Name : Debarati Das

Test Data set size: 10 and Train data set size : 140

The following table summarizes the test result that is submitted.

K value	Distance metric	K nearest neighbour classifier accuracy
K = 1	Manhattan	100%
	Euclidean	100%
	Supremum	90%
K = 2	Manhattan	100%
	Euclidean	100%
	Supremum	100%
K=3	Manhattan	100%
	Euclidean	100%
	Supremum	100%
K=4	Manhattan	100%
	Euclidean	100%
	Supremum	100%
K=5	Manhattan	100%
	Euclidean	100%
	Supremum	100%

Comments and observations:

1. We have divided the iris.csv into train.csv and test.csv based on logic that uses a random number generator with seed=31. This decides which data element will be part of test data and which is part of train data.
2. We have tried out with different seed values in the random generator and as a result, we have found that the train and test dataset changes to an extent leading to a slightly different values of the accuracy.
3. We have also tried out different test and train division i.e. for example 50 test and 100 train data items. That also leads to slightly different prediction accuracy of the classifier.
4. In general, as number of neighbours increases, the accuracy increases as we have more number of nearest neighbours to participate in the “voting”
5. The effect of distance metric is NOT very prominent but in general, the supremum distance measure based classifier seems to less accurate compared to the other.
6. The somewhat similar accuracy is due to the fact that the measurements are not vastly different in value. Had it been so, the Euclidean distance (due to the squared difference being taken) will have a bigger impact in accuracy. So, we can expect more different

accuracy with a much larger variation in each co-ordinate of the dataset.

7. Supremum metric is bad for small K clearly but there is too little test data for actual verification.