ECE 1395 - Spring 2025

Homework Assignment 4: Regularization, KNN, and One-vs-all Classification

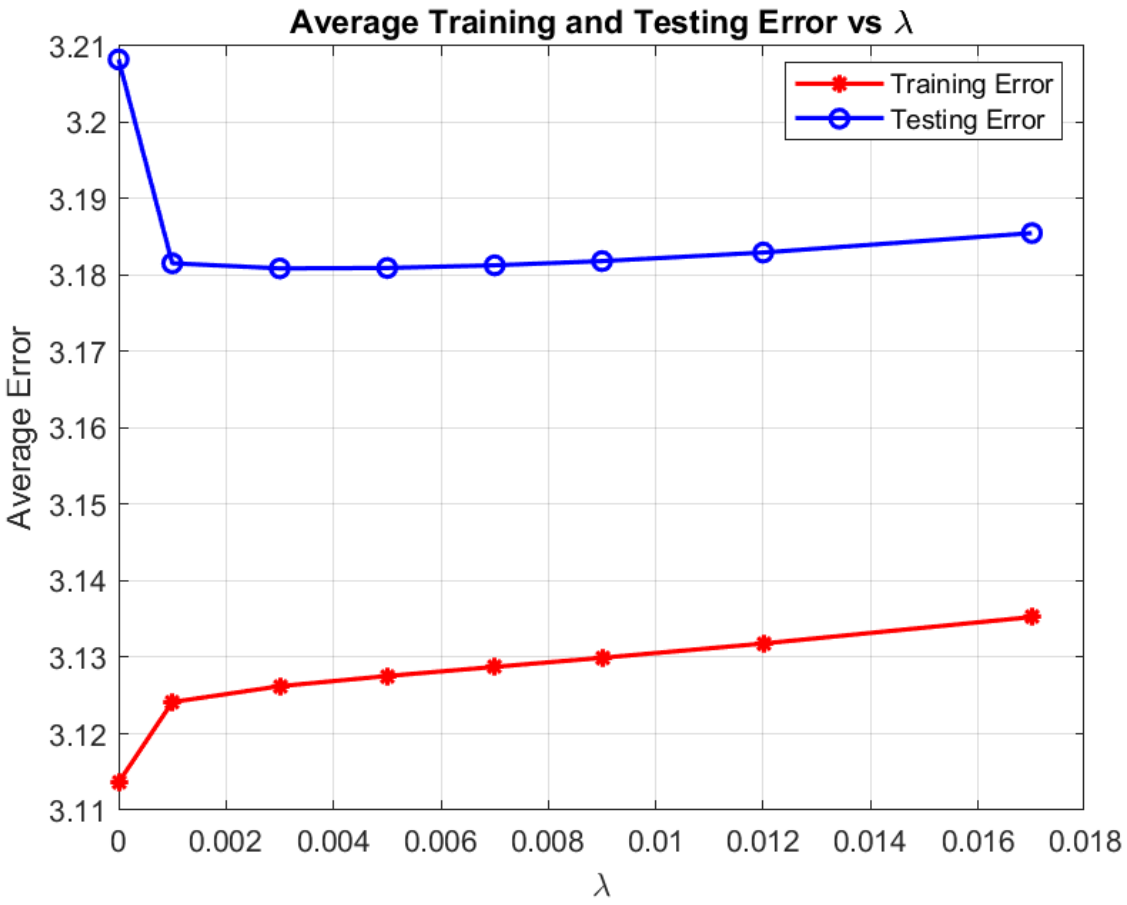
Name: Omar Morsy

ID: oam15@pitt.edu

1. Question 1: Regularization

a) Regularized Normal Equation Implementation

The function Reg_normalEqn.m has been implemented to compute the closed-form solution to linear regression using normal equation with regularization.



