Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ciencias y Sistemas
Inteligencia Artificial 1
Sección A

Algoritmos

Juan Pablo García Ceballos
201901598

Lugar y Fecha: Guatemala, Sacatepéquez 31/10/2024

REGRESIÓN LINEAL

La **regresión lineal** es un modelo estadístico que permite analizar cómo una variable independiente (X) afecta a una variable dependiente (Y).

- **X**: Variable independiente o exógena.
- **Y**: Variable dependiente o endógena.

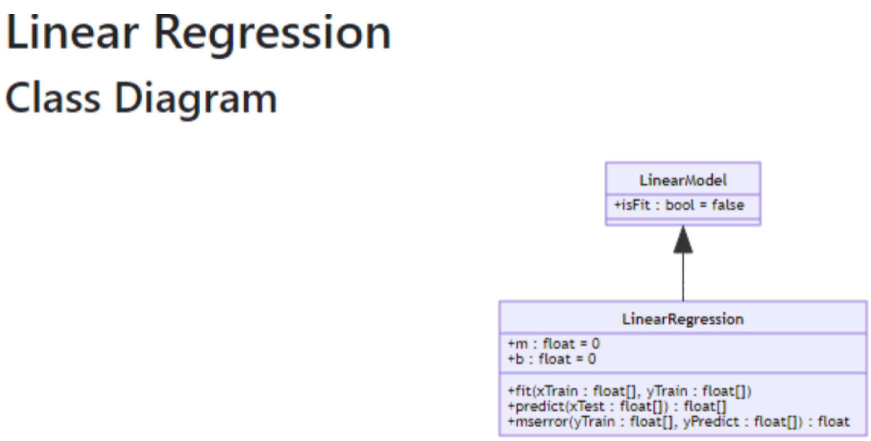
El propósito de este modelo es proporcionar estimaciones razonables para Y en función de distintos valores de X, usando una muestra de pares de datos (x_1, y_1), …, (xn, yn).

Ejemplos prácticos de aplicación de la regresión lineal:

- Estudiar cómo la altura del padre puede influir en la altura de los hijos.
- Calcular el precio de una vivienda basado en su tamaño.
- Relacionar la tasa de mortalidad con el consumo medio de cigarrillos.
- Analizar la abundancia de un parásito según la temperatura.
- Aproximar la nota en una materia a partir del tiempo semanal de estudio.
- Estimar el tiempo de ejecución de un programa según la velocidad del procesador.

Objetivos al aprender y comprender la regresión lineal:

- Construir un modelo de regresión lineal simple para describir cómo X influye en Y.
- Obtener estimaciones puntuales de los parámetros de un modelo de regresión.
- Calcular el valor promedio de Y para cualquier valor dado de X.
- Predecir valores futuros de la variable dependiente Y.



REGRESIÓN POLINOMIAL

La **regresión polinomial** es un método de predicción para una variable de respuesta cuantitativa en función de una variable predictora cuantitativa, donde la relación entre ellas se representa mediante una función polinomial de grado nnn.

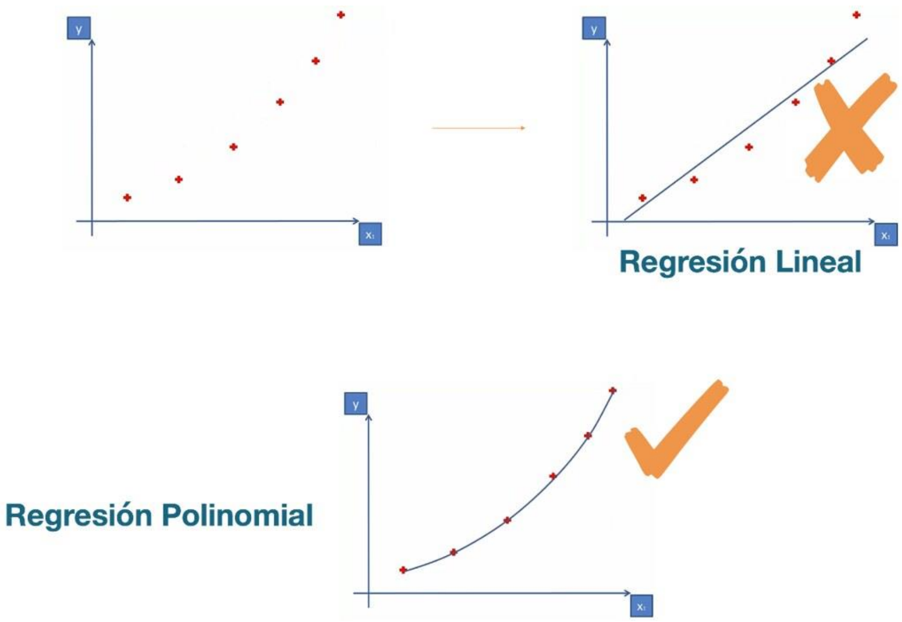
A diferencia de la regresión lineal, este modelo permite captar relaciones más complejas entre las variables mediante términos elevados a potencias mayores que uno (por ejemplo, x^2, x^3, etc.). Para calcular la regresión polinomial, se suelen emplear métodos de mínimos cuadrados, resolviendo incógnitas relacionadas con errores como el error absoluto y el error cuadrático. Esto implica definir una función objetivo adecuada, y se puede implementar en un lenguaje de programación de libre elección para mostrar los polinomios de mínimos cuadrados y sus valores de aproximación.

