

Ruixiang Qi HW#3 CS534

1.a.

The training set contains 351 records.

After shuffling, we divide the data into 5 folds, with 70 records($1/5$ of the data) in each of the first 4 folds and 71 records in the last fold.

Fold 1: Index 1-70

Fold 2: Index 71- 140

Fold 3: Index 141-210

Fold 4: Index 211-280

Fold 5: Index 281-251

Our dataset looks like:

Fold1	Fold 2	Fold 3	Fold4	Fold 5
Index 1 - 70	Index 71- 140	Index 141-210	Index 211-280	Index 281-351

Cross Validation Starts:

Outer Loop:

1st Iteration:

Training Folds + Validation Folds				Testing Fold
Index 1 - 70	Index 71- 140	Index 141-210	Index 211-280	Index 281-351

Inner Loop: (use each of the 4 folds as validation fold in each iteration)

1st Iteration:

Training Folds			Validation Fold	Testing Fold
Index 1 - 70	Index 71- 140	Index 141-210	Index 211-280	Index 281-351

2nd Iteration:

Training Folds			Validation Fold	Testing Fold
Index 1 - 70	Index 71- 140	Index 211-280	Index 141-210	Index 281-351

3rd Iteration:

Training Folds			Validation Fold	Testing Fold
Index 1 - 70	Index 141-210	Index 211-280	Index 71- 140	Index 281-351

4th Iteration:

Training Folds			Validation Fold	Testing Fold
Index 71- 140	Index 141-210	Index 211-280	Index 1 - 70	Index 281-251

2nd Iteration:

Training Folds + Validation Folds				Testing Fold
Index 1 - 70	Index 71- 140	Index 141-210	Index 281-351	Index 211-280

Inner Loop: (use each of the 4 folds as validation fold in each iteration)

1st Iteration:

Training Folds			Validation Fold	Testing Fold
Index 1 - 70	Index 71- 140	Index 141-210	Index 281-351	Index 211-280

2nd Iteration:

Training Folds			Validation Fold	Testing Fold
Index 1 - 70	Index 71- 140	Index 281-351	Index 141-210	Index 211-280

3rd Iteration:

Training Folds			Validation Fold	Testing Fold
Index 1 - 70	Index 281-351	Index 141-210	Index 71- 140	Index 211-280

4th Iteration:

Training Folds			Validation Fold	Testing Fold
Index 281-251	Index 141-210	Index 71- 140	Index 1 - 70	Index 211-280

3rd Iteration:

Training Folds + Validation Folds				Testing Fold
Index 1 - 70	Index 71- 140	Index 211-280	Index 281-351	Index 141-210

Inner Loop: (use each of the 4 folds as validation fold in each iteration)

1st Iteration:

Training Folds			Validation Fold	Testing Fold
Index 1 - 70	Index 71- 140	Index 211-280	Index 281-351	Index 141-210

2nd Iteration:

Training Folds			Validation Fold	Testing Fold
Index 1 - 70	Index 71- 140	Index 281-351	Index 211-280	Index 141-210

3rd Iteration:

Training Folds			Validation Fold	Testing Fold
Index 1 - 70	Index 281-351	Index 211-280	Index 71- 140	Index 141-210

4th Iteration:

Training Folds			Validation Fold	Testing Fold
Index 281-251	Index 211-280	Index 71- 140	Index 1 - 70	Index 141-210

4th Iteration:

Training Folds + Validation Folds				Testing Fold
Index 1 - 70	Index 141-210	Index 211-280	Index 281-251	Index 71- 140

Inner Loop: (use each of the 4 folds as validation fold in each iteration)

1st Iteration:

Training Folds			Validation Fold	Testing Fold
Index 1 - 70	Index 141-210	Index 211-280	Index 281-351	Index 71- 140

2nd Iteration:

Training Folds			Validation Fold	Testing Fold
Index 1 - 70	Index 141-210	Index 281-351	Index 211-280	Index 71- 140

3rd Iteration:

Training Folds			Validation Fold	Testing Fold
Index 1 - 70	Index 281-351	Index 211-280	Index 141-210	Index 71- 140

4th Iteration:

