MATH560 HW5

Jon Griffith and Carla Ellefsen

April 14, 2025

1. 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?

   Answer: Best subset selection has the smallest training RSS since it has a much larger search space of every possible model for a total of 2^p models. Forward and backward stepwise selection on the other hand only compute $1 + p(1 + p)/2$ models so they are more computationally efficient, but probably won't produce the model with the smallest training RSS.

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

   Answer: The smallest test RSS for k predictor models will probably be produced by the best subset method but it is not necessarily guaranteed since there is a risk of best subset overfitting the data since it considers every possible model. Forward and backward stepwise k models will probably be comparable to each other in their test RSS values. In theory, any of the three could produce the smallest test RSS but we'd expect best subset method to be the winner more times than not.

   (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.

      Answer: True. Each progressive $k + 1$ step contains the previous k variables of the previous step plus one additional.

(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.

Answer: True. Each $k-$variable model is contained in the previous $k+1$ variable model from the previous step.

(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.

Answer: False. It is not guaranteed that the models will retain the same variables for each $k-$variable model. They often will differ for higher values of p.

(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.

Answer: False. Same as answer above, these models are not guaranteed to retain the same variables for $k-$variable models.

(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.

Answer: False. Best subset selection reevaluates all $\binom{p}{k}$ variables at each step to find the k variables with the lowest training RSS and doesn't guarantee any variables from the preceding step survive the current step.