

Exercise 9

Classification and Discriminant Analysis

December 17, 2014

Load the data `SAheart` from the package `ElemStatLearn`. The data contain information about males in a heart-disease high-risk region in South Africa (see help). The goal is to apply nearest neighbor methods to split the data into groups according to the variable `chd` (coronary heart disease). Use the binary variable `famhist` in the exercise. Scale the data to have zero mean and unit variance.

1. *k-means with prototypes*: function `kmeans()`
 - (a) Randomly select a training data set of 300 observations and apply k-means clustering for each group with 5 prototypes. Assign the test data with nearest prototypes to predict the group membership (classes of nearest prototypes) for the test data and compute the misclassification rate.
 - (b) Repeat the procedure 100 times and visualize the resulting misclassification rates with a boxplot.
2. *Learning Vector Quantization*: function `lvq` in `library(class)`
 - (a) Randomly select a training data set of 300 observations. Compute a *codebook* by using `lvqinit()` consisting of 10 prototypes. Now apply different algorithms (`lvq1`, `lvq2`, `lvq3`, `olqv1`), predict the group membership for the test data by using `lvqtest` and compute the misclassification rates.
 - (b) Repeat the procedure 100 times and visualize the resulting misclassification rates with a boxplot.
3. *knn classification*: function `knn` in `library(class)`
 - (a) Randomly select a training data set of 300 observations. Use the function `knnEval` from the `library(chemometrics)` to specify the optimal k (the number of nearest neighbors).
 - (b) Apply `knn()` on the training set with k from (a), predict the group membership for the test data and compute the misclassification rate.
 - (b) Repeat the procedure 100 times and visualize the resulting misclassification rates with a boxplot.

Please, send your R scripts with the solution as a text file saved as "Surname9.R", via email to

`kynclova@statistik.tuwien.ac.at`

at latest until December 15.