Training Folds			Validation Fold	Testing Fold
Index 281-251	Index 211-280	Index 141-210	Index 1 - 70	Index 71- 140

5th Iteration:

Training Folds + Validation Folds				Testing Fold
Index 71- 140	Index 141-210	Index 211-280	Index 281-251	Index 1 - 70

Inner Loop: (use each of the 4 folds as validation fold in each iteration)

1st Iteration:

Training Folds			Validation Fold	Testing Fold
Index 71- 140	Index 141-210	Index 211-280	Index 281-351	Index 1 - 70

2nd Iteration:

Training Folds			Validation Fold	Testing Fold
Index 71- 140	Index 141-210	Index 281-351	Index 211-280	Index 1 - 70

3rd Iteration:

Training Folds			Validation Fold	Testing Fold
Index 71- 140	Index 281-351	Index 211-280	Index 141-210	Index 1 - 70

4th Iteration:

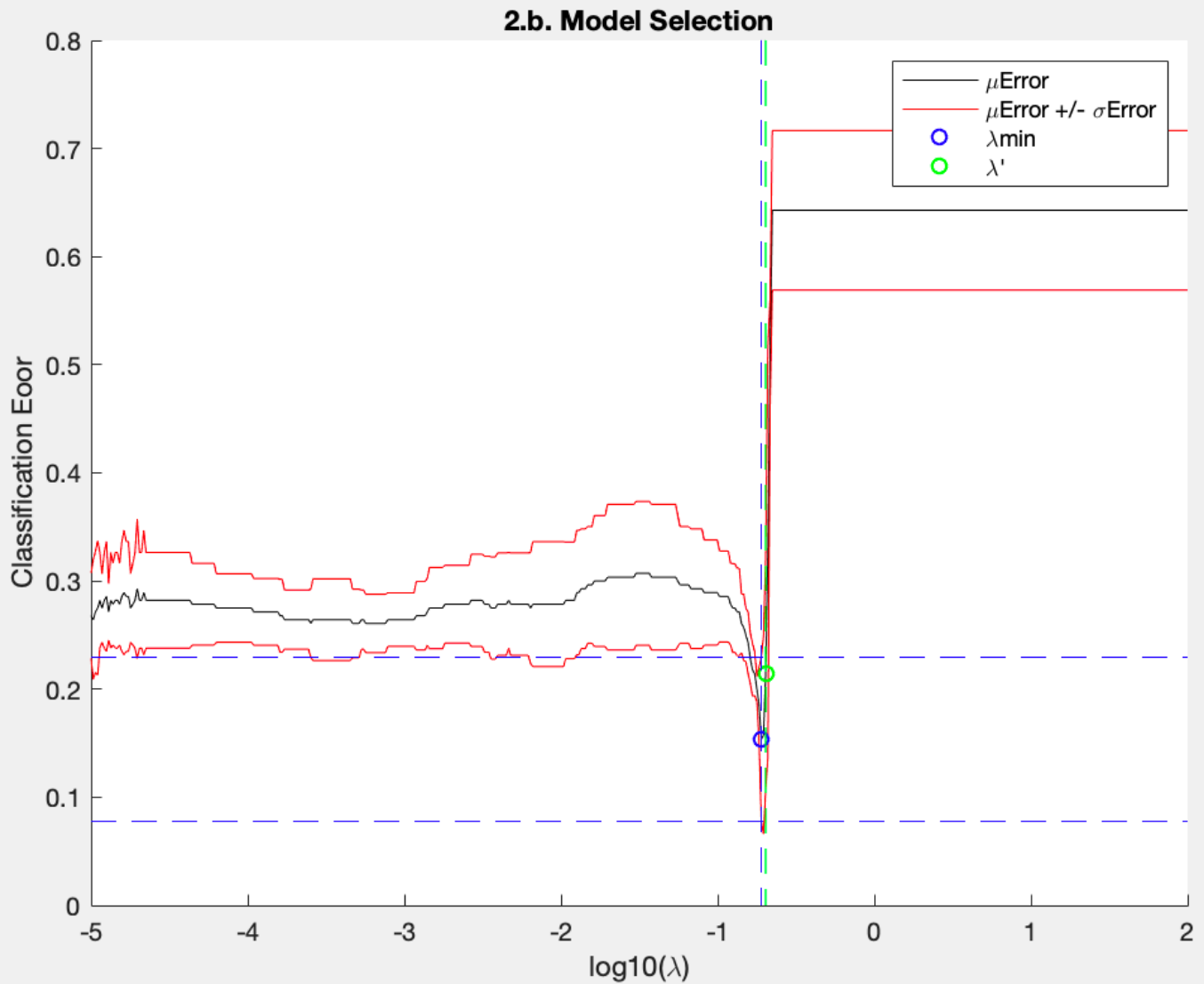
Training Folds			Validation Fold	Testing Fold
Index 281-251	Index 211-280	Index 141-210	Index 71- 140	Index 1 - 70

