Hometask 12

A healthcare company wants to predict the muscle mass percentage of individuals based on three factors: daily protein intake (in grams), exercise hours per week and hydration level (in liters per day). The dataset is as follows:

Protein Intake	Exercise	Hydration	Muscle Mass Percentage
80	5	3	45
100	7	3.5	50
90	6	3.2	48
110	8	4	55
95	6.5	3.8	50

(a) Perform 3 iterations of Multivariate Linear Regression using Gradient Descent with learning rate $\eta = 0.0001$.

- Initialize all weights w_1, w_2, w_3 and bias b to 0
- After each iteration, report the updated coefficients b, w_1, w_2 and w_3.
- Calculate the Mean Squared Error (MSE) after each iteration
- Use the final model to predict the muscle mass percentage for an individual consuming 105g of protein daily, exercising 7.5 hours per week and hydrating with 3.6L per day.
- What are the pros and cons of using Gradient Descent for training linear regression models? What is the impact of choosing a very large or very small learning rate?

(b) Perform 3 iterations of Ridge Regression on the same dataset using Gradient Descent with regularization parameter $\alpha = 0.1$ and learning rate $\eta = 0.0001$.

- Report the updated values of b, w_1, w_2 and w_3 after each iteration.
- Calculate the Mean Squared Error (MSE) after each iteration
- Use the final model to predict the muscle mass percentage for an individual consuming 105g of protein daily, exercising 7.5 hours per week and hydrating with 3.6L per day.
- Explain the main differences between L1 and L2 regularization. How does regularization help prevent overfitting in linear models?