BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

HYDERABAD CAMPUS

BITS F464 MACHINE LEARNING

ASSIGNMENT - II

Donkada Vishal Dheeraj	2018A7PS0239H
Thakur Shivank Singh	2018A7PS0439H
Pranav Reddy Pesaladinne	2018A7PS0238H

Comparative Analysis of Classification Algorithms

Models:

Fisher Linear Discriminant

Linear Perceptron,

Naive Bayes,

Logistic Regression,

Artificial Neural Networks,

Support Vector Machines

The models are imported from Sklearn Package with default hyperparameters and a 7 Fold Cross Validation is used over the dataset. Train and Test accuracies are calculated over each fold for all the six algorithms and represented below.

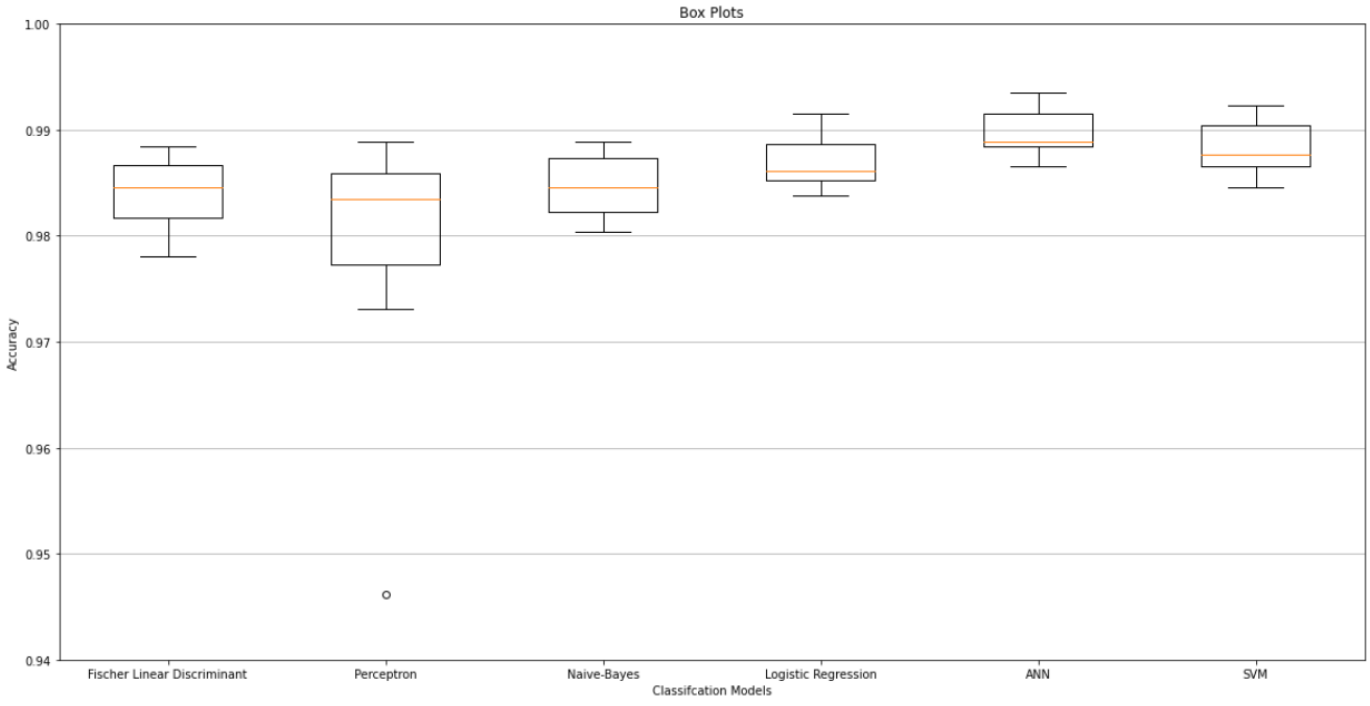
Train Accuracies over 7 Folds

	Linear Discriminant	Perceptron	Naive-Bayes	Logistic Regression	ANN	SVM
0	0.984474	0.987682	0.985372	0.987425	0.989671	0.988773
1	0.983576	0.980753	0.983512	0.986335	0.988773	0.988259
2	0.983255	0.982165	0.983704	0.986784	0.989286	0.987939
3	0.984089	0.982165	0.984603	0.986976	0.989350	0.988452
4	0.984153	0.971386	0.985501	0.987618	0.989863	0.989029
5	0.985052	0.941682	0.985308	0.987746	0.990697	0.989286
6	0.984090	0.985373	0.984732	0.987426	0.989543	0.988966

Test Accuracies over 7 Folds

	Linear Discriminant	Perceptron	Naive-Bayes	Logistic Regression	ANN	SVM
0	0.980370	0.985373	0.980370	0.985758	0.988838	0.987683
1	0.988453	0.981524	0.988838	0.991532	0.991917	0.991147
2	0.988453	0.988838	0.987683	0.989222	0.993457	0.992302
3	0.984604	0.983449	0.986913	0.988068	0.991147	0.989607
4	0.983064	0.973056	0.984604	0.984604	0.988068	0.986143
5	0.978060	0.946112	0.980370	0.983834	0.986528	0.984604
6	0.984983	0.986523	0.984213	0.986138	0.988833	0.986908

Box Plots



Interpretations:

From the Box Plot it can be seen that the **Artificial Neural Network** has higher accuracy than others across 7 folds with higher median value among all the algorithms. This may be because ANN can learn non-linear relationships better and has a lot of flexibility to fit complex data. This provides a better model to predict efficiently.

Perceptron seems to perform poorer among all the models as the median line lies lower than the other algorithms and an accuracy datapoint is outside the box. This may be because the algorithm provides only a linear boundary to classify data and hence works badly with complex relationships. Another issue is that it is similar to Stochastic Gradient in the sense that every update to the weight vector is done using only a single sample from the dataset each time. This may reduce the loss with respect to one sample but the overall loss may increase.