



UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERÍA
DEPARTAMENTO DE CIENCIAS COMPUTACIONALES

Seminario de Solución de Problemas de Sistemas Basados en Conocimiento

Práctica No. 2

Nombre: Hurtado González Edgar Arturo
Código: 212597894

Introducción

La regresión lineal simple es un método estadístico utilizado para modelar la relación entre una variable dependiente (la que se quiere predecir) y una variable independiente (la que se utiliza para la predicción) mediante una línea recta.

En su forma más básica, la ecuación de la regresión lineal simple es:

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

Donde:

- Y es la variable dependiente que se quiere predecir.
- X es la variable independiente.
- β_0 es la intersección con el eje Y (el valor de y cuando x es igual a 0).
- β_1 es la pendiente de la línea, que indica cómo cambia y cuando x aumenta en una unidad.
- ϵ es el término de error, que representa la diferencia entre el valor real de y y el valor predicho por el modelo.

El objetivo de la regresión lineal es encontrar los valores de β_0 y β_1 que minimizan la suma de los cuadrados de los errores (mínimos cuadrados), es decir, encontrar la línea que mejor se ajuste a los datos observados.

Para hacer esto, se utilizan técnicas estadísticas como el método de mínimos cuadrados o algoritmos de optimización para calcular los coeficientes de β_0 y β_1 . Una vez que se ha ajustado el modelo, se puede utilizar para predecir los valores de la variable dependiente y para nuevos valores de la variable independiente x .

La regresión lineal simple es una herramienta poderosa y ampliamente utilizada en análisis de datos, predicción y modelado estadístico debido a su simplicidad y facilidad de interpretación. Sin embargo, es importante tener en cuenta sus suposiciones, como la linealidad y la independencia de los errores, para obtener resultados precisos y significativos.

Desarrollo

En la Práctica se realizó lo siguiente:

1. **Carga de datos:** Se utilizó la librería “**pandas**” para cargar un conjunto de datos desde un archivo CSV, que contiene información sobre años de experiencia y salarios.

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

# Carga de datos
data = pd.read_csv("Salary_dataset.csv")
```

2. **Extracción de variables:** Se tomaron las columnas relevantes del conjunto de datos para las variables “years_experience” y “salary”.

```
# Extracción de variables
years_experience = data["YearsExperience"]
salary = data["Salary"]
```

3. **Cálculo de promedios:** Se calcularon los promedios de años de experiencia y salarios utilizando la función “**np.mean()**” de la librería “**numpy**”.

```
# Cálculo de promedios
mean_years_experience = np.mean(years_experience)
mean_salary = np.mean(salary)
```

4. **Cálculo de la pendiente y el intercepto:** Se implementó una función “**calculate_slope()**” para calcular la pendiente (coeficiente β_1) utilizando la fórmula de regresión lineal. También se calculó el intercepto (coeficiente β_0).

```
# Función para calcular la pendiente (coeficiente beta_1)
def calculate_slope(x, y, mean_x, mean_y):
    numerator = np.sum((x - mean_x) * (y - mean_y))
    denominator = np.sum((x - mean_x) ** 2)
    return numerator / denominator

# Cálculo de la pendiente (coeficiente beta_1)
slope = calculate_slope(years_experience, salary, mean_years_experience, mean_salary)

# Cálculo del intercepto (coeficiente beta_0)
intercept = mean_salary - slope * mean_years_experience
```

5. **Predicción para nuevos datos:** Se definió un nuevo valor de años de experiencia “**new_experience**” y se utilizó la ecuación de la línea de regresión (basada en la pendiente y el intercepto calculados) para predecir el salario correspondiente a ese nuevo valor.

```
# Predicción para nuevos datos
new_experience = 6.5
predicted_salary = intercept + slope * new_experience
print("Predicción del Salario con", str(new_experience), "Años de Experiencia:", predicted_salary)
```

6. **Gráfica de dispersión y línea de regresión:** Se utilizó “matplotlib” para visualizar los datos originales como una gráfica de dispersión y trazar la línea de regresión lineal obtenida sobre estos datos. Además, se añadió el punto predicho y su valor en la gráfica para ilustrar la predicción realizada.

```
# Gráfica de dispersión y línea de regresión
plt.figure(figsize=(8, 6))
plt.scatter(years_experience, salary, label="Datos Originales")
plt.plot(years_experience, slope * years_experience + intercept, color="red", linewidth=1, label="Línea de Regresión")
plt.xlabel('Años de Experiencia')
plt.ylabel('Salario')
plt.title('Regresión Lineal (Salario según la Experiencia)')
plt.legend()

# Agregando el valor predicho en la gráfica
plt.text(new_experience, predicted_salary, f'Predicción: {predicted_salary:.2f}', color='red', fontsize=9)

# Agregando el punto predicho en el grid
plt.scatter(new_experience, predicted_salary, color='purple', marker='o', s=100, label='Predicción')

plt.grid(True)
plt.show()
```

Conclusión/Resultados

El ejercicio realizado muestra un proceso básico de regresión lineal simple para predecir salarios basados en años de experiencia. A través del uso de datos reales, se ilustró cómo se pueden calcular la pendiente y el intercepto de una línea de regresión para modelar la relación entre dos variables. La regresión lineal simple es una herramienta poderosa para analizar y predecir relaciones entre variables cuando se asume una relación lineal.

Predicción del Salario con 6.5 Años de Experiencia: 86272.95905598119