b) Feature Matrix Size

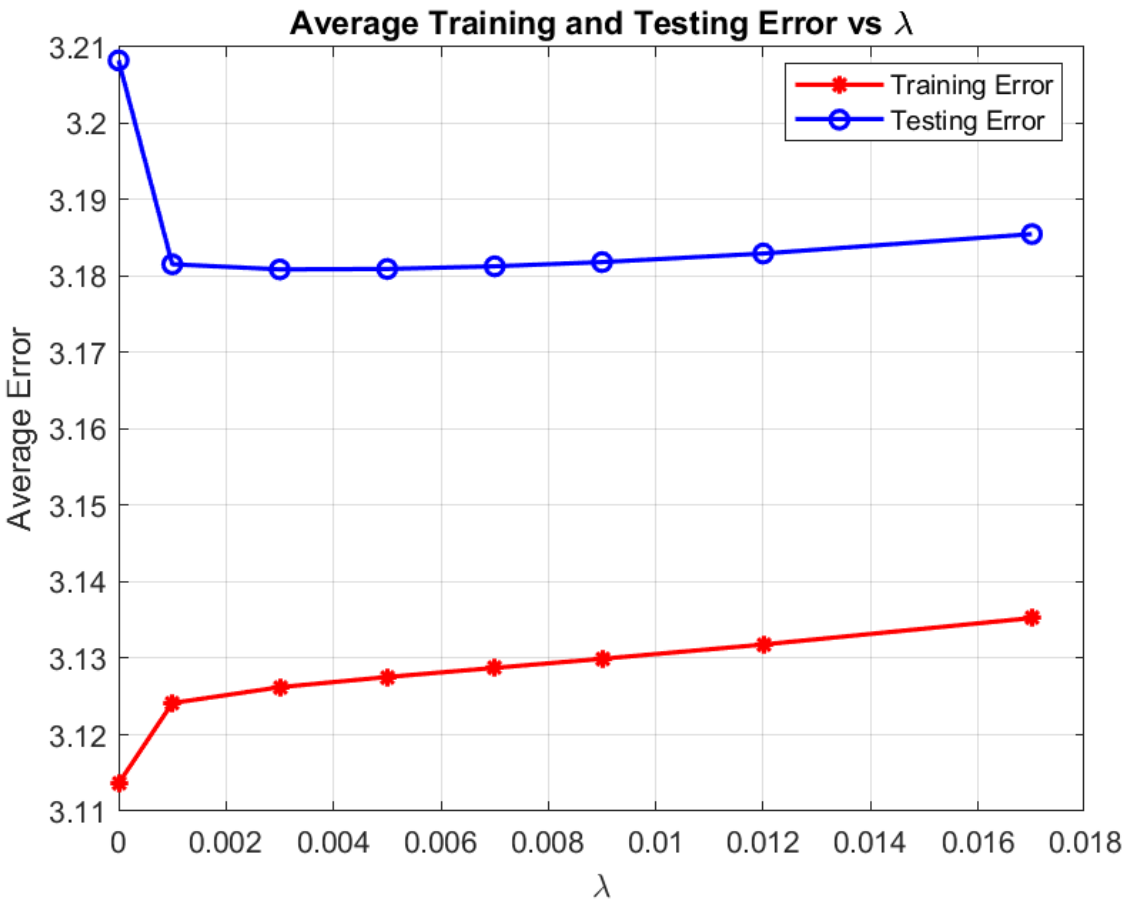
The size of feature matrix: 1001 x 500

```
x =

  struct with fields:

    X_data: [1001×500 double]
         y: [1001×1 double]
```

