Master BeNeFri in Computer Science

Course: Statistical Learning Methods
Spring 2016

Exercise #9. Classification: LDA, Logistic regression

Download from the ILIAS website the dataset `Vertebral` dataset (filename: VertebralData.2C.txt). This dataset is composed by various biomedical variables that can be used to predict the orthopedic class of the patient (variable `Status`) which is a binary variable (`Normal` (100 observations), `Abnormal` (210 observations)). You have total 310 observations (patients) with six predictors. There are no missing values. (The correct classification of this dataset is sometimes hard). Other information is given in the file VertebralDescription.pdf.

1. Use logistic regression to predict variable `Status`.

2. Use LDA to predict variable `Status`.

3. Compare the predictions you obtained with the logistic regression and LDA. Use a fair methodology to compare the classifiers (and explain your choice).

   Can you estimate the error rate for those strategies? Which classifier is the best? Why?