

## Statistical Learning, Exercise 8 : Comparing schemes, logistic regression and LDA

### Question 1

Using the *Cars2* data set (Cars2Data.txt) to :

1. Create three different (glm) linear regression models (at least one using a polynomial as predictor ( $mpg \sim poly(predictor, i)$  with  $i \neq 1$ );
2. Using 10-fold cross validation, estimate the test error of the models created in (1);
3. Applying t-test, can you know which model is the best?

### Question 2

Download the *Cancer* data set (Cancer.txt), you can get more informations in the file CancerDescription.txt.

1. Predict the category variable *diagnosis* using logistic regression and interpret the most important values you obtained with R;
2. Estimate the error rate of your model;

### Question 3

Download the *Vertebral* data set (VertebralData.2C.txt) composed of various bio-medical variables used to predict the orthopedic class of the patient (variable *Status*). This variable is a binary and can have the values *Normal* (100 observations) and *Abnormal* (210 observations). There is 310 observations in the data set with six predictors and there is no missing values. You can read the file VertebralDescription.pdf for more information.

1. Predict the variable *Status* using logistic regression;
2. Predict the variable *Status* this time using LDA;
3. Compare the results you obtained in (1) and (2). Choose a methodology and explain your strategy to evaluate the two models;