Linear Regression: A Quick Revision

Generated by collegeAi

Introduction

- Supervised learning algorithm for predicting a continuous target variable.
- Models the relationship between independent variable(s) and a dependent variable.
- Goal: Find the "best fit" line (or hyperplane in higher dimensions) through the data.

Section 1: Simple Linear Regression

- One independent variable (x) and one dependent variable (y).
- Equation: $y = mx + b$
- y: predicted value (dependent variable)
- x: independent variable
- m: slope (coefficient)
- b: y-intercept (constant)
- Aims to minimize the sum of squared errors (least squares method).

Section 2: Multiple Linear Regression

- More than one independent variable (x1, x2, x3, ...).
- Equation: y = b0 + b1*x1 + b2*x2 + b3*x3 + ...
- b0 is the y-intercept.
- b1, b2, b3,... are coefficients for each independent variable.
- Still aims to minimize the sum of squared errors, but with more dimensions.

Example

- Simple Linear Regression: Predicting house prices based on square footage.
- Independent variable (x): Square footage
- Dependent variable (y): House price
- Multiple Linear Regression: Predicting sales based on advertising spend across different channels (TV, radio, social media).
- Independent variables (x1, x2, x3): TV spend, Radio spend, Social Media spend
- Dependent variable (y): Sales

Python Code Example (if applicable)

- ```python
- from sklearn.linear_model import LinearRegression
- import numpy as np
- # Sample Data (replace with your actual data)
- X = np.array([[1], [2], [3], [4], [5]]) # Independent variable
- y = np.array([2, 4, 5, 4, 5]) # Dependent variable
- # Create and train the model
- model = LinearRegression()
- model.fit(X, y)
- # Make a prediction
- new_x = np.array([[6]])
- predicted_y = model.predict(new_x)
- print(f"Predicted y for x=6: {predicted_y[0]}")

Summary

- Linear Regression models linear relationships between variables.
- Simple Linear Regression uses one independent variable. Multiple Linear Regression uses multiple.
- The goal is to minimize the sum of squared errors between predicted and actual values.