Objetivo: El objetivo es encontrar una función polinómica que se ajuste lo mejor posible a los datos proporcionados, de modo que la curva generada represente bien el comportamiento de la variable de respuesta. En este contexto, el error absoluto se define como la suma de las diferencias al cuadrado entre la función modelada y la función polinómica ajustada, aplicando este cálculo a los valores de la variable independiente.

Ecuación de la regresión polinomial:

$$y=a+bx+cx^2+dx^3+\dots$$

Aquí, los coeficientes a,b,c,d,a, b, c, d,a,b,c,d, etc., representan los pesos de cada término de xxx en el polinomio, que se ajustan para minimizar el error y proporcionar la mejor aproximación posible.



ÁRBOLES DE DECISIÓN

Los **árboles de decisión** son herramientas visuales que representan posibles resultados de una serie de decisiones interrelacionadas, permitiendo comparar opciones según sus costos, probabilidades y beneficios. Este modelo facilita tanto la toma de decisiones en organizaciones como la creación de algoritmos que matemáticamente determinen la opción más favorable.

Un árbol de decisión suele comenzar con un **nodo inicial** y se ramifica en múltiples resultados posibles. Cada resultado puede llevar a otros nodos adicionales que continúan ramificándose, dándole una estructura similar a la de un árbol.

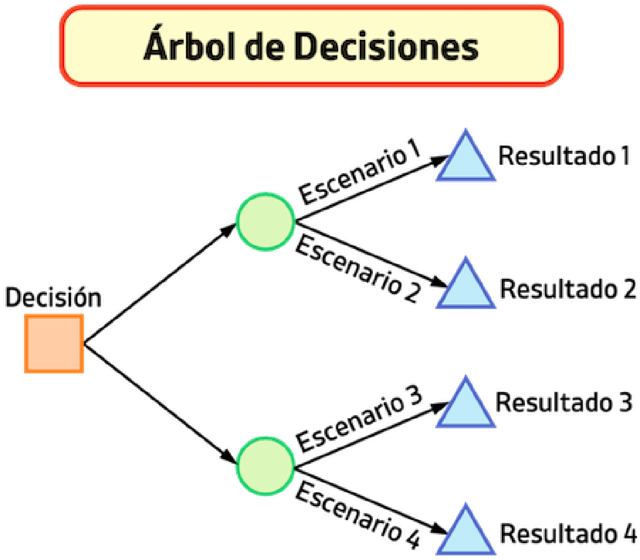
Tipos de Nodos en Árboles de Decisión

1. **Nodo de probabilidad**: Representado con un **círculo** y utilizado para indicar las probabilidades de diferentes resultados en un punto específico.
2. **Nodo de decisión**: Representado con un **cuadrado** e indica un punto en el que se toma una decisión.
3. **Nodo terminal**: Representado al final de cada ruta, muestra el **resultado final** de una secuencia de decisiones.

Simbología de los Árboles de Decisión

- **Círculo**: Nodo de probabilidad.
- **Cuadrado**: Nodo de decisión.
- **Triángulo o Nodo Terminal**: Representa el fin de una rama con el resultado final del análisis o decisión.

