**Machine Learning**

**Laboratory Experiment - Week11**

*Submitted by: Yash Mittal (12012002) & Rahul Shrimali (12012008)*

***11.2***: For the same data set, you have done classification using 4 methods (Logistic Regression, Dec-Tree, kNN and SVM). Analyze and report which performed best. Explain the possible reasons for that model performing best.

***Ans****:* As per the output of the code for various models trained, we have the following metrics to report -

| **Model** | **Accuracy (f1-score)** |
| --- | --- |
| **Logistic Regression** | 0.66 |
| **KNN** | 0.71 |
| **Random Forest** | 0.77 |
| **Ada-boost** | 0.65 |
| **SVM - Linear Kernel** | 0.67 |
| **SVM - Poly Kernel (with degree 3)** | 0.64 |
| **SVM - RBF Kernel** | 0.94 |
| **SVM - Sigmoid Kernel** | 0.45 |

From the above table we can easily see that the SVM Model with RBF kernel performs the best. Probable reason for this model performing best is , it allows model to have non-linear boundaries and according to this dataset the linear boundaries may not give the best result, as we know logistic regression and svm linear kernel they allow only linear boundaries which may not fit perfectly in this scenario, whereas in comparison to knn we know that knn does not perform that well in case of high-dimensional data whereas SVM with rbf kernels is more efficient in that case. Random forest

***11.3****:* Whether any ensembling technique performed better than the model reported in reply to above Q 11.2. If so, report the results.

***Ans****:* No, SVM with RBF (Radial Base Function) which is not an ensemble technique performs best. The accuracy results of SVM with different kernels are as follows:

| **Kernel** | **Accuracy (f1-score)** |
| --- | --- |
| **Linear Kernel** | 0.67 |
| **Poly Kernel (with degree 3)** | 0.64 |
| **RBF Kernel** | 0.94 |
| **Sigmoid Kernel** | 0.45 |

One thing to notice here is, if we compare the kernels other than “RBF” i.e. linear, poly (with degree 3) and sigmoid, then we can see that they are not performing better than the models, since the models have better accuracy then these kernels as reported in the table given with Q 11.2. Based on the data in the said table it can be seen that ***”Random Forest”*** performs the best with a f1-score of “0.77”, this shows that the performance of ensembling techniques being better than the models is based on the selection of suitable kernel function as per the given data.

**NOTE:** *The dataset used for the above analysis is:* ***hr\_data.csv*** *provided in Week 9.*