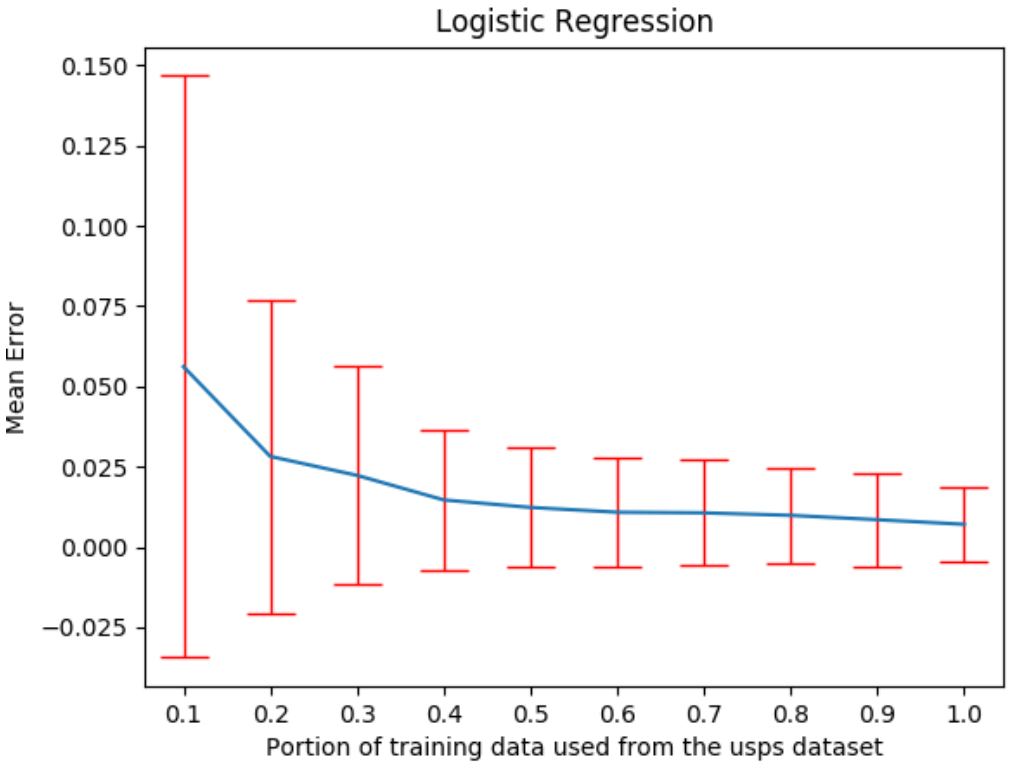
CSCI B-555: Machine Learning
Programming Project 3 Report

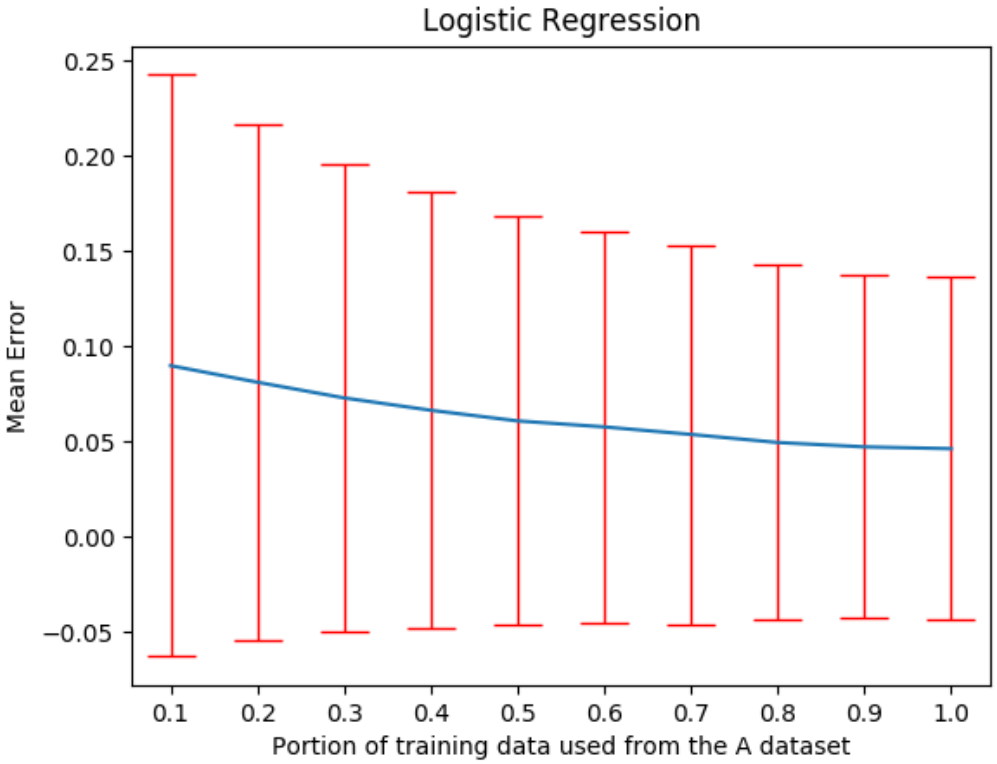
Aniruddha Patil (2000578987)

