Mark Chafin

IST – 5535

Homework 2

Chapter 2 section 2.4 (pages 52-53)

1. Question 1
   1. Flexible would be better. A flexible method is able to take more data from the large “n”. The large “n” reduces the risk of overfitting. Overfitting is the largest problem within flexible methods.
   2. Inflexible method would perform better. The small “n” increases the risk of overfitting which is not good.
   3. A flexible method would be best here because inflexible methods do not perform as well in nonlinear relationships.
   4. Inflexible method will work best since there is a high variance of error. With the high variance of error there will likely be a lot of “noise” and the inflexible method will reduce the risk of overfitting.
2. Question 5
   1. The bias-variance tradeoff states that neither flexible or inflexible is better than the other. Flexible is good when the relationship is non-linear and there are a lot of data points. Inflexible is good for linear relationships and you don’t have many data points.
3. Question 7
   1. Compute the Euclidean distance between each observation and the test point, X1 = X2 = X3 = 0
      1. 3
      2. 2
      3. 3.16
      4. 2.24
      5. 1.41
      6. 1.73
   2. The nearest neighbor to (0,0,0) is 5 at (-1,0,1) with an Euclidean distance of 1.41. 5 is green so I am predicting (K = 1) test point will be green.
   3. The nearest neighbor to (0,0,0) is 5, 6, and 2. Since 5 = Green, 6 = Red, and 2 = Red. I am predicting (K =3) test point will be Red with the majority.
   4. The smaller value of K, the better. Since the Bayed boundary is highly non-linear there is a less advantage to generalize further because of the high variance.