Nearest Neighbours Algorithm

Abrie Greeff
B.Sc Hons (Computer Science)
Department of Computer Science
University of Stellenbosch

June 26, 2006

Contents

1   Question 1

This question was implemented in Java 1.5.0. To execute the program type *java knn* in the console. This will present you with a list of possible parameters that can be passed to apply the k-nearest-neighbours algorithm.

2   Question 2

2.1   Part a

This provided an error rate of 100% which is unacceptable. The reason for this is the way the iris data set is organized. Not one of the possible target labels were specified in the training set and thus the algorithm were not able to predict the test data. This shows us that the nearest neighbours algorithm is very dependant on the quality of the training set. The following output (only part of it) was generated for this question.

```
Confusion matrix
Iris-setosa      0 0 0
Iris-versicolor  0 0 0
Iris-virginica   41 9 0
```