

Statistical Data Mining I

Concept Quiz

1. In k-NN, if the “true” decision boundary were highly non-linear, then would we expect the best value of K to be *large* or *small*? Why (brief)?

2. TRUE OR FALSE:
In a linear model, adding as many predictors as possible, even extraneous or irrelevant ones, will lower the training error.

3. TRUE OR FALSE:
In a linear model, adding as many predictors as possible, even extraneous or irrelevant ones, will lower the test error.

4. Suppose that we perform best subset (exhaustive), forward stepwise and backward stepwise selection on a single data set. Which of the three approaches of subset selection is guaranteed to have the smallest training RSS (residual sum of squares) in a k -variable model?

5. TRUE OR FALSE:
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.

6. TRUE OR FALSE:
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.