UANL
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas

FCFM
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

Regresión Lineal

Equipo 3 (Grupo 003 Lunes):

Jennifer Priscila de León Flores	1860533
Enrique Alejandro Gallegos Luna	1858788
Mario Alberto Rodríguez Morales	1860043
Luis Angel Ruiz Ramírez	1862717

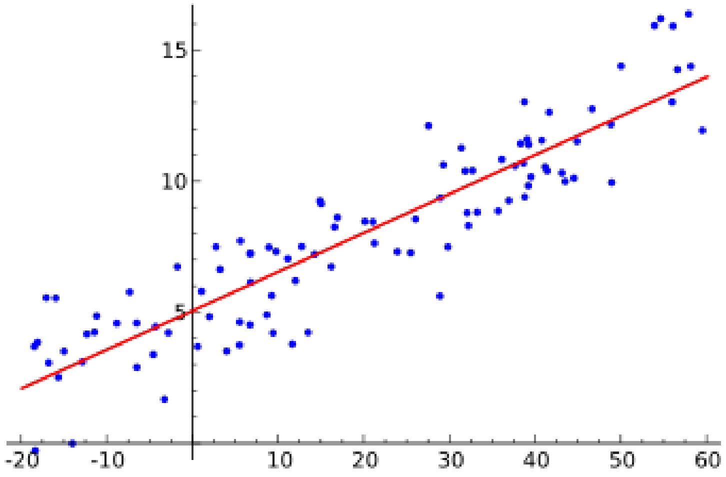
Regresión Lineal

La regresión lineal es un campo de estudio que enfatiza la relación estadística entre dos variables continuas conocidas como variables de predicción y respuesta.

- La variable predictora se denota con mayor frecuencia como X y también se conoce como variable independiente.
- La variable de respuesta se denota con mayor frecuencia como Y y también se conoce como variable dependiente.

Existen diferentes tipos de regresión lineal que se clasifican de acuerdo a sus parámetros:

- Cuando hay solo una variable predictora se conoce como regresión lineal simple.
- Cuando hay más de una variable predictora se conoce como regresión lineal múltiple.



Regresión Lineal

La regresión lineal tiene diversas métricas, una métrica es una función que define una distancia entre cada par de elementos de un conjunto. Entre las que utilizaremos se encuentran:

- Error Cuadrático Medio: Representa a la raíz cuadrada de la distancia cuadrada promedio entre el valor real y el valor pronosticado.

$$ECM = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2}$$

- Coeficiente de Determinación: También denominado como R^2, es un estadístico usado en el contexto de un modelo estadístico cuyo principal propósito es predecir futuros resultados o probar una hipótesis.

$$R^2 = \frac{\sum_{t=1}^{T}\left(\hat{Y}_t - \bar{Y}\right)^2}{\sum_{t=1}^{T}(Y_t - \bar{Y})^2}$$

Regresión Lineal Simple

Este modelo sólo está conformado por dos variables estadísticas llamadas X y Y. Para la regresión lineal simple, se asume que X y Y se relacionan mediante la relación funcional:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Donde Y es la variable de respuesta, X es la variable predictora, β_0 y β_1 son constantes desconocidas llamadas coeficientes de regresión y ε es un valor aleatorio. En estadística, β_0 y β_1 pueden estimarse como:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} \qquad \widehat{\beta_0} = \bar{y} - \hat{\beta}_1 \bar{x}$$

Luego el modelo ajustado de regresión lineal simple es:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

Regresión Lineal Múltiple

La regresión lineal permite analizar la relación entre dos o más variables a través de ecuaciones, lo que se denomina regresión múltiple o regresión lineal múltiple. Este modelo cuenta con varias variables predictoras, por lo que cuenta con varios parámetros, para la regresión lineal múltiple, se asume que la variable de respuesta Y se relaciona con las variables predictoras $X_0, X_1, \ldots, X_n$ mediante una relación funcional, en la cual sus parámetros se pueden estimar como:

$$\begin{pmatrix} y_1 & x_{11} & \cdots & x_{1p} \\ y_2 & x_{21} & & x_{2p} \\ \vdots & & \ddots & \vdots \\ y_n & y_{n1} & \cdots & x_{np} \end{pmatrix} \Rightarrow \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & & x_{2p} \\ \vdots & & \ddots & \vdots \\ 1 & y_{n1} & \cdots & x_{np} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix} \Rightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$$