*I plotted graphs in 1b and 1c in the same Matlab file

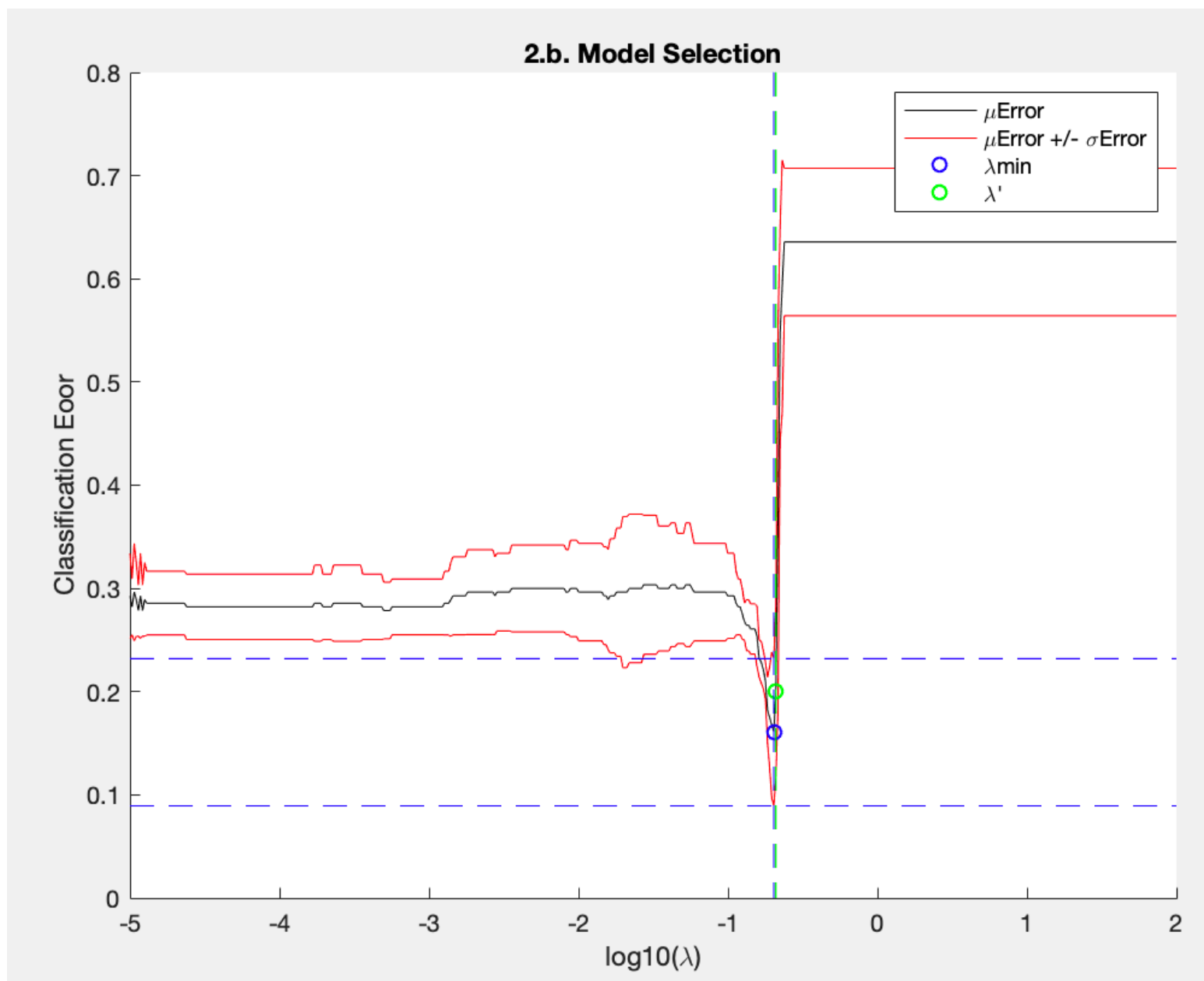
1.b

The following graphs are classification error vs. $\log(\lambda)$ for each of the 5 