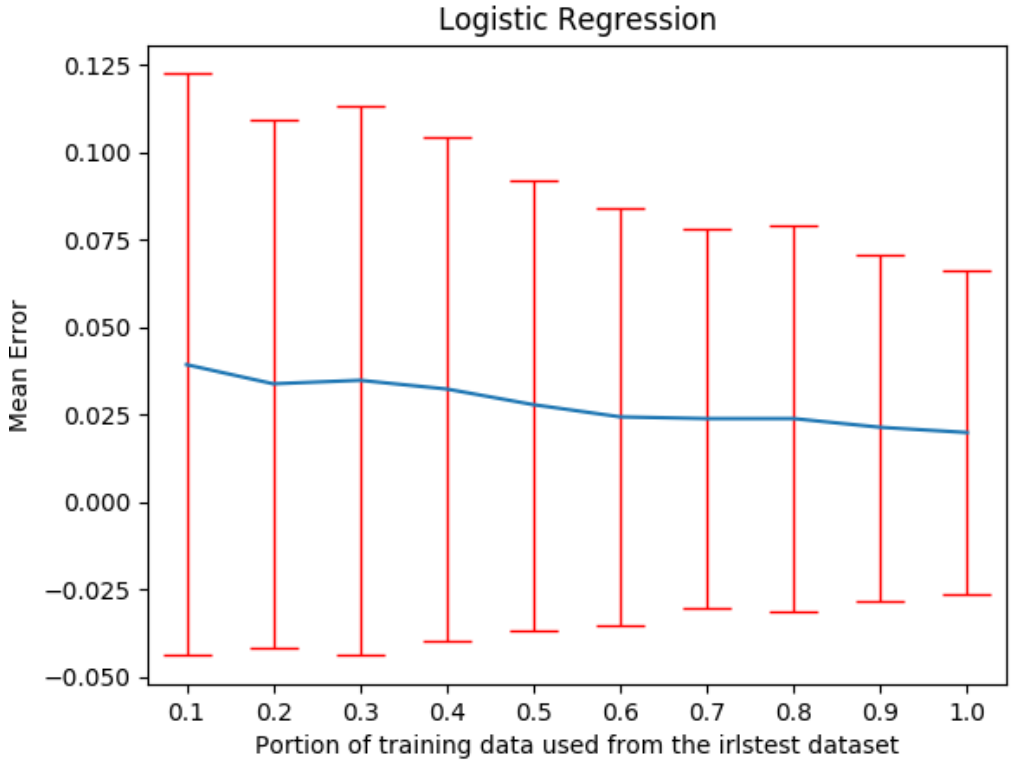
11/6/2019

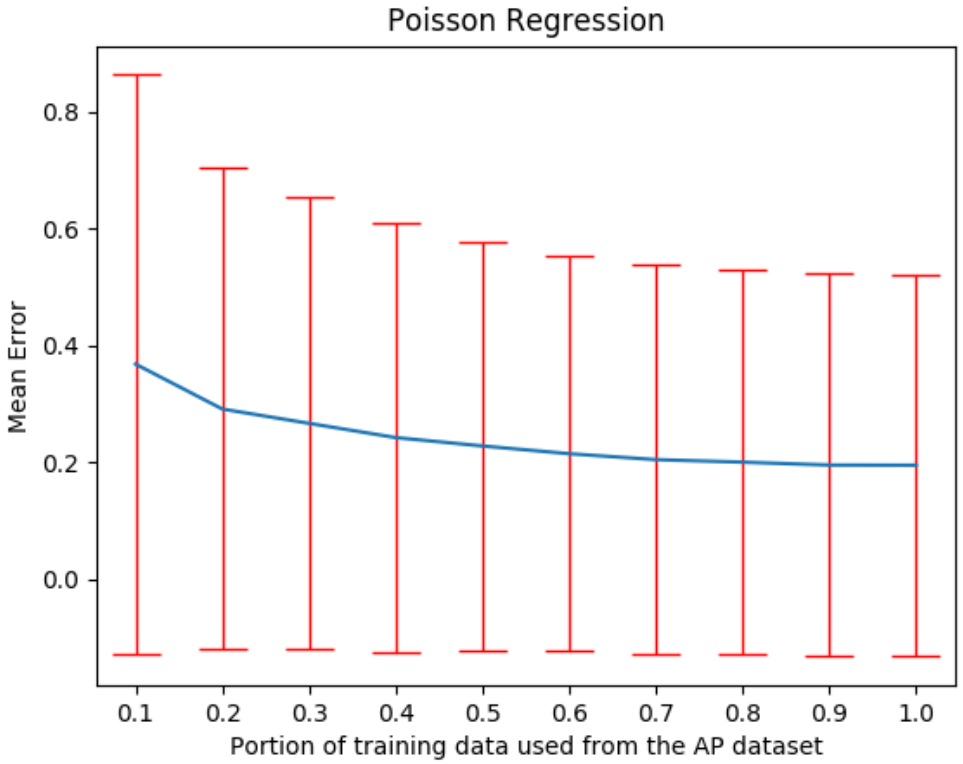


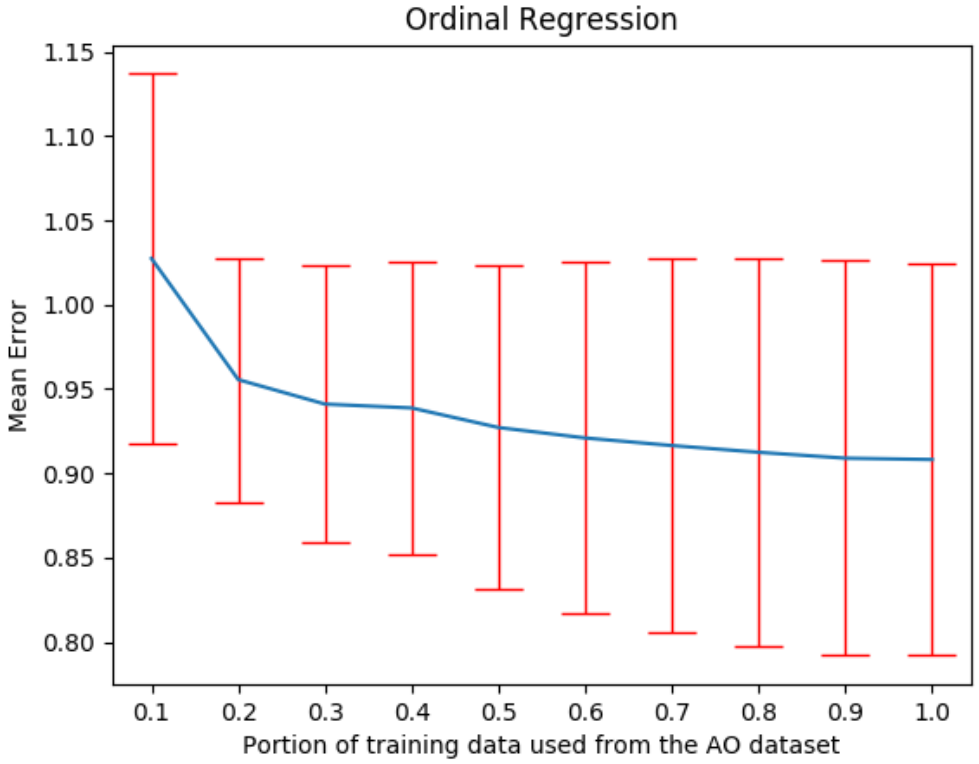
Dataset	Runtime (sec)	Iterations
A	5.733	6.83
irlstest	0.377	6.01
usps	13.325	8.397
AP	3.513	3.867
AO	14.25	4.267

Table 1: Table of average runtimes and mean iterations to converge on given datasets


Logistic Regression
Mean Error
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Portion of training data used from the A dataset


Logistic Regression
Mean Error
0.125
0.100
0.075
0.050
0.025
0.000
−0.025
−0.050
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Portion of training data used from the irlstest dataset


Poisson Regression
Mean Error
0.0
0.2
0.4
0.6
0.8
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Portion of training data used from the AP dataset


Ordinal Regression
Mean Error
0.80
0.85
0.90
0.95
1.00
1.05
1.10
1.15
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Portion of training data used from the AO dataset

Model Selection

I tried α values from $0.1, 0.2, ..., 1, 2, ..., 10, 20, ..., 200$ and found that $\alpha = 0.1$ gives the least error when all the data in the train split is used in each of the dataset.

Conclusion

In all the tasks except ordinal regression, we can see the standard deviation of the errors reducing as we use more samples for training. This is because more training samples helps us learn a better representation.

The standard deviation of the errors in the ordinal regression task might be high as the model is more sensitive than the others.