linearRegression

November 14, 2024

Linear Regression - Linear regression is a supervised machine learning algorithm used for predicting a continuous dependent variable based on one or more independent variables. It establishes a linear relationship between the dependent and independent variables.

Key Concepts - Linear regression tries to model the relationship between two (or more) variables by fitting a linear equation to observed data. The primary goal is to find the best-fitting line (regression line) that represents this relationship. Dependent Variable (Y): The variable we are trying to predict. Independent Variable (X): The variable(s) used to predict the dependent variable. Simple Linear Regression Equation: Y = b0 + b1*X + e

How it works and how is it used? - Linear regression works by finding the best-fitting line that minimizes the residual sum of squares (RSS) between the observed values and the predicted values. It uses statistical techniques like Ordinary Least Squares (OLS) to determine the optimal values of the coefficients.

- Simple Linear Regression: Involves only one independent variable. Multiple Linear Regression: Involves two or more independent variables. How it is used: Linear regression is often used in forecasting, trend analysis, and risk assessment in fields like finance, healthcare, real estate, and economics.

List of Important Types - Simple Linear Regression: A linear relationship between a single independent variable and a dependent variable. - Multiple Linear Regression: A linear relationship between two or more independent variables and a dependent variable. - Ridge Regression: A type of regression that adds a penalty to the size of coefficients to reduce overfitting. - Lasso Regression: Similar to Ridge Regression but uses L1 regularization, which can reduce coefficients to zero, thus performing variable selection. - ElasticNet Regression: A combination of Lasso and Ridge regression techniques.

Goal - The primary goal of linear regression is to establish a relationship between the independent and dependent variables and to make predictions based on that relationship. It aims to minimize the error between the predicted and actual values.

Important Formula - The formula for a simple linear regression is: 1. Simple Linear Regression Equation: - Y = b0 + b1*X + e

2. Simple Linear Regression Variables:

- Y = Predicted dependent variable
- b0 = Intercept
- b1 = Slope of the regression line
- X = Independent variable
- e = Error term

3. Multiple Linear Regression Equation:

- Y = b0 + b1*X1* + *b2*X2 + … + bn*Xn + e

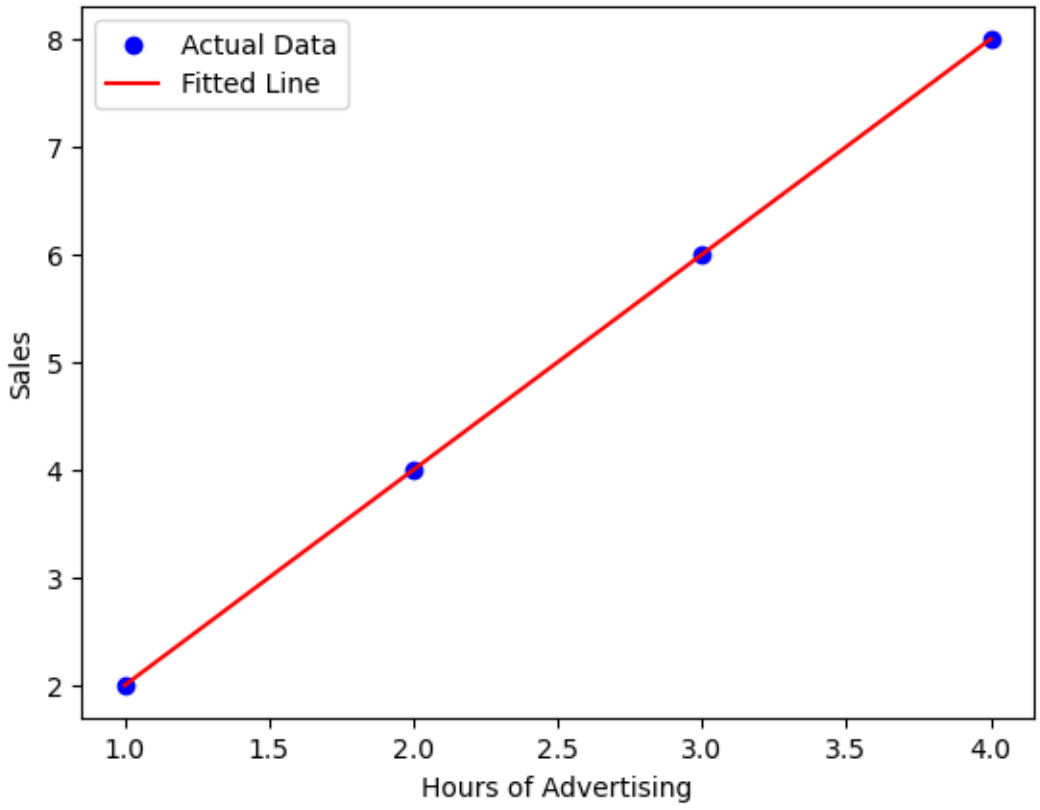
4. Multiple Linear Regression Variables:

- X1, X2, …, Xn = Multiple independent variables

Example - Suppose you are trying to predict the sales of a company based on the number of hours of advertising. By applying linear regression, you would find the best-fitting line (equation) that predicts sales based on advertising hours.

[5]:
```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

# Data
X = np.array([[1], [2], [3], [4]])  # Hours of Advertising
Y = np.array([2, 4, 6, 8])  # Sales
# Model
model = LinearRegression()
# Fit the model
model.fit(X, Y)
# Predictions
Y_pred = model.predict(X)
# Plotting the result
plt.scatter(X, Y, color='blue', label='Actual Data')
plt.plot(X, Y_pred, color='red', label='Fitted Line')
plt.xlabel('Hours of Advertising')
plt.ylabel('Sales')
plt.legend()
plt.show()
# Coefficients
print(f"Intercept: {model.intercept_}")
print(f"Slope: {model.coef_[0]}")
```




```
Intercept: -1.7763568394002505e-15
Slope: 2.000000000000001
```

Real World Scenario - Predicting house prices based on features like square footage, number of bedrooms, etc. You want to predict the price of a house based on its size (in square feet). Application: Linear regression helps establish the relationship between the house price (dependent variable) and its size (independent variable), allowing real estate agents to predict prices based on

Problem Statement - Given a dataset of advertising hours and sales for a retail company, the task is to predict future sales based on advertising efforts.

How it can help - Linear regression can help in understanding the relationships between variables and making predictions. For example, in business, it can help forecast sales, assess the impact of marketing campaigns, or predict customer behavior based on historical data.

Alternate Solution - If the data is not linear or has multicollinearity, polynomial regression (a more flexible version of linear regression) or decision tree regression could be used as alternative models. For problems with high-dimensional data, regularization techniques like Ridge or Lasso can improve performance.