Homework5

Name:
Student ID:
Statistical Learning for Data Science

Due time: June 7, 2023 (Wednesday) 16:00pm

1 Proof

Choosing the Optimal Model

- AIC

In the case of the linear model with Gaussian errors, maximum likelihood and least squares are the same thing, and C_p and AIC are equivalent.

Prove this.（HW）

2

We now review k-fold cross-validation.

(a) Explain how k-fold cross-validation is implemented.

(b) What are the advantages and disadvantages of k-fold cross-validation relative to:

i The validation set approach?

ii LOOCV?

3

Suppose that we use some statistical learning method to make a prediction for the response Y for a particular value of the predictor X. Carefully describe how we might estimate the standard deviation of our prediction.

4

We perform best subset, forward stepwise, and backward stepwise selection on a single data set. For each approach, we obtain $p + 1$ models, containing $0, 1, 2, \ldots, p$ predictors. Explain your answers:

(a) Which of the three models with k predictors has the smallest training RSS?

(b) Which of the three models with k predictors has the smallest test RSS?

(c) True or False:

i The predictors in the k-variable model identified by forward stepwise are a subset of the predictors in the $(k+1)$-variable model identified by forward stepwise selection.

ii The predictors in the k-variable model identified by backward stepwise are a subset of the predictors in the $(k+1)$ variable model identified by backward stepwise selection.

iii The predictors in the k-variable model identified by backward stepwise are a subset of the predictors in the $(k+1)$ variable model identified by forward stepwise selection.

iv The predictors in the k-variable model identified by forward stepwise are a subset of the predictors in the $(k+1)$-variable model identified by backward stepwise selection.

v The predictors in the k-variable model identified by best subset are a subset of the predictors in the $(k+1)$-variable model identified by best subset selection.

The electronic version should be sent in PDF format in the form of "homework5-name-student ID" to 12032795@mail.sustech.edu.cn