

Regresión lineal

Hiram Misael Cerda Casas

Matrícula: 1911548

March 31, 2025

1 Introducción

La regresión lineal es una técnica estadística utilizada para modelar la relación entre una variable dependiente y una o más variables independientes. Su principal objetivo es encontrar una función lineal que mejor represente la relación entre las variables, permitiendo realizar predicciones y analizar tendencias. La ecuación general de la regresión lineal simple se expresa como:

$$y = mx + b \quad (1)$$

donde y es la variable dependiente, x es la variable independiente, m es la pendiente de la recta y b es la intersección con el eje y .

2 Metodología

En esta sección se describirán los pasos seguidos para implementar el modelo de regresión lineal.

De primeras, se tuvo que instalar el entorno de Python para poder crear la notebook sin complicaciones. Una vez descargado todo, se procedió al desarrollo del ejercicio.

El ejercicio consistió en analizar un conjunto de datos sobre artículos relacionados con Machine Learning, con el fin de predecir cuántas veces será compartido un artículo en función de sus características.

Se cargó el conjunto de datos en formato CSV, el cual contiene columnas como el número de palabras, enlaces, comentarios, imágenes y número de veces compartido.

```
# Cargamos los datos de entrada
```

```
data = pd.read_csv("./articulos_ml.csv")
```

```
# Filtramos la cantidad de palabras y compartidas
```

```
filtered_data = data[(data['Word-count'] <= 3500) & (data['#-Shares'] <= 80000)]
```

```
# Creamos nuestros datos de entrada
```

```
dataX = filtered_data[["Word-count"]]
X_train = np.array(dataX)
y_train = filtered_data['#-Shares'].values
```

```
# Creamos el objeto de Regresion Lineal
regr = linear_model.LinearRegression()
```

De la ecuación de la recta $y = mX + b$, se obtuvo que la pendiente m es 5.69 y el término independiente b es 11200.

Para probar el algoritmo, suponemos que queremos predecir cuántas veces será compartido un artículo sobre Machine Learning de 2000 palabras:

```
y_Dosmil = regr.predict([[2000]])
print(int(y_Dosmil))
```

El resultado obtenido fue de ****22595 veces compartido**** para un artículo de 2000 palabras.

3 Resultados

Se presentan los resultados obtenidos del modelo de regresión lineal aplicado.

Los valores obtenidos para los coeficientes del modelo son los siguientes:

- **Coefficiente (m):** 5.6977
- **Término independiente (b):** 11200.30
- **Error cuadrático medio:** 372,888,728.34
- **Varianza explicada (R^2):** 0.06
- **Predicción para 2000 palabras:** 22,595 "compartidas"

4 Conclusión

El aprendizaje sobre modelos de regresión lineal permite comprender cómo las variables pueden estar relacionadas y cómo hacer predicciones basadas en datos históricos. Esta herramienta es fundamental en el análisis de datos y en la toma de decisiones basada en información cuantificable.