

Comprehensive Comparison

-Rahul Mehndiratta 2017A7PS1479H

-BVS Ruthvik2019A7PS0017H

-R Vedang. 2019A7PS0150H

Problem Statement:

1. As for this assignment, you need to do a comparative study and analysis of the ML models you have studied till now, i.e. Fisher Linear Discriminant, Linear Perceptron, Naive Bayes, Logistic Regression, Artificial Neural Networks and Support Vector Machines. Note, only for this assignment you can use the Sklearn library to directly import models/methods and use them.

2. For all the models, use 7-fold cross validation and generate a box-plot over the test set accuracy over each fold. Visualize the all box-plots in a single image. That is, the image must contain six box-plots, one for each model with the box-plot denoting the variation of test set accuracy over each fold.

What needs to be documented:

1.A comparative analysis of the models and their accuracies (train and test).

```
Mean accuracy - svm for test set = 0.9903218024197908
Mean accuracy - svm for train set = 0.9905324921586549
```

```
Mean accuracy - nb for test set = 0.9836129007232456
Mean accuracy -nb for train set = 0.9836770258672074
```

```
Mean accuracy -logistic_regression for test set = 0.9892220341882013
Mean accuracy -logistic_regression for train set = 0.989267714838196
```

```
Mean accuracy -fda for test set = 0.9865825142081011
Mean accuracy -fda for train set = 0.9867839853432158
```

```
Mean accuracy -linear_perceptron for test set = 0.9789943018535078
Mean accuracy -linear_perceptron for train set = 0.9801117549597327
```

```
Mean accuracy for -ann test set = 0.9903218447666163
Mean accuracy for -ann train set = 0.9907524533295239
```

2.The model that performed best and one the that performed worst. (Do mention reasons why that certain model may have given the best or worst results.)

SVM and ANN were the models that had the best performance in terms of test set accuracy with accuracies of approximately 99.03% each. Linear perceptron didn't do so well as compared to other models with an accuracy of 97.8%. The reason for these observations probably has to do with the nature of decision boundaries created by these models. ANN and SVM have the capability to create non-linear decision boundaries which helps in better classification where linear perceptron creates a linear decision boundary which is clearly not as efficient. Perhaps for this reason(that linear perceptron is limited to creating linear decision boundaries), the box plot of it is more "spread out", since the accuracies over each fold are at the mercy of the extent of linear separability of the training data in that fold.

3.The image containing box-plots for each model.