c) Training and Testing Error Analysis

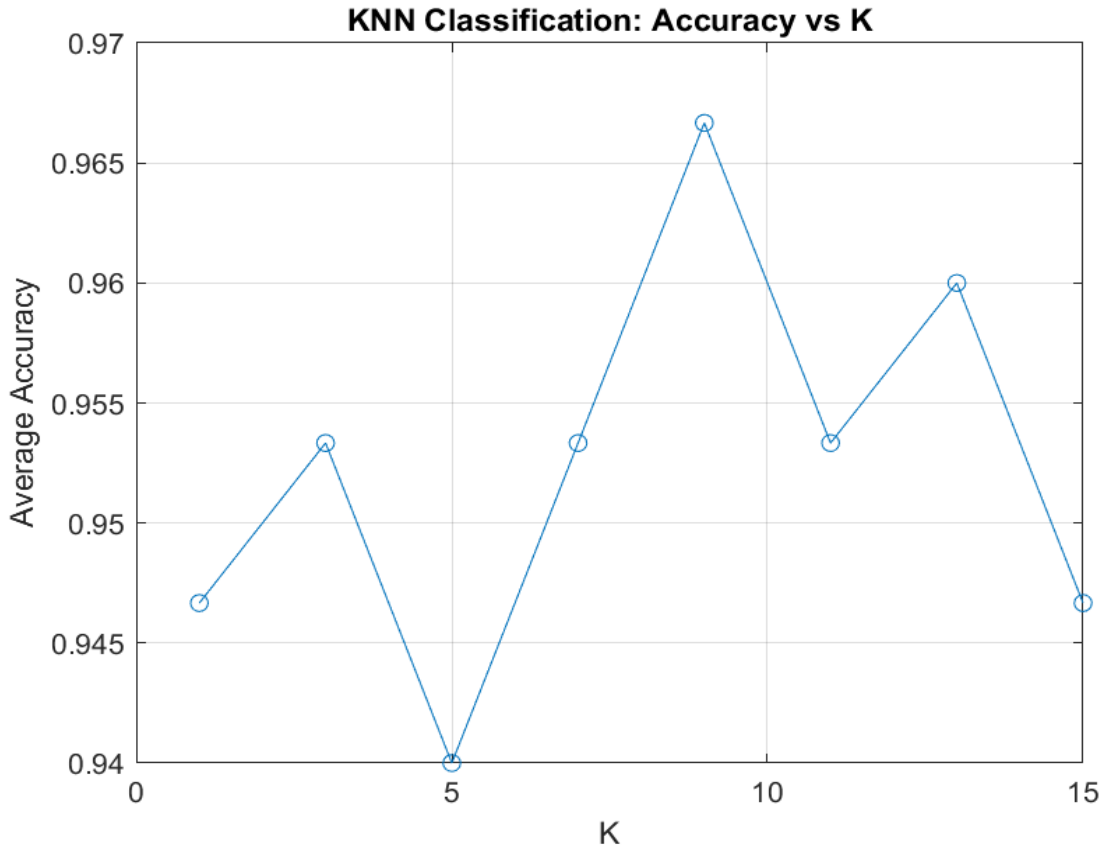


Based on the graph of training and testing errors versus λ, I suggest using λ=0.003-0.005 as it achieves the minimum testing error while maintaining a reasonable gap with training error.

Without regularization (λ=0), the model overfits as shown by high testing error, while larger λ values lead to underfitting with both errors increasing. The chosen λ value provides the optimal balance between model complexity and generalization ability, as evidenced by the convergence of training and testing errors at this point

2. Question 2: KNN - Effect of K

The figure below shows the average accuracy vs K for the KNN classifier:



K=9 achieves the highest accuracy of 0.966 and is the recommended value for this specific dataset. However, this K value is not necessarily robust for other classification problems since the optimal K depends heavily on each dataset's unique characteristics like size, dimensionality, and class distribution. For any new problem, cross-validation should be performed to find the optimal K value, as what works well for one dataset may not be optimal for another.

3. Question 3: One-vs-all Classification

Part a:

The one-vs-all classification approach has been implemented in logReg_multi.m using logistic regression as the base classifier.

Part b:

Training Accuracy: 0.9360 Testing Accuracy: 0.8400

The one-vs-all classifier shows signs of overfitting with 93.60% training accuracy dropping to 84.00% testing accuracy, indicating that while the model learns the training data well, it doesn't generalize as effectively to new data. Despite the 9.6% performance gap suggesting room for improvement through regularization, the model still achieves reasonable performance with 84% accuracy on unseen test data.