Business Analytics - ETC3250 2018 - Lab 8

The bootstrap

Souhaib Ben Taieb

19 April 2018

Exercise 1

Do the exercise 2 in Section 5.4 of ISLR.

Bootstrap confidence interval of the correlation coefficient

We will find a 95% confidence interval for the correlation coefficient of Median House value and average number of rooms in the Boston data set from the `MASS` package.

The functions `cor` and `cor.test` will compute the correlation and an asymptotic 95% confidence interval for it. This interval is based on Fisher's z transform

$$z = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$$

which is approximately normally distributed with variance $1/(n-3)$ where n is the number of observations. So if z_L and z_U are upper and lower limits for z, then

$$r_L = \frac{\exp(2z_L)-1}{\exp(2z_L)+1} \qquad \text{and} \qquad r_U = \frac{\exp(2z_U)-1}{\exp(2z_U)+1}$$

are upper and lower limits for r.

We will use the bootstrap to test if this is a good approximation in this case.

Exercise 2

Check that the confidence interval returned by `cor.test` is computed using the above transformation.

Exercise 3

Compute a 95% bootstrap confidence interval for the correlation. You will need to sample rows of the `Boston` matrix.

Exercise 4

Write a function that will return a bootstrap confidence interval for the correlation of any two numeric variables of the same length. Your function should take four arguments:

- `x`: a numeric vector of data
- `y`: a numeric vector of data
- `level`: the probability coverage of the confidence interval with default value of 0.95
- `B`: the number of bootstrap samples with default value of 1000.