Machine Learning (SS20)

Homework-4

Priyanka Goenka

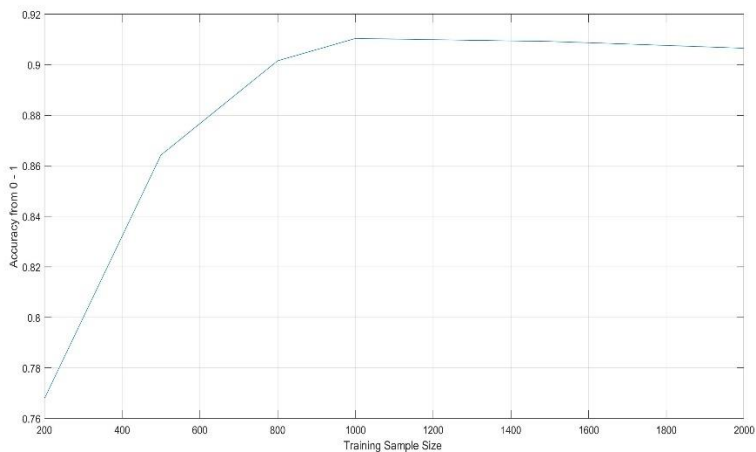GitHub Link for all codes of homework 4: https://github.com/PriyankaGoenka/CSE847-Homework4

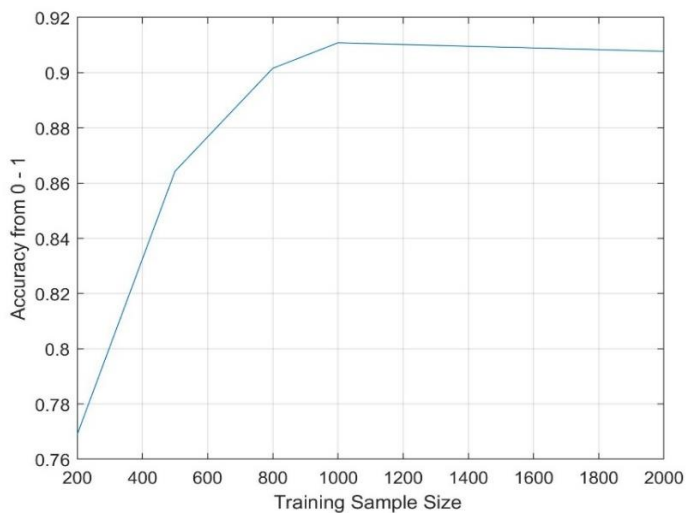Question 1: Logistic Regression

All the graphs represented below, shows different accuracies for different learning rate.