outer loop iterations. Since the sample was randomized before the cross validation, it is reasonable that the 5 graphs look similar.

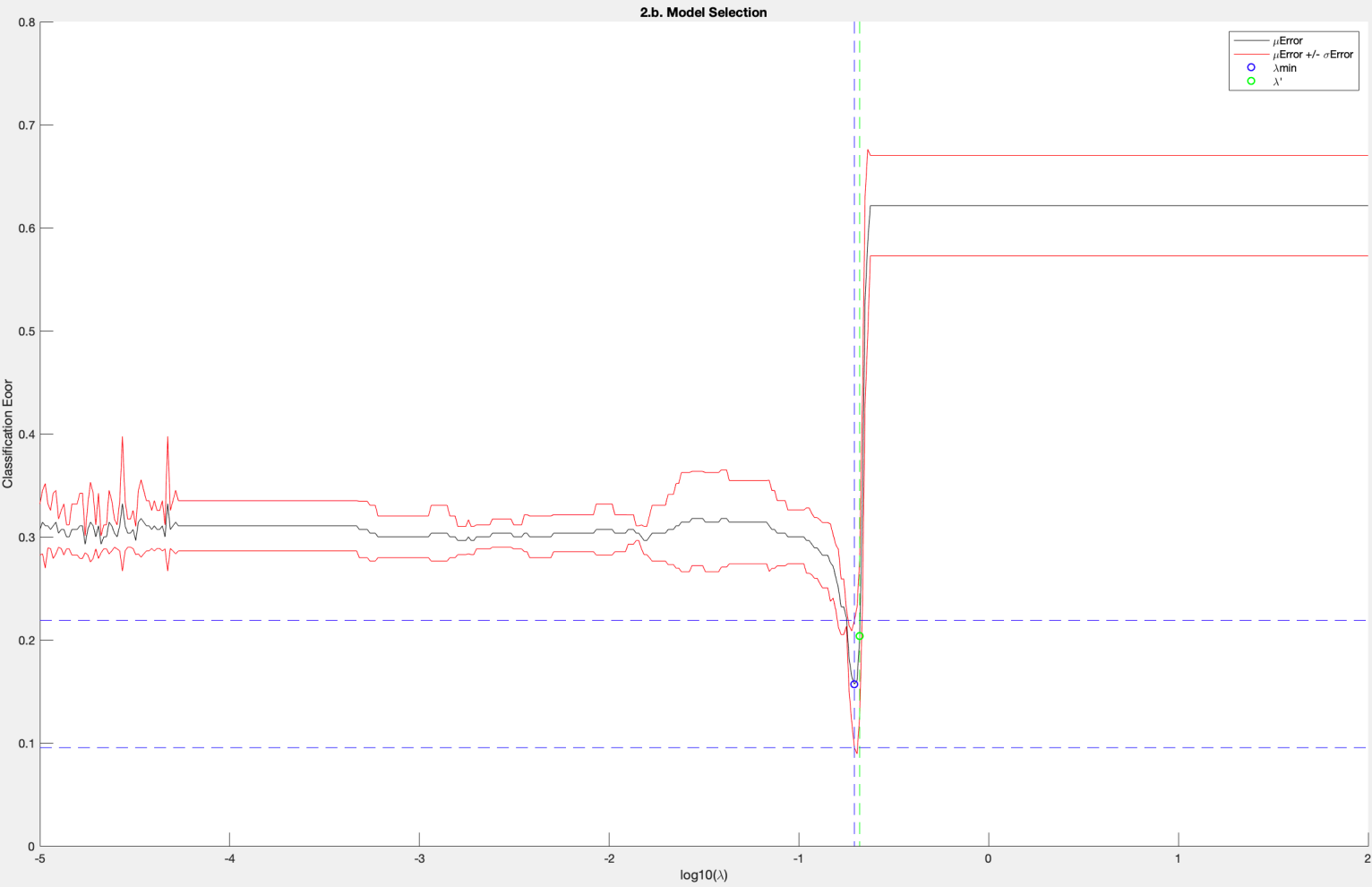
Iteration 1:



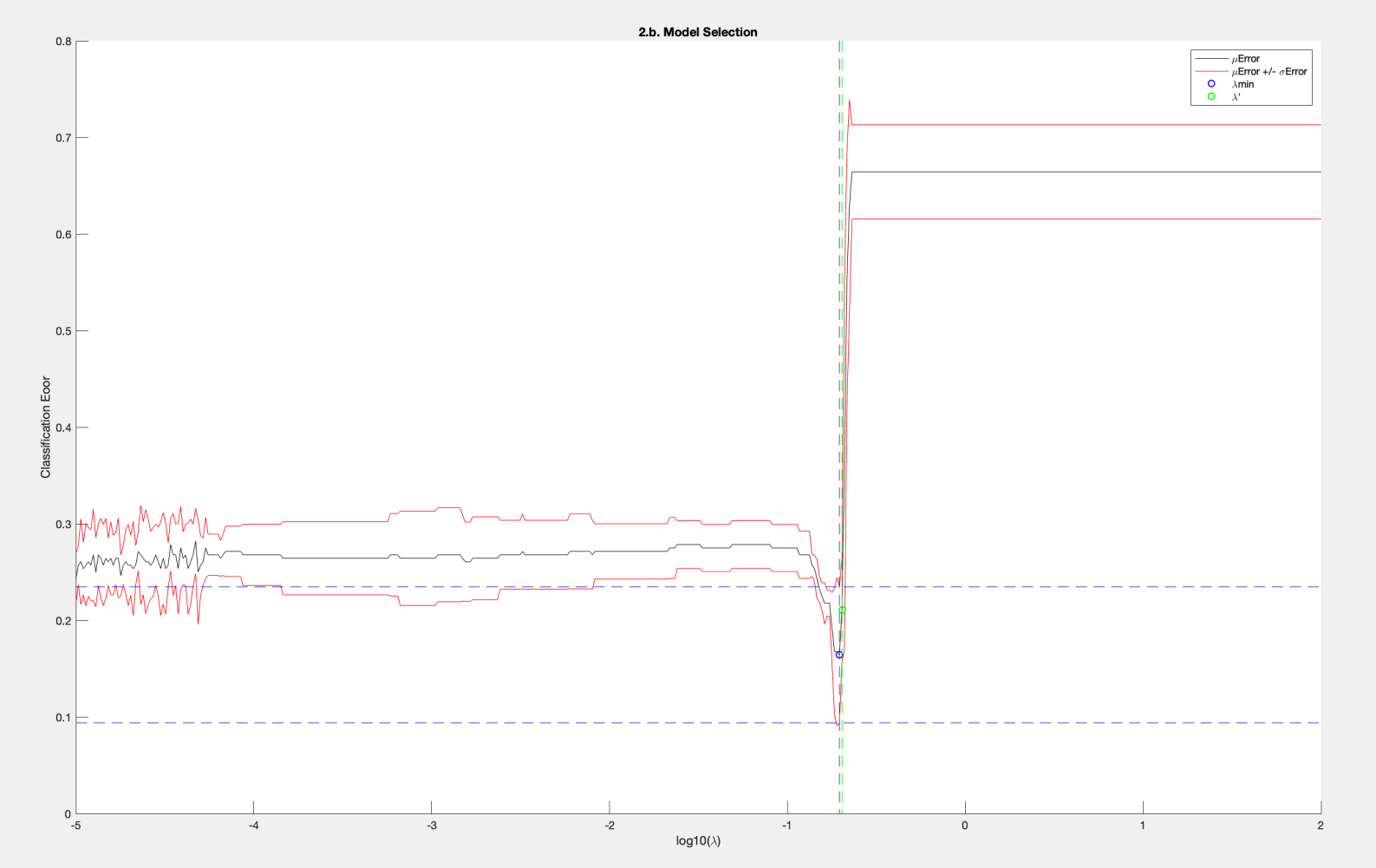
Iterations 2:



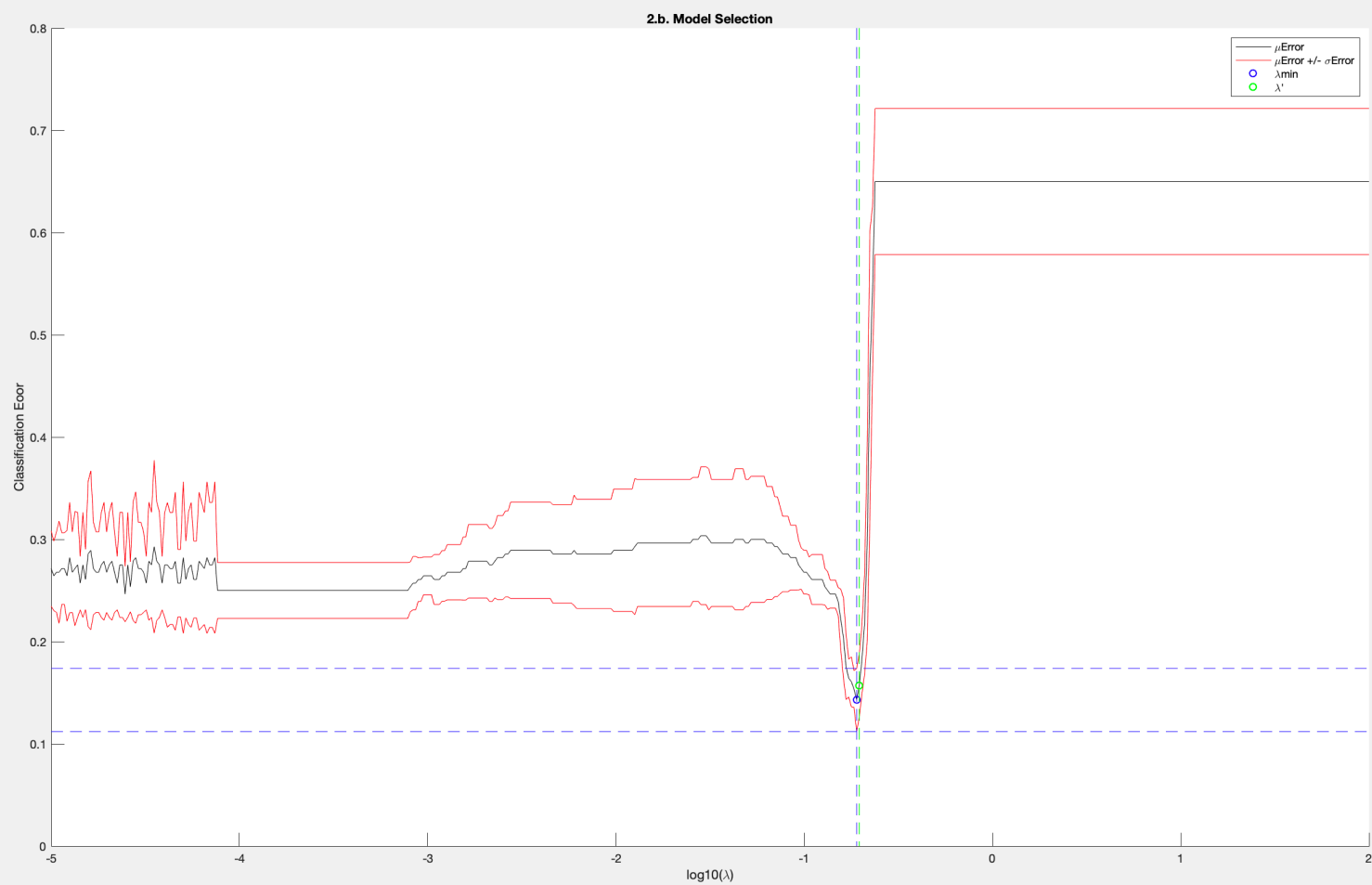
Iteration 3:



