CAB320 – Assignment 2

Machine Learning

### Report

Submitted By:

n9400001 Moira Quinn

n7226209 Maurice Cafun

n9708651 Christopher O'Rafferty

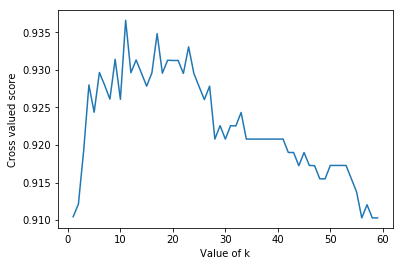
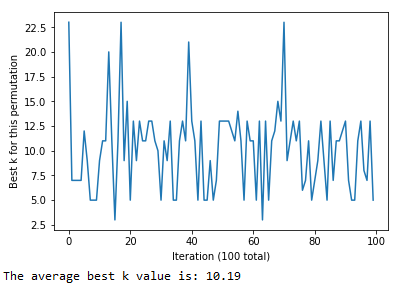
04 June 2017

## Section 1: Optimising the classifiers:

# Nearest Neighbours Classifier

In the Sklearn Library, the K-Nearest Neighbours classifier has a parameter called k, which refers to the number of nearest neighbours to be used for classification. This parameter can influence the score that the classifier will achieve when tested against data that is tested with it. By using cross-validation, we can optimise the k parameter by testing a range of numbers, and determining which has the highest score. This means that the final classifier to be returned by the function can have the most optimised k parameter for the dataset being tested.

By using the test data supplied, “medical\_records.data”, we were able to find the most optimised k value for many permutations of the data. We decided to choose the largest k value that gives the highest score, meaning that if two k values give the same highest score, the larger k value will be chosen. This is to suppress the effects of noise.

The graph on the left shows the highest scores for this permutation are given by k values around the range 5 – 20, with the best value being 11. The graph on the right shows the trend of the best k-values for 100 random permutations of this dataset. The best average k value for our dataset is 10.19, which rounds to 10.

# Support Vector Machine Classifier

In the Sklearn Library, the K-Nearest Neighbours classifier has a parameter called k