Rushabh Kanadiya
180070026

1.1

2.1

Here, 1-vs-rest can be directly used as a model that outputs the "class memberships" of the given data point, however for 1-vs-1 we need to deduce the class memberships from the output of the $C*(C-1)/2$ different classifiers (for example we can assign the most predicted class among all the classifiers to data point, or if they predict the class memberships, we could take the argmax for the sum of scores for all the classes). Although both the methods are quite easy to implement, 1-vs-1 approach should give better results because the model is more complex. However on draw back is that it doesn't scale well with data, because of the number of perceptrons/classifiers it uses.

3.1

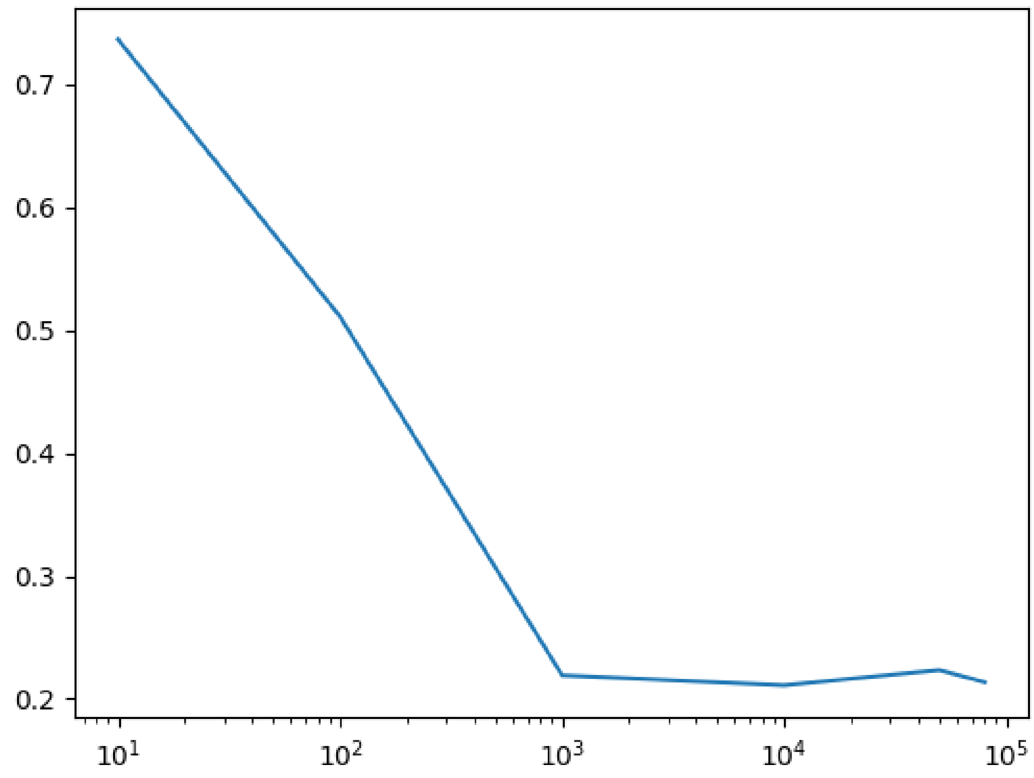
a) Increasing the value of lambda will decrease the variance and increase the bias. Because of more regularization, the model becomes less sensitive to variations in the data points, and more entries of w (weight vector) tend to zero.

b) Adding more training points to the perceptron should decrease the variance. Geometrically, perceptron algorithm tries to find a hyperplane that separates the two classes(the margin). The data is sampled from the universe with some underlying distribution on the X itself(Randomness introduced by some underlying force / noise in the measurements). As we feed more and more data in the perceptron, the more experienced it becomes, when a large number of data points are accumulated we can say that the perceptron has seen almost all the examples that there could be, in this sense, the variance should decrease because adding more data points or training the perceptron on a different set of sample points with the same cardinality should give a similar experience to our perceptron, and thus the resulting hyperplane should not differ much from the previous one.

c) NA

3.2

a)



b) As the degree increases, the bias decreases. However this causes overfitting by making the predictor more complex than needed and thus increasing the variance (making the output highly dependent on the sample set).

As expected the train mse decreases but the test error first decrease and then increases with the degree. The right bias variance tradeoff is achieved at degree 5, where the test set mse is the least.

Graph of Test error vs degree