Homework 3.1 – solution

Example code that loads the data and splits it into training and test portions is available in the file load-col-data.R

1) Simple Decision Trees models

m.dt.11

m.dt.11 <- rpart(Class ~ A1+A2+A3+A4+A5+A6+A7+A8+A9+A10+A11,
data=train, method="class")

test accuracy = 87.8%
m.dt.10 <- rpart(Class ~ A1+A2+A3+A4+A5+A6+A7+A8+A9+A10,  
data=train, method="class")

test accuracy = 85.6%
2) Simple Naïve Bayes models

**m.nb.11**

m.nb.11 <- naiveBayes(as.factor(Class) ~ A1+A2+A3+A4+A5+A6+A7+A8+A9+A10+A11, data=train)

test accuracy = 85.3%

**m.nb.10**

m.nb.10 <- naiveBayes(as.factor(Class) ~ A1+A2+A3+A4+A5+A6+A7+A8+A9+A10, data=train)

test accuracy = 85.6%

**Comparing NB models via cross-validation**

Example code that does the cross-validation is available in the file do-cv.R

***** Results of cross-validation process *****

m.nb.11 -- cross-validation accuracies:
[1] 0.860 0.875 0.847 0.862 0.833 0.866 0.865 0.847 0.859 0.847

m.nb.10 -- cross-validation accuracies:
[1] 0.844 0.874 0.858 0.864 0.828 0.854 0.856 0.835 0.850 0.852

*** Another run:

> round(m.nb.11.acc,3)
[1] 0.853 0.859 0.863 0.871 0.832 0.848 0.863 0.860 0.850 0.849

> round(m.nb.10.acc,3)
[1] 0.851 0.848 0.862 0.871 0.835 0.836 0.860 0.859 0.841 0.843