x-axis is the training sample number and y-axis represents the accuracy ranging from $0 - 1$.
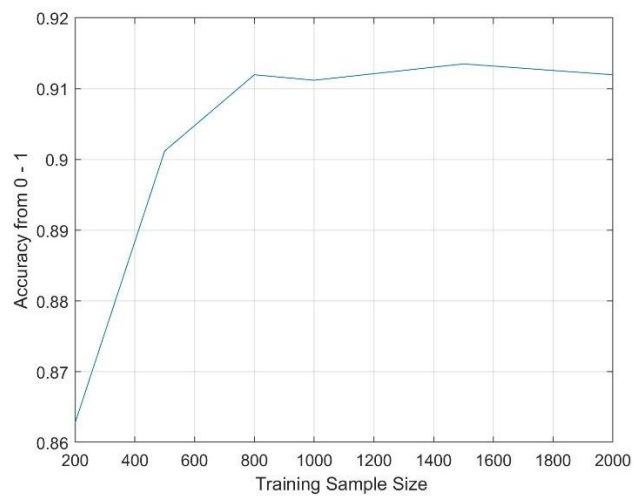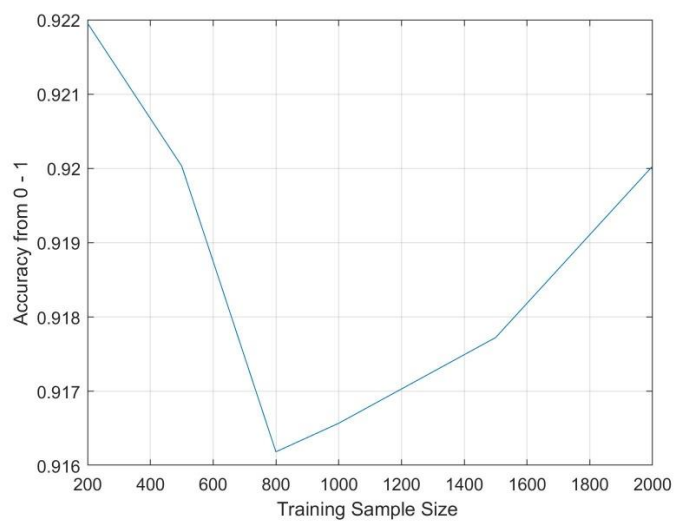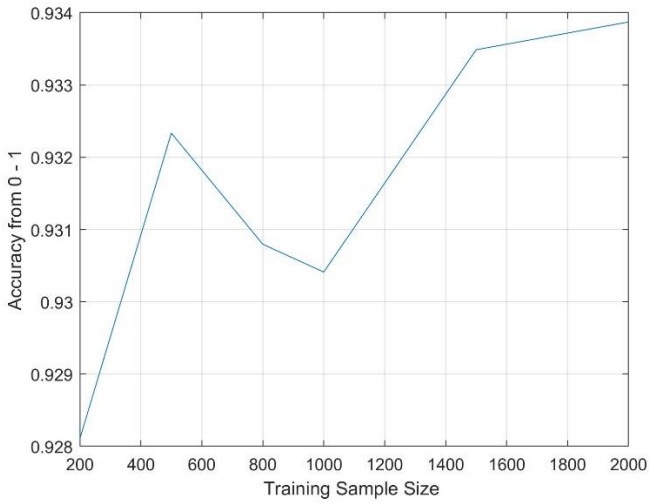
1) Learning Rate: 0.00001
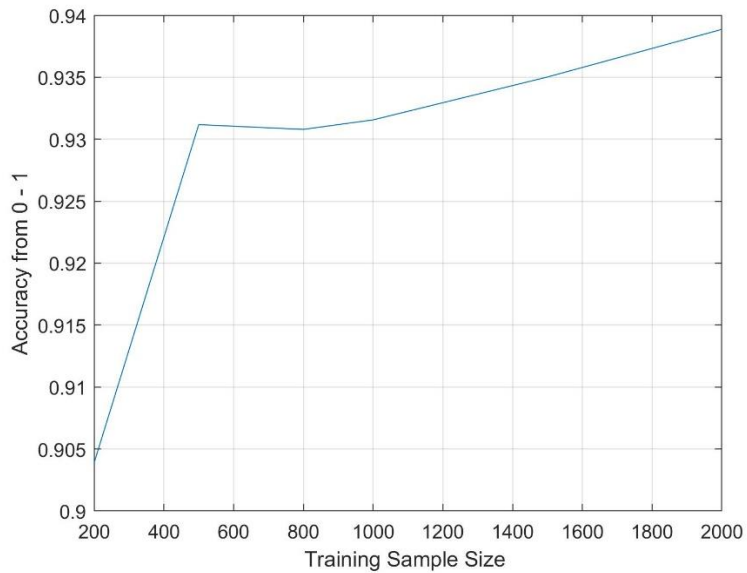

2) Learning Rate: 0.0001

3) Learning Rate: 0.001

4) Learning rate: 0.01

5) Learning rate: 0.1

6) Learning rate: 1.0

From all plots with different learning rates, we can observe that all graph converges to an accuracy above 90%.
In this data set, the learning rate:1.0 and training size = 2000 gave the best accuracy.

GitHub Link: https://github.com/PriyankaGoenka/CSE847-Homework4/blob/master/LogisticRegression.m

Question 2: Sparse Logistic Regression

Testing Parameters : 1e-8 , 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1
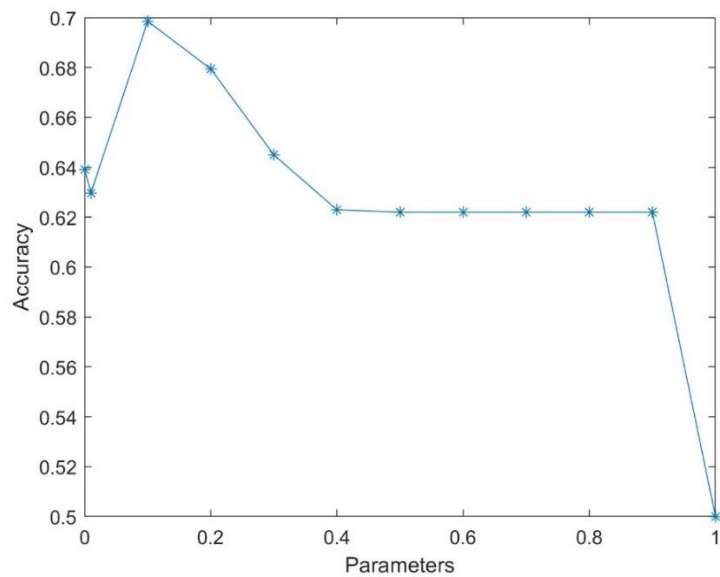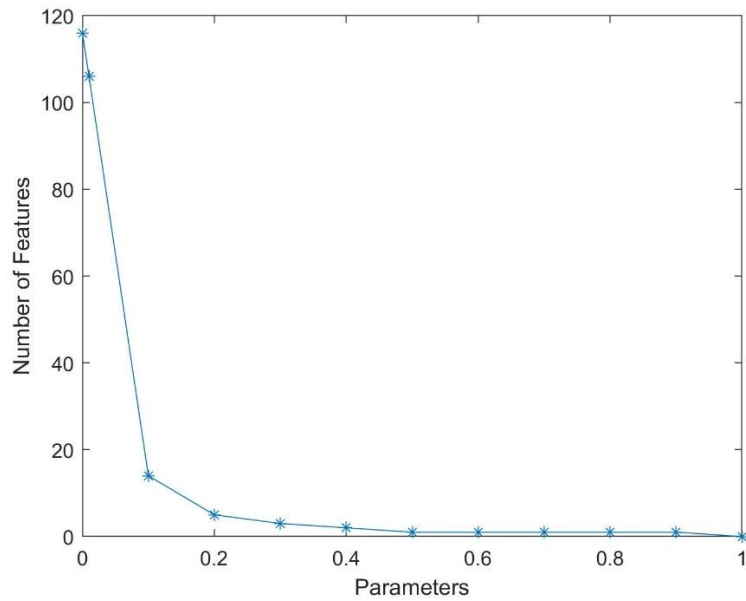
From this subset of parameters, we can conclude that the best accuracy occurs when the parameter is 0.1 and number of features is 14.

