MATHEMATICS FOR MACHINE LEARNING

LAB 9 – 5%

Regression Analysis

Solve the following questions. You can use Python or solve them on paper (Or you can solve some using Python and some on paper). If you are using Python to solve, please make sure to attach the output in your submissions. Like the other labs, please hand in PDF and/or Python code with output.

Dataset

For Simple Linear Regression (Questions 1, 2, 4, 5, 6, and 7)

The following dataset represents the relationship between years_of_experience and salary (in dollars).

Years of Experience	Salary ($)
1	30,000
2	35,000
3	40,000
4	45,000
5	50,000
6	55,000
7	60,000
8	65,000
9	70,000
10	75,000

For Multiple Linear Regression (Questions 3, 8, 9, and 10)

The dataset below includes three independent variables:

Years of Experience – Number of years the employee has worked

Education Level – Level of education (1 = High School, 2 = Bachelor's, 3 = Master's)

Location – Categorical representation of the work location (1, 2, or 3)

Salary ($) – Dependent variable

Question 1: Simple Linear Regression

You are given the dataset above, where the independent variable is years_of_experience, and the dependent variable is salary. Implement a simple linear regression model to predict salary based on years of experience. Calculate the regression coefficients and plot the best-fitting line.

Question 2: Interpretation of the Slope and Intercept

Using the simple regression model created in Question 1, explain the meaning of the slope and intercept of the regression line. What do they represent in the context of this dataset?

Question 3: Multiple Linear Regression

The dataset contains years_of_experience, education_level, and location as independent variables and salary as the dependent variable. Implement a multiple regression model and explain the relationship between these independent variables and salary.

Question 4: Error Estimation in Linear Regression

Using the regression model from Question 1, calculate the Mean Squared Error (MSE) and the RMSE to evaluate the model's performance.

Question 5: Best Fit Line Calculation

Given the dataset for simple linear regression, calculate the best-fit line using the formula for simple linear regression.

Question 6: Predicting Values

Using the best-fit line obtained in Question 5 to predict the salary for an employee with 6 years of experience.

Question 7: Residuals Analysis

For the simple regression model in Question 1, calculate the residuals (the difference between actual and predicted values). Plot the residuals to check if the assumptions of linear regression are met.

Question 8: Correlation and Regression Coefficients

Using the dataset for multiple linear regression, calculate the correlation coefficient between each independent variable and the dependent variable (salary).

Question 9: Regularization in Multiple Regression

In a multiple regression scenario with several predictors, apply Lasso (L1 regularization) and Ridge (L2 regularization) to prevent overfitting. Compare the coefficients obtained using these methods.