This project is mainly about using python to implement cross validation for ridge regression. We tune the regularization parameter lambda to see how it affects the training loss and the test loss of ridge regression.

Coding: For this project, we mainly use sklearn. (1)Apply KFold function from sklearn to split the data into 10 folds. (2)For each lambda value, we use Ridge.fit and Ridge .predict (also from sklearn) to train on 9 folds and predict on the remaining 1 fold. (3) Finally, we calculate the Root Mean Squared Error (RMSE) between model predictions and true labels.

Results: In this project, we analyze the performance of ridge regression using RMSE. Using 10-fold cross validation, we train the model on 9 folds of data and see the test loss on the remaining 1 fold. From the result, we can see that at first, when lambda increases, the RMSE decreases, which means the model has better generalization to the new test data. But when we further increase lambda, the RMSE increases, the reason is probably that too large lambda leads to large bias, i.e. large difference between predicted f\_hat and true f\_star.