Bookmark this page

5.2.R1

0/1 point (graded)

Suppose we want to use cross-validation to estimate the error of the following procedure:

Step 1: Find the k variables most correlated with y

Step 2: Fit a linear regression using those variables as predictors

We will estimate the error for each k from 1 to p, and then choose the best k.

True or false: a correct cross-validation procedure will possibly choose a different set of k variables for every fold.

- TRUE ✔
- FALSE ✖

Explanation

True: we need to replicate our entire procedure for each training/validation split. That means the decision about which k variables are the best must be made on the basis of the training set alone. In general, different training sets will disagree on which are the best k variables.

Submit

Answers are displayed within the problem