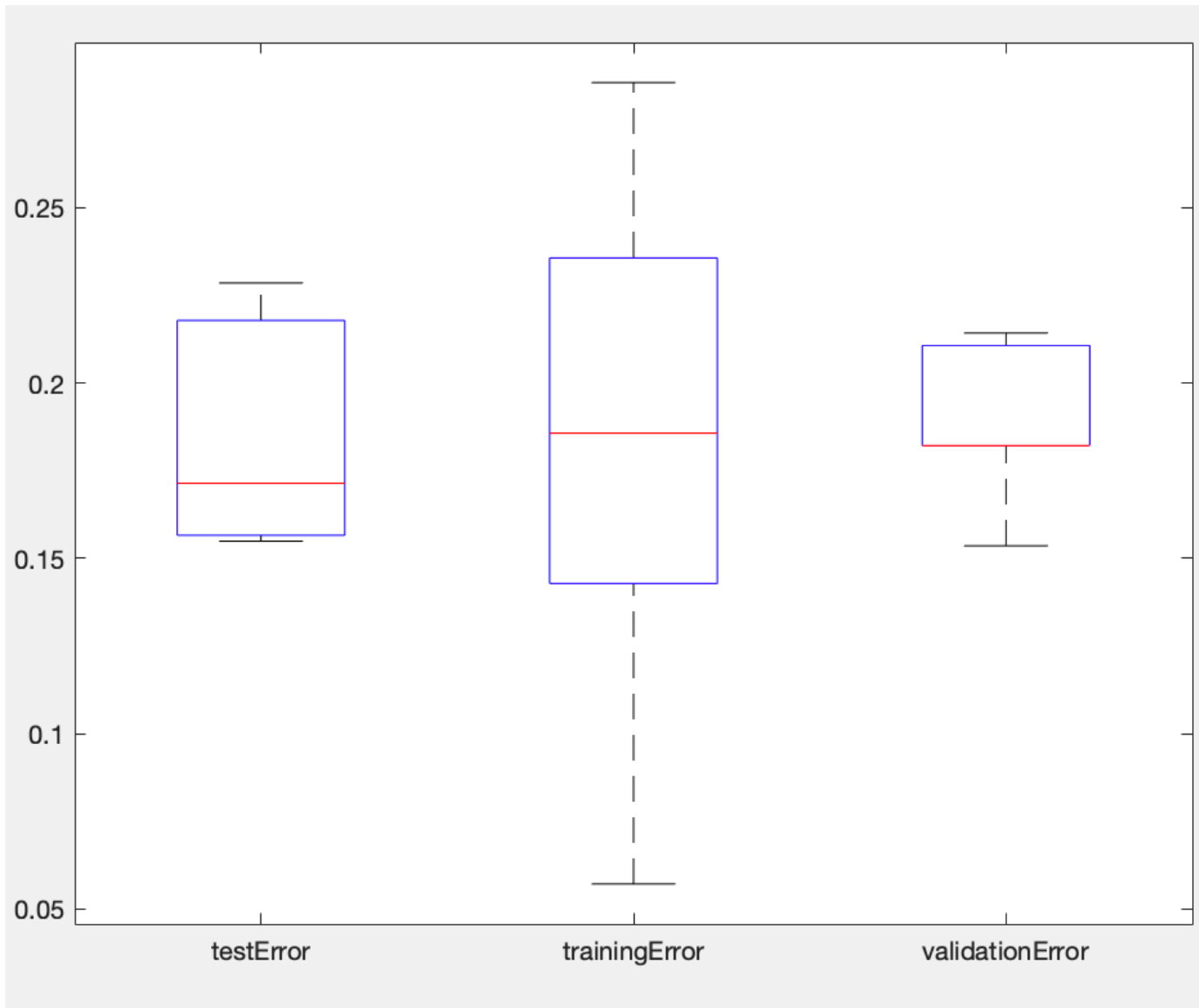
Iteration 4:



Iteration 5 :



1.c



Discussion:

The three graphs have similar mediums of errors.

However, test errors have a smaller span than the other two types of errors. This means that test error has smaller variance than training errors.

We could notice that validation error has smaller ranger and is closer to test error than training error. It suggests that validation error (cross validation) reflects test errors more accurately.