# 

**1:**

## Introduction

The main objective of performing this assignment is to create classification models, train, test, tune and compare between them using evaluation metrics.

A data set is provided with 18 features. The first 8 features are kinematic properties measured by particle accelerator and the remaining 10 features are functions of the first 8 attributes in determining the existence of a particle.

For the first two models it is trained with all the attributes and the next two models utilizes feature selection thus dropping 4 attributes from its 8 kinematic values of data.

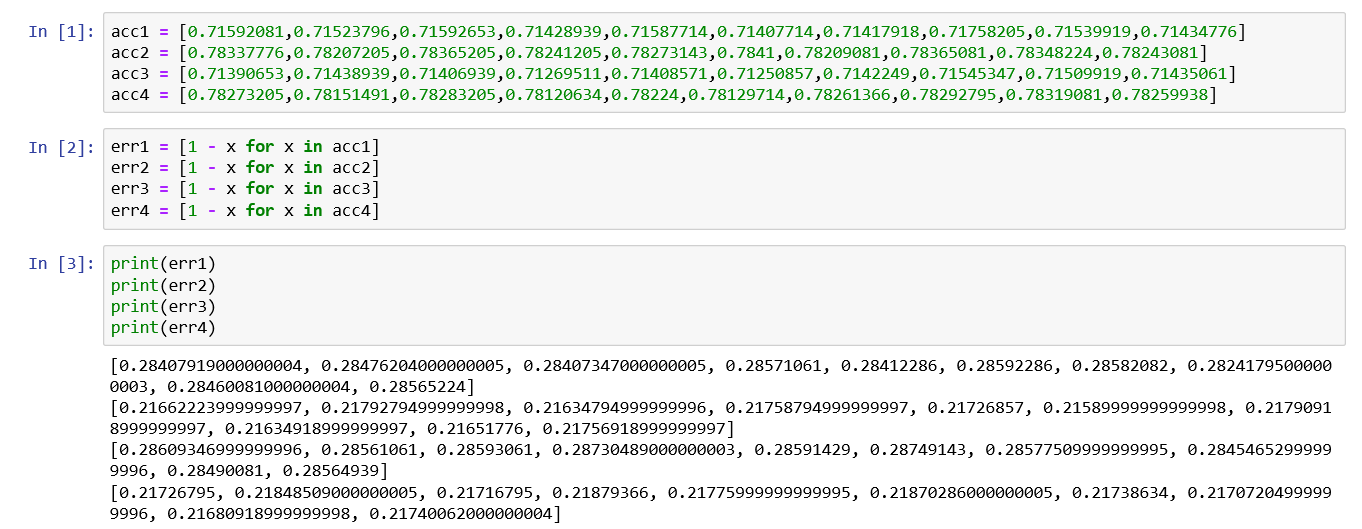
The data set is related to Particle accelerators. It is the detection of fundamental particles, their classification if the measurement data is by a particle or background noise needs to be performed by the classifier. The classifier falls into two categories of the class label where 1 corresponds to presence of particle and 0 corresponds to absence of particle.

Once 4 models are trained and tested it needs to be compared using hypothesis testing.

Sections 1.2 ,1.3,1.4 and 1.5 are enclosed as a ipynb notebook. Section 1.2 is a decision tree classifier for all 18 attributes. Section 1.3 is a bagging classifier for all 18 attributes. Section 1,4 is a decision tree classifier for selected attributes as given by the course leader. Section 1.5 is a Bagging classifier for selected attributes of the classifier.

**Note 1.2 ,1.3, 1.4 and 1.5 are enclosed**

**1.6 Comparative evaluation of the Classifiers**



**Figure** Computing error values from accuracies obtained in the above sections

The above figure consists of codes to compute errors required to perform comparisons. Accuracies obtained in the previous sections after cross validation are put into a list and subtracted from 1 to get error (since error = 1- accuracy). Note that the subfix of acc and err represent respective models. So, acc1 is for decision tree model , acc2 for bagging classifier , acc3 are accuracies of selected decision tree and so on.

**T test**

T test is used to compare t models. T- values obtained from t test determine how close the two models being compared are. The t values are then compared with confidence limit z, if t>z or t < -z , then the null hypothesis is rejected. The formula to T values is given by

Or we can utilize scipy’s ‘ttest\_rel’ which does the same. Six different inter comparisons can be performed on the four models.



**Figure** Performing Ttest on scipy

Since n folds = 10, degree is ‘k-1’ = 10 – 1 = 9.

Assuming significance level as 0.5, we have **z = 2.26** ( from t distribution table)

Let’s consider null hypothesis to be that **Both models are similar.**

Using t- values from the above figure,

1. Model 1 vs 2

t = 202.18

202.18 > 2.26 i.e t > z

Null hypothesis rejected.

1. Model 1 vs 3

t = -4.47

-4.47 < -2.26 i.e t < -z

null hypothesis is rejected.

1. Model 1 vs 4

t = 224.75

which is larger than z

null hypothesis is rejected

1. Model 2 vs 3

t = -183.13

which is lesser than -z

Null hypothesis rejected.

1. Model 2 vs 4

t = -2.42< -2.26

Thus , t is lesser than -z

Null hypothesis rejected

1. Model 3 vs 4

t = 382.11

which is greater than z

Therefore, null hypothesis rejected.

Thus none of the models are similar to one another since all of them fail the null hypothesis. There are no two classifiers which are same however model 2 and 4 is the closest among them

**1.7 Conclusion**

After training, testing , evaluating and comparing four classifiers it can be said that each model is different from one another. Which means that they’re different from one another even though 1, 3 and 2 , 4 models had the same classifier, change in attributes caused the change in model. And though models 1,2 and 3,4 had same number of attributes the change in classifiers caused the model to change.

The performance of the model can be further be improved by

* Changing the classifier. Using Random forest or boosting ensemble methods
* Hyperparameter tuning. Using adequate amount of estimators and other parameter.
* Careful and calculated feature selection: The features selected for model 3 and 4 are random and unthought for, which is the reason why there is a difference of only .1 % in accuracy. Careful feature selection or combination of features based on their contribution can yield better results