Linear Regression Project

Name: Swastika Das

Registration Number: 2023PGDM1286

Subject: Machine Learning with Python

Date: 01-07-2025

# 1. Introduction

Linear Regression is a fundamental statistical technique in predictive modeling, commonly used in business analytics to forecast trends, demand, and evaluate risk. As an MBA student, understanding linear regression empowers us to make data-driven decisions. It helps estimate relationships between variables, which is crucial for operations, finance, and marketing.

# 2. Objective

The objective of this project is to implement a simple Linear Regression model in Python to predict outcomes based on independent variables. We'll demonstrate it using a sample dataset.

# 3. Methodology

The steps followed in this project include:  
1. Importing libraries  
2. Loading the dataset  
3. Visualizing the data  
4. Splitting data into training and testing sets  
5. Training the Linear Regression model  
6. Predicting and evaluating the model

# 4. Python Code with Explanation

# Importing libraries  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
from sklearn.model\_selection import train\_test\_split  
from sklearn.linear\_model import LinearRegression  
from sklearn.metrics import mean\_squared\_error, r2\_score  
  
# Creating sample data  
data = {  
 'Experience': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],  
 'Salary': [35000, 40000, 45000, 50000, 60000, 65000, 70000, 75000, 80000, 85000]  
}  
df = pd.DataFrame(data)  
  
# Visualizing the data  
plt.scatter(df['Experience'], df['Salary'])  
plt.title('Experience vs Salary')  
plt.xlabel('Years of Experience')  
plt.ylabel('Salary')  
plt.show()  
  
# Splitting data  
X = df[['Experience']]  
y = df['Salary']  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=0)  
  
# Model training  
model = LinearRegression()  
model.fit(X\_train, y\_train)  
  
# Prediction  
y\_pred = model.predict(X\_test)  
  
# Evaluation  
print("Mean Squared Error:", mean\_squared\_error(y\_test, y\_pred))  
print("R^2 Score:", r2\_score(y\_test, y\_pred))  
  
# Visualization of Predictions  
plt.scatter(X\_test, y\_test, color='red')  
plt.plot(X\_test, y\_pred, color='blue')  
plt.title('Actual vs Predicted')  
plt.xlabel('Experience')  
plt.ylabel('Salary')  
plt.show()

# 5. Conclusion

This project demonstrates a basic application of Linear Regression using Python. The model successfully learns the relationship between experience and salary, helping us predict future salaries. Such predictive techniques are vital in business scenarios for budgeting, forecasting, and strategic planning.

# 6. References

1. scikit-learn documentation - https://scikit-learn.org  
2. Python Official Documentation - https://docs.python.org  
3. Linear Regression, Wikipedia - https://en.wikipedia.org/wiki/Linear\_regression