CS 256: Homework1

SJSU ID: 012447851
Name: Mayuri Wadkar

Instructions to run the program from command line:

$ python mayuri_wadkar_knn.py <absolute_path_of_datasets> <dataset_name> <k>

Where,

<absolute_path_of_datasets> : Absolute path of the folder containing 10 fold test and training files, 10 each (Note : Please enclose the path in double quotes)

<dataset_name> : Name of the dataset being processed (Note: please enter lowercase)

<k> : Value of K

E.g.:

$ python mayuri_wadkar_knn.py "C:\Users\mayur\Google Drive SJSU\SJSU\CS 256\Homework1\haberman-10-fold" haberman 9

My python implementation requires following external libraries:

1. Matplotlib
2. Numpy

Output Description:

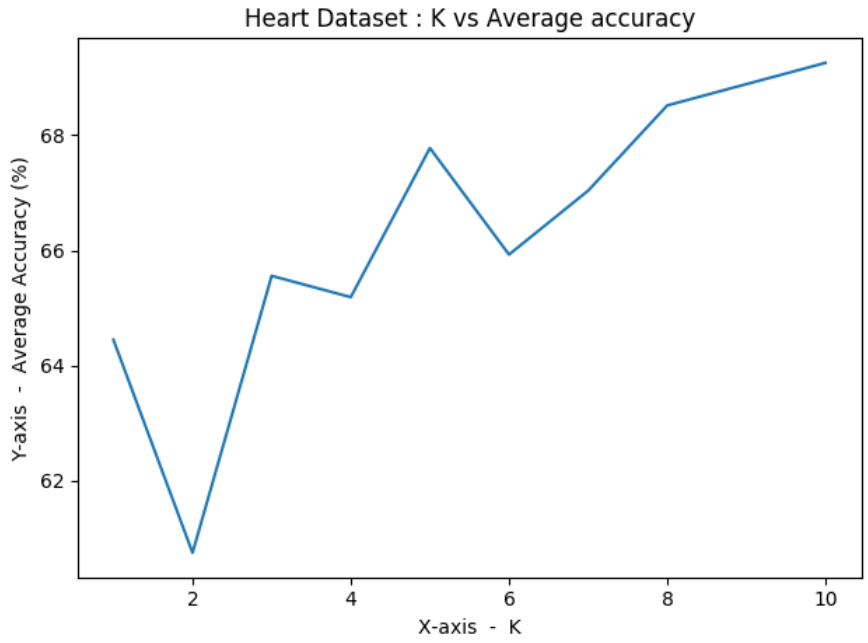
I have tested my KNN implementation on 5 datasets from KEEL and found out average accuracy of a classifier over all 10 folds for each dataset.

Please find on next pages the computation summary of Average Accuracies for different datasets.

Analysis of average accuracy of a classifier for different values of K:

