1. Compared to OLS, lasso imposes additional restriction to the coefficients, and thus is less flexible and will give improved prediction accuracy when its increase in bias is less than its decrease in variance.By setting it , thus less flexible. As lambda increases, spurious coefficients are forced to 0, and at a certain point all coefficients are removed , all coefficients are shrunk to zero. thus test RSS will start to increase.

compared to OLS, Ridge regression imposes additional restriction to the coefficients, and thus is less flexible and will give improved prediction accuracy when its increase in bias is less than its decrease in variance.By setting it , thus less flexible. Increasing lambda will cause spurious coefficients are forced to 0, and at a certain point all coefficients are removed, thus test RSS will start to increase.

2

Best subsets selection is most practical when p is not large, and it can comprehensively cover all the combinations of the regressors. However, it does have drawbacks, since it is not suitable for large p in terms of computational issue as well as overfitting/ high variance due to the huge search space.

3

This is because by definition, adjusted R squared is a function of RSS. Max adjusted R squared is equivalent to Min the RSS. Thus we can expect the best model selected using RSS will be the same as the best model chosen using adjusted R2.

4.

This is false, because since we used K fold CV, the optimal lambda is obtained. This lambda might not be big enough to result in a sparse solution.

5.

This is true, since standardizing the predictors will not change the products of Xj\*Betaj for all j. Thus by definition, R squared as the ratio of explained sum of squares and total sum of squares is a function of the unchanged Xj\*Begaj, thus won't change.

6.

Principal components transform the data to a new coordinate system such that the greatest variance by some scalar projection of the data comes to lie on the respective coordinates, and thus the components are obtained by linear combinations of covariates. However, variable selection approaches select the best model using certain criteria, which will result in fewer variables included. In the final regression, PC uses new variables (principal components) while Variable selection use same subset of the variables.