DS502- HW1

Mahdi Alouane and Rahul Pande

1. 2.4 (question 1)

- (a) The sample size n is extremely large, and the number of predictors p is small.

As the sample size is extremely large the certainty of the veracity of the sample mean is high. Therefore the variance will be low, since an unknown sample will not deviate a lot from the sample mean of a large sample. The low number of predictors will prevent overfitting and thus reduce variance. Since we have two factors that reduce variance, in this case, a flexible statistical learning method is expected to be better because it will reduce the bias. The higher variance of more flexible learning method will be countered by the above two factors.

- (b) The number of predictors p is extremely large, and the number of observations n is small.

The large number of predictors will tend to overfit and perform poor on unseen data, thus increasing variance. Similarly with less number of samples there could be a lot of variablility in the least sqaures which would mean higher variance for unseen data. Therefore in this case an inflexible learning method is expected to perform better since it will counter the high variance.

- (c) The relationship between the predictors and response is highly non-linear.

Where the relationship between predictors and response is highly non-linear, an inflexible model will have high bias given that it won't be able to capture the complex relationship between the predictors and response. Therefore a flexible method is expected to perform better in this case since it will reduce the bias.

- (d) The variance of the error terms, i.e. $\mu^2 = \text{Var}()$, is extremely high.

As the variance of error terms is extremely high, the certainty of the veracity of the sample mean is very low. Therefore the model variance will be high, since an unknown sample could deviate a lot from the sample mean. Hence, in this case, an inflexible statistical learning method is expected to be better because it will reduce the variance.

1. 2.4 (question 3)

1. 2.4 (question 4)