Chapter 2

Q1.

(a) Better. A flexible method will perform better because it can extract more information from large data n. Also risk of overfitting is low since n is large.

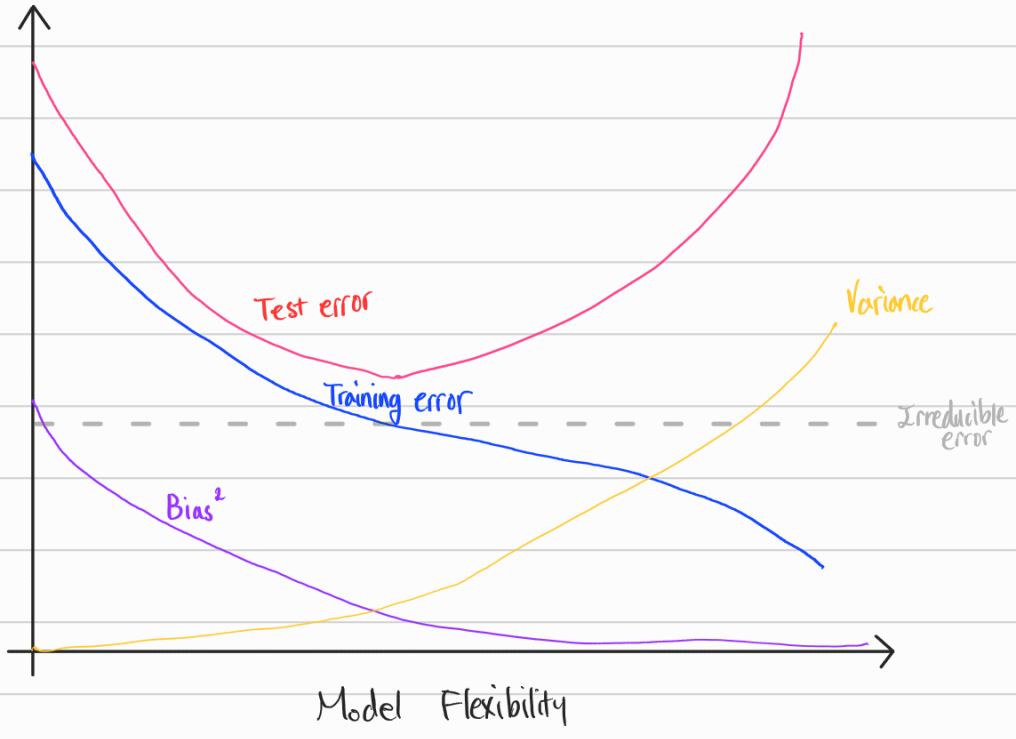
(b) Worse. Overfitting is very likely to happen with flexible methods since n is small.

(c) Better. A flexible method will perform better since inflexible methods are not suitable finding non-linear relationships

(d) Worse. High variance of error terms shows us that the data is noisy. Flexible methods will overfit to the noise

Q3.

(a)



Q3 (b) $bias^2$ - keep decreasing because increase in flexibility will produce a closer fit

Variance - keep increasing because higher flexibility will overfit

training error - continously decreases because higher flexibility will produce closer fit

test error - decreases until it reaches to the optimal point of bias-variance tradeoff then it increases due to overfitting.

Irreducible error - it defines the lower limit. Test error is bounded below by the irreducible error. Training error lower than irreducible error indicates the model is overfitted.

Chapter 3

Q 1. Null hypothesis for TV and radio is that they have no effect on sales while holding other predictors fixed.
Since p-values of both TV and radio are almost 0. Thus their null hypothesis is false. $\Rightarrow$ changing budget will have impact on sales
Newspaper also has similar null hypothesis. However, the p-value for newspaper is high. Thus we do not have enough evidence to reject null hypothesis.

Q 5

Variable names are interchangable.

$$\hat{y}_i = x_i \hat{\beta} = x_i \frac{\sum x_j y_j}{\sum x_k^2} = \sum_{j=1}^{n} \left(\frac{x_i x_j y_j}{\sum_{k=1}^{n} x_k^2} \right)$$

$$\Rightarrow \sum_{j=1}^{n} \left[\left(\frac{x_i x_j}{\sum_{k=1}^{n} x_k^2} \right) y_j \right] \quad \leftarrow \text{in the form of } \sum a_i y_i$$

$$a_j = \frac{x_i x_j}{\sum_{k=1}^{n} x_k^2}$$