Business Analytics - ETC3250 2018 - Lab 5

Souhaib Ben Taieb

23 March 2018

Exercice 1

Read and run the code in Sections 4.6.1 to 4.6.6 of ISLR.

Assignment - Question 1

Exercise 7 in chapter 4.7 of ISLR.

Assignment - Question 2

Exercise 8 in chapter 4.7 of ISLR.

Assignment - Question 3

Exercise 10 in chapter 4.7 of ISLR.

Assignment - Question 4

Download the file "data_lab5.Rdata" which contains two datasets D1 and D2, each with n = 800 points, $x \in \mathbb{R}^2$ and $y \in \{0, 1\}$.

- (1) Plot the data D1 with the class labels given by y. Run logistic regression, using the glm function in R. What is the training misclassification rate?

- (2) Draw the decision boundary in $\mathbb{R}_2$ of the logistic regression model from (1), on top of your plot from (1). What shape is it? Does this boundary look like it adequately separates the classes?

- (3) Run logistic regression on the predictors x_1 and x_2, as well as the predictor x_1^2. This is analogous to adding a quadratic term to a linear regression. What is the training misclassification rate? Why is this better than the model from (1)?

- (4) Do (1), (2) and (3) for dataset D2. What additional predictors can you pass to logistic regression to improve classification? (Hint: draw a curve between the classes by eye. What shape does this have?)

TURN IN

- Your `.Rmd` file (which should knit without errors and without assuming any packages have been pre-loaded)
- Your Word (or pdf) file that results from knitting the Rmd.
- DUE: April 1, 11:55pm (late submissions not allowed), loaded into moodle