<u>Testing Parameters</u>: 1e-8, 1e-5, 5e-5, 1e-4, 5e-4, 1e-3, 5e-3, 1e-2, 5e-2, 0.1, 0.2, 0.3
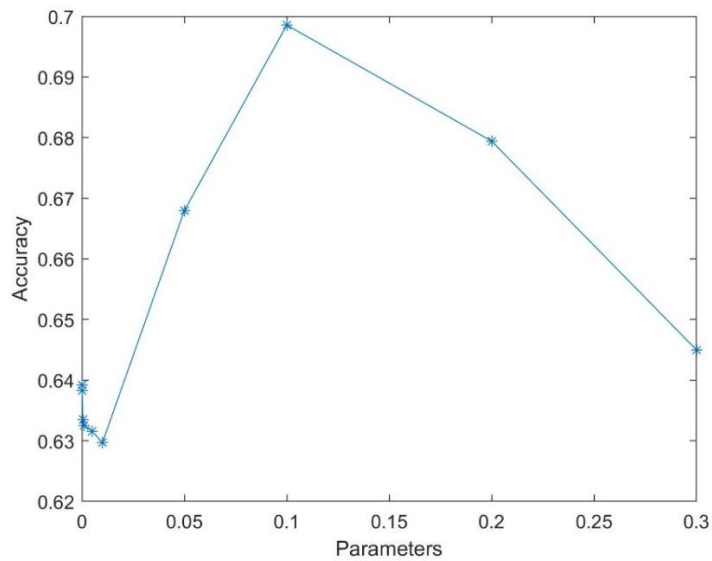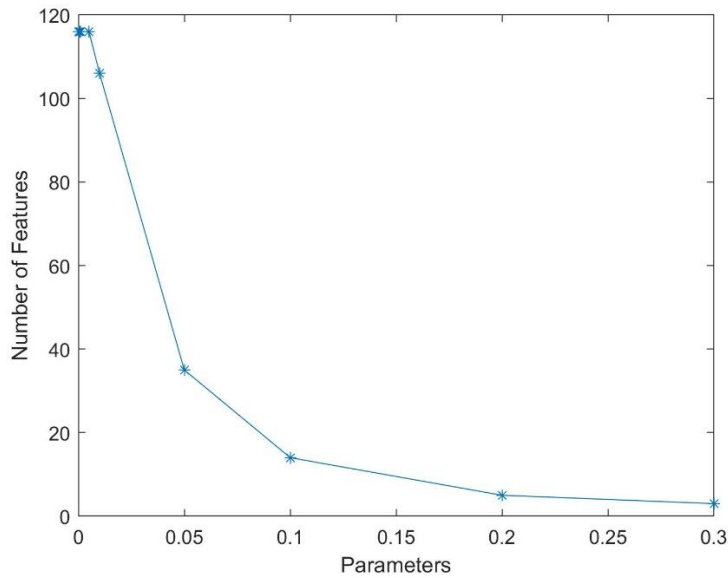
This subset of parameters shows the zoomed version of the above graph. We conclude that, the best accuracy obtained is 70% when the parameter value is 0.1 and the number of features is 14. Other 110 features can be ignored.

GitHub Link: https://github.com/PriyankaGoenka/CSE847-Homework4/blob/master/alzheimers.m

THANK YOU!