CSE 512-HW3: Machine Learning Report and README

Name: AJAY GOPAL KRISHNA SBU ID: 112688765

1 Programming

1.1 Results

The dataset was split into 80-20. 455 points was used in training i.e to calculate the distance for the test point and 144 points was used in testing.

1.2 Observation and Best performance

K-neighbors	Test Size	Miss-classified - Error	Accuracy
21	144	6	94.7368

The best performance on a 80-20 split data was observed when I use $k = 21$, it gave an accuracy of 94.7% This was tested without shuffling the dataset.

Execute **python knn.py -d path-to-data -k 21 -t 0.8** to get the best performance

2 README

knn.py supports the below parameters:

Option	Description	Default-Value
-h, –help	show this help message and exit	
-d, –dataset	***path_to_dataset***	Breast_cancer_data.csv
-k, –knn	K-nearest Neighbors	3
-t, –train_size	Train-Test split. Give between (0-1)	0.8
-s, –shuffle	Shuffle the data set randomly	False

To run the k-nn algorithm with $k = 5$ and train-test split of 85% without shuffling the data, execute
python knn.py -d path_to_data -k 5 -t 0.85

Note to shuffle the data-set add **-s** or –**shuffle** flag while running the command.