmo de aprendizaje.

VIII. APÉNDICE: TABLA DE DERIVADAS

A. Funciones Básicas

k: constante; x: variable

Función f(x)	Derivada f'(x)	Ejemplo
k	0	f(x) = 7, f'(x) = 0
kx	k	f(x) = 5x, f'(x) = 5
x^2	2x	f(x) = x^2, f'(x) = 2x
x^n	n·x^(n-1)	f(x) = x^3, f'(x) = $3x^2$
e^x	e^x	f(x) = e^x, f'(x) = e^x
ln(x)	1/x	f(x) = ln(x), f'(x) = 1/x
sin(x)	cos(x)	f(x) = sin(x), f'(x) = cos(x)
cos(x)	-sin(x)	f(x) = cos(x), f'(x) = -sin(x)
x	1	f(x) = x, f'(x) = 1

B. Reglas de Derivación

Regla	Función f(x)	Derivada f'(x)
Regla de la suma	u(x) + v(x)	u'(x) + v'(x)
Regla del producto	u(x)·v(x)	u'(x)·v(x) + u(x)·v'(x)
Regla del cociente	u(x)/v(x)	$\frac{[u'(x)\cdot v(x) - u(x)\cdot v'(x)]}{[v(x)]^2}$
Regla de la cadena	f(g(x))	f'(g(x))·g'(x)

Ejemplos

f(x) = x^2 + 3x, f'(x) = 2x + 3

f(x) = x·2x, f'(x) = 1·2x + x·2 = 4x

f(x) = x/2x, f'(x) = [1·2x - x·2]/$[2x]^2$ = 0

f(x) = $(2x)^2$, f'(x) = 2(2x)·2 = 8x