B. Prabhakar

==========

201505618

==========

Assignment I

Data Sets used:

Iris	
Data Set Characteristics	Multivariate
Number of Instances	150
Attribute Characteristics	Real
Number of Attributes	4
Associated Tasks	Classifiation
No of Classes	3 1. Iris Setosa 2. Iris Versicolour 3. Iris Virginica
Wine	
Data Set Characteristics	Multivariate
Number of Instances	48842
Attribute Characteristics	Real
Number of Attributes	13
Associated Tasks	Classifiation
No of Classes	3
Tic-Tac-Toe Endgame Data Set	
Data Set Characteristics	Multivariate
Number of Instances	958
Attribute Characteristics	Real
Number of Attributes	9
Associated Tasks	Classifiation
No of Classes	2 1. Positive 2. Negative
In the *Tic-Tac-Toe* dataset all the given values of each of the feature are replaed with a unique integer and then the value of the features that are thus obtained are used as an input to the algorithm(KNN, K-Fold) for permorming the classification.	

What have I implemented ?

- Implemented the KNN algorithm with Random sub-sampling method and then with 5-Fold cross verification method.
- Tested the above mentioned algorithm with the Iris, Tic-Tac-Toe End game and Wine data-sets and calculated the accuracy. Performed the testing with each data-set for 10 times and obtained the mean and standard-deviation for the accuracy obtained in those trials.
- Plotted the Decision-Boundary for the Iris data-set by taking Sepal-width on X-axis and Petal-width on Y-axis.

Python Libraries that are used

- *matplotlib* for plotting the libraries.
- *numpy* for arrays.
- *Texttable* for printing the Confusion-matrix as a table.