

2.4 Exercise

1.

First, we can know that the test MSE can be divided into three parts, the irreducible error, the squared bias and the variance:

$$MSE = E \left[\left(Y - \hat{f}(x) \right)^2 \mid (x = x_0) \right] = Var[\hat{f}(x_0)] + [Bias(\hat{f}(x_0))]^2 + Var(\varepsilon)$$

Sine the $Var(\varepsilon)$ is the irreducible error, and it is the minimum of the test MSE, so it requires low variance and low squared bias to reduce MSE. From math model, we can know that more flexible methods have less bias and have higher variance, because more flexible the model is, the better it will fit the real function. But a small change in the training set might lead to large changes in the model, so it has a higher variance.

(a).

A flexible model will be better, because more sample can be used to fit the data.

(b).

An inflexible method will be better, because the flexible could not fit the data well.

(c).

A flexible model will be better, because the inflexible model would have a large bias.

(d).

An inflexible method will be better, because the flexible may overfit the real data and have a large bias.

2.

(a).

It is a regression problem because the response and the factor are quantitative. And we are most interested in inference because we are interested in understanding which factors affect CEO salary, not the method to predict salary. (p=3, n=500)

(b).

It is a classification problem because there are tow categories of the response variable. We are more interested in prediction because we want to know whether a new launch will be a success or a failure. (p=13, n=20)

(c).

A regression problem, since the variable we want to predict is a number, and we are more interested in prediction. ($p=3$, $n=52$)