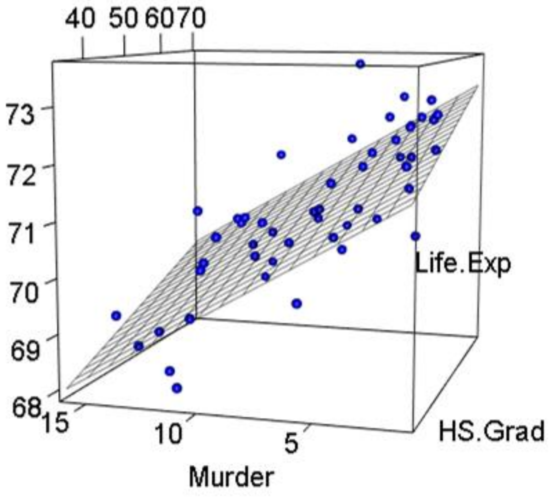
Entonces el modelo ajustado de regresión lineal múltiple, es la relación funcional:

$$\hat{Y} = \hat{\beta}_0 + \sum_{i=0}^{m} \hat{\beta}_i X_i$$

Regresión Lineal en Minería de Datos

- Permite predecir a partir de un muestreo aleatorio y se adapta a una amplia variedad de situaciones.
- La precisión de la regresión es directamente proporcional al numero de datos disponibles.

Por ejemplo, una constructora desea determinar el costo adecuado para los departamentos que va a construir y vender en una determinada zona. Para ello la constructora se puede basar en el costo de viviendas cercanas a la zona. El modelo ayudaría a predecir el costo factible de sus nuevos departamentos.



Regresión Lineal en Python

Se utilizan diversas librerías de las cuales se utilizan diversas funciones, una de ellas es la librería sklearn la cual incluye varios algoritmos de clasificación, regresión y análisis de grupos, entre las funciones a utilizar de esta se encuentran las siguientes:

- linear_model: Implementa una variedad de modelos lineales.
  - ✓ LinearRegression: Regresión lineal por mínimos cuadrados ordinarios.
  - ✓ fit: Ajusta el modelo lineal.
  - ✓ predict: Predecir usando el modelo lineal.

- metrics: Implementa funciones que evalúan el error de predicción para propósitos específicos.
  - ✓ mean_squared_error: Calcula el Error Cuadrático Medio
  - ✓ r2_score: Calcula el Coeficiente de Determinación.

Regresión Lineal en Python

Otra librería utilizada es la conocida como matplotlib, la cual es una biblioteca completa para crear visualizaciones estáticas, animadas e interactivas en Python, entre las funciones que utilizaremos se encuentran:

- pyplot: Está diseñado principalmente para gráficos interactivos y casos simples de generación de gráficos.
  - ✓ scatter: Crea una diagrama de dispersión de Y vs X.
  - ✓ plot: Crea el gráfico de Y vs X como líneas.
  - ✓ show: Muestra los gráficos abiertos (generados).

Fuentes Consultadas

- Data, S. B. (2019, 7 diciembre). ¿Qué es la regresión lineal? Parte 1. sitiobigdata.com. https://sitiobigdata.com/2019/10/25/que-es-la-regresion-lineal/#
- Colaboradores de Wikipedia. (2006, 20 junio). Regresión lineal. Wikipedia, la enciclopedia libre. https://es.wikipedia.org/wiki/Regresi%C3%B3n_lineal
- A. (2018, 25 diciembre). Regresión lineal en Python. MachineLearningParaTodos.com. https://machinelearningparatodos.com/regresion-lineal-en-python/