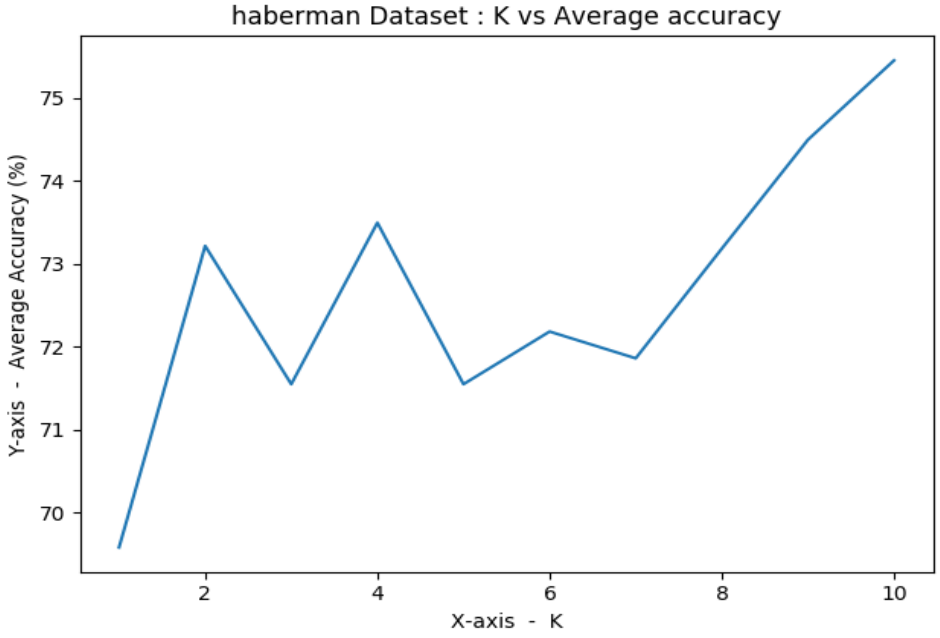
1. Heart

K	1	2	3	4	5	6	7	8	9	10
Average Accuracy	64.4444	60.7407	65.5555	65.1851	67.7777	65.9259	67.0370	68.5185	68.8888	69.2592



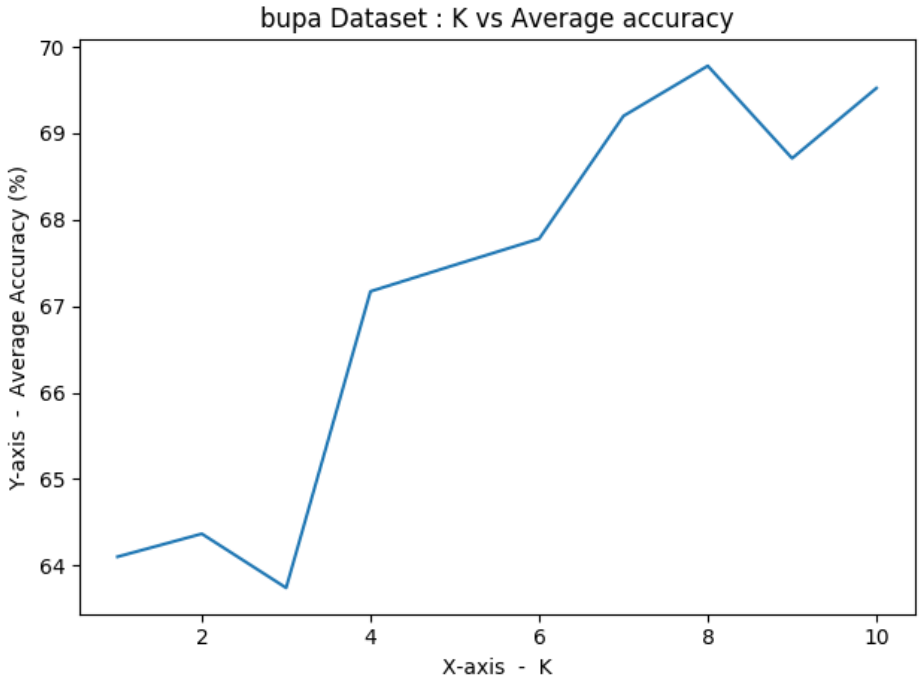
2. Haberman

K	1	2	3	4	5	6	7	8	9	10
Average Accuracy	66.6666	73.2150	71.5483	73.4946	71.5483	72.1827	71.8602	73.1827	74.4946	75.4516



3. bupa

K	1	2	3	4	5	6	7	8	9	10
Average Accuracy	64.1019	64.3675	63.7424	67.1710	67.4768	67.7793	69.1995	69.7793	68.7102	69.5220



4. Australian

K	1	2	3	4	5	6	7	8	9	10
Average Accuracy	61.4492	60.4347	64.0579	61.0144	64.6376	64.3478	67.9710	65.7971	67.6811	67.3913