Homework Report 1

Problem 1

Briefly explain how you used batch stochastic gradient descent with regularization to learn the weights. Think about how the regularization is incorporated into the loss function and how that affects the gradient when updating weights.

L2 regularization is a technique where a penalty term is added to the loss function that is proportional to the square of the model's parameters. Therefore, when calculating gradient, we add 2 * self.lmbda * self.weights to it, which means that when updating the weights, the regularization term is subtracted from the gradient. This encourages the model to have smaller weights, which helps prevent overfitting.

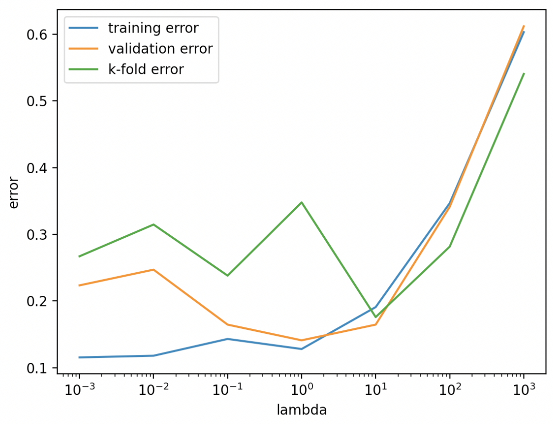
Problem 2

Think back to when you implemented Logistic Regression on the Census dataset. How would it have been different if you applied Tikhonov regularization? Specifically, how would the regularization affect the accuracy and the types of errors?

This regularization technique is typically used to prevent overfitting and improve the model's generalization performance. By adding a regularization term to the loss function, the model's parameters are encouraged to have smaller magnitudes, which can help reduce the variance of the model and, in turn, reduce overfitting. In terms of the accuracy of the model, Tikhonov regularization can help improve the model's performance by reducing overfitting. This means that the model may have a better ability to generalize to unseen data, which can improve the accuracy of the model on the test set. In terms of the types of errors, the regularization term can help reduce false positives and false negatives by improving the model's generalization performance. Note that there may be a risk of underfitting if the regulaization is not chosen carefully for the best bias-complexity tradeoff.

Problem 3

Use plotError(), which we have implemented for you, to produce a model selection curve. Include your plot here. Then, conclude what the best value of lambda is and explain why. NOTE: It takes about five minutes to generate a graph. Please set your default lambda in the constructor to your optimal lambda you discovered for TA testing purposes.



The above is the plot produced by running on 1000 epochs and batch size 1. We can conclude that the best lambda is 1, since the validation error and training error is the lowest there. This means that the average model performance is the best when the lambda value is 1.

Problem 4

In this project, you used validation data to select a model. Suppose that each patient might've had multiple samples (e.g., multiple lab tests or x-rays) collected and entered into the dataset. Would you need to account for this when splitting your train-validation-test data? If yes, how? If no, why not? (3-5 sentences)

Yes, we need to account for this. One way to account for this correlation is to ensure that samples from the same patient are not split across different sets. This can be achieved by assigning all samples from the same patient to the same set. Another way is to utilize the k-fold cross-validation strategy, where the data is split into K folds, and each fold includes all samples from the same patient.This would help minimize the risk of overfitting to specific patients.