BITS F464 - Machine Learning

Assignment-2C: Comprehensive Comparison

Made by -

Tanay Gupta (2018AAPS0343H)
Anusha Agarwal (2018A3PS0032H)
Kriti Jethlia (2018A7PS0223H)

Model Results

- Fisher Linear Discriminant
    - Train Accuracies

      Over cross folds : 0.986, 0.987, 0.986, 0.986, 0.986, 0.986

      Mean : 0.986

    - Test Accuracies

      Over cross folds : 0.984, 0.982, 0.990, 0.988, 0.985, 0.987, 0.988

      Mean : 0.986

- Linear Perceptron Results
    - Train Accuracies

      Over cross folds : 0.976, 0.964, 0.988, 0.983, 0.867, 0.965, 0.974

      Mean : 0.956

    - Test Accuracies

      Over cross folds : 0.978, 0.975, 0.979, 0.867, 0.960, 0.970, 0.956

      Mean : 0.954

- Naive Bayes Results
    - Train Accuracies

      Over cross folds : 0.983, 0.984, 0.983, 0.983, 0.983, 0.983, 0.983

      Mean : 0.983

    - Test Accuracies

Over cross folds : 0.984, 0.979, 0.984, 0.984, 0.984, 0.984, 0.983

Mean : 0.983

- Logistic Regression Results
  - Train Accuracies

    Over cross folds : 0.989, 0.989, 0.989, 0.989, 0.989, 0.989, 0.989

    Mean : 0.989

  - Test Accuracies

    Over cross folds : 0.990, 0.989, 0.989, 0.988, 0.986, 0.988, 0.989

    Mean : 0.989

- SVM
  - Train Accuracies

    Over cross folds : 0.989, 0.988, 0.989, 0.989, 0.989, 0.989, 0.989

    Mean : 0.989

  - Test Accuracies

    Over cross folds : 0.987, 0.991, 0.990, 0.986, 0.986, 0.988, 0.990

    Mean : 0.988

- ANN
  - Train Accuracies

    Over cross folds : 0.917, 0.988, 0.971, 0.982, 0.982, 0.973, 0.982

    Mean : 0.971

  - Test Accuracies

    Over cross folds : 0.913, 0.987, 0.974, 0.979, 0.985, 0.976, 0.986

    Mean : 0.972

Comparison :

- Both Logistic regression and SVM seem to perform well enough with very high testing accuracy and low variance across the folds as visible in the box plot. Higher accuracy of SVM can be attributed to its ability to avoid overfitting.

- Perceptron performs worst with very high variance, followed by Naive Bayes. Since the perceptron algorithm depends a lot on the order in which the training examples are encountered and is also affected by initialized weights, it tends to give a higher variance. As for Naive Bayes, the features might not be independent of each other. Logistic regression, which performs better on the given dataset, on the other hand, doesn't assume that the features are independent.
- Neural networks for this binary classification problem might be performing worse than logistic regression because neural networks are more difficult to train and are more prone to overfitting than logistic regression. They require more data to train in comparison to Logistic regression.
- Overall Logistic regression tends to perform better than LDA, which is very sensitive to outliers.

Box Plot :