Este enfoque visual y estructurado ayuda a evaluar decisiones complejas, presentando todas las alternativas y consecuencias posibles en un formato claro y fácil de seguir.



K-means

El **algoritmo K-means** es un método de agrupamiento no supervisado que organiza datos en **K grupos o clústeres** en función de sus características, de modo que los datos dentro de cada grupo sean lo más similares posible entre sí, y lo más diferentes posibles de los datos de otros grupos. Es ampliamente utilizado en análisis de datos y aprendizaje automático para identificar patrones en conjuntos de datos grandes y sin etiquetar.

Funcionamiento del Algoritmo K-means

1. **Selección del número de clústeres (K)**: Se define el número de grupos en los que se quiere dividir el conjunto de datos.
2. **Inicialización de centroides**: Se eligen aleatoriamente K puntos en el espacio de datos, que actuarán como los centroides iniciales de cada clúster.
3. **Asignación de puntos a clústeres**: Cada punto de datos se asigna al clúster cuyo centroide esté más cercano, basándose en la distancia (generalmente la distancia euclidiana).
4. **Reubicación de centroides**: Después de asignar todos los puntos, se recalcula el centroide de cada clúster como el promedio de todos los puntos asignados a él.
5. **Iteración**: Los pasos de asignación y reubicación se repiten hasta que los centroides de los clústeres ya no cambien significativamente o hasta alcanzar un número máximo de iteraciones.

Objetivo

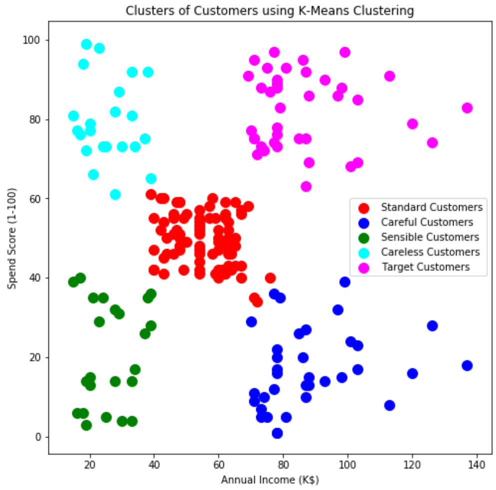
Minimizar la suma de las distancias cuadradas entre los puntos y su centroide asignado en cada clúster, también conocida como **inercia** o **suma de los errores al cuadrado** (SSE). Esto hace que los puntos dentro de cada grupo estén lo más cerca posible del centroide del grupo.

Aplicaciones de K-means

- Segmentación de clientes en marketing.
- Agrupamiento de imágenes en categorías.
- Análisis de patrones de comportamiento en redes sociales.
- Compresión de datos mediante reducción de dimensionalidad.

Consideraciones

La elección de K y la inicialización de los centroides son críticos para el rendimiento del algoritmo, y pueden afectar la estabilidad y precisión del agrupamiento.



Teorema de Bayes

El **Teorema de Bayes** permite calcular la probabilidad de que ocurra un evento A, dado que se tiene información sobre otro evento B que afecta o condiciona la probabilidad de A. Es una herramienta fundamental en estadística y probabilidad, ya que permite actualizar las probabilidades a medida que se obtiene nueva información.

A diferencia del **Teorema de la Probabilidad Total**, que evalúa la probabilidad de B a partir de varios eventos A, el Teorema de Bayes calcula la probabilidad de A dado que B ha ocurrido, es decir, la probabilidad de A condicionada a B.

Fórmula del Teorema de Bayes

La fórmula del Teorema de Bayes se expresa matemáticamente como:

$$P\ (A/B) = \frac{P(B/A) * P(A)}{P(B)}$$

donde:

- $P(A|B)$ es la probabilidad de que ocurra A dado que B ha ocurrido.
- $P(B|A)$ es la probabilidad de que ocurra B dado que A ha ocurrido.
- $P(A)$ es la probabilidad a priori de A (la probabilidad de A antes de saber que B ocurrió).
- $P(B)$ es la probabilidad total de B, calculada como la suma ponderada de $P(B|A) \cdot P(A)$ para todos los posibles eventos A.