Polytechnique Montréal

INF8245AE – Machine Learning

Assignment 1 – Linear Regression

Mattéo Colavita - 2142009

2025-09-26

1 Question 1 : Linear and Weighted Ridge Regression

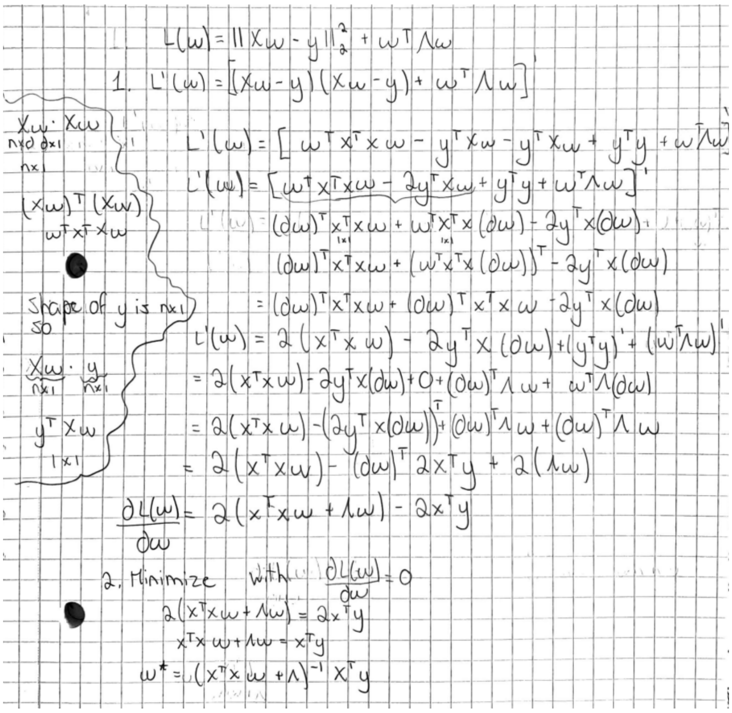

FIGURE 1 – Derivative of the loss function for ridge regression with respect to the weights.

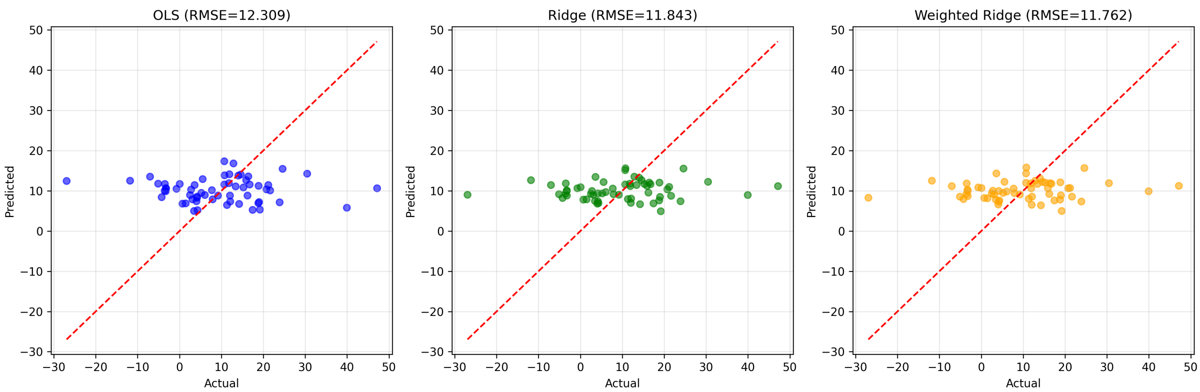

FIGURE 2 – Comparison of predictions from linear, ridge and weighted ridge regression on the test set.

2 Question 2 : Cross-Validation

Metric	Best λ	λ=0.01	λ=0.1	λ=1	λ=10	λ=100
MAE	10	7.381	7.316	7.140	7.110	7.817
MaxError	100	27.758	27.681	27.559	27.476	27.095
RMSE	10	9.855	9.772	9.577	9.532	10.101

TABLE 1 – Mean MAE, MaxError, and RMSE scores obtained via 5-fold cross-validation for different values of λ in ridge regression.

3 Question 3 : Gradient Descent for Ridge Regression

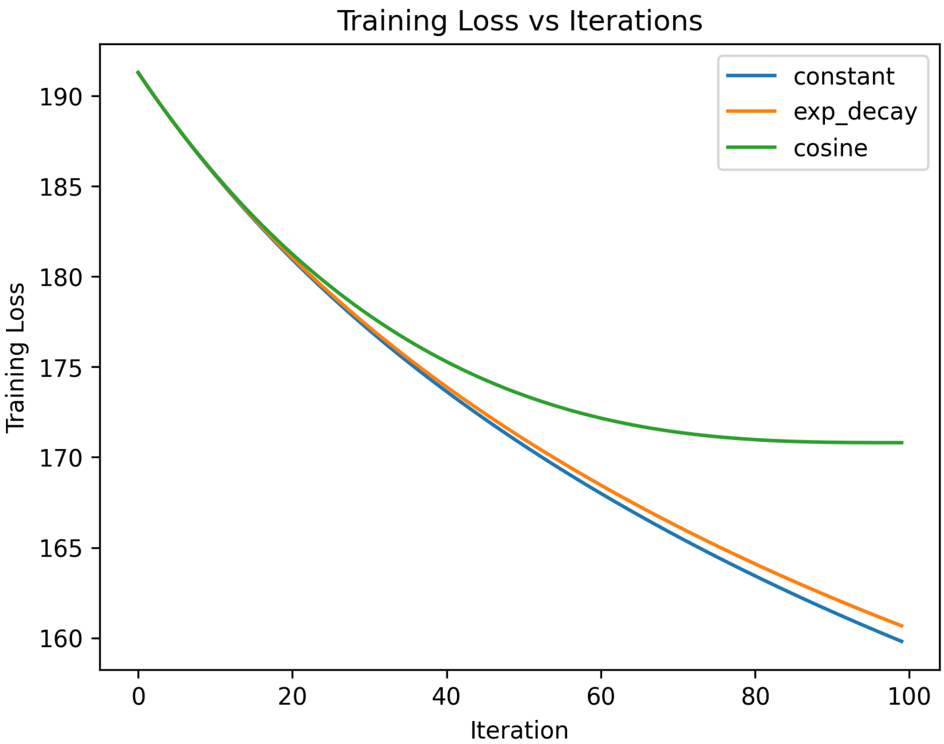

FIGURE 3 – Training loss vs iterations for different learning rate schedules in gradient descent for ridge regression.

Learning Rate Schedule	**RMSE**
Constant	13.8910
Exponential Decay	13.9283
Cosine Annealing	14.3473

TABLE 2 – RMSE results for different learning rate schedules in gradient descent for ridge regression.

Among the tested learning rate schedules, both constant and exponential decay schedules show a fast decrease in training loss and achieved low RMSE values of 13.89 and 13.93, suggesting great generalization. In contrast, cosine annealing plateaued early, indicating premature convergence at a suboptimal point. The rapid decay of the learning rate within the given amount of iterations may have caused smaller step sizes before reaching the minimum, preventing stable convergence despite ridge regression's convex loss function.