Review Week 2 Answers

1. We are asked to used the nearest neighbor method to calculate the color of point Y which has coordinates, $X_1 = X_2 = X_3 = 0$. First calculate the distance from Y to each of the points in the data using the formula, $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$. For instance in the case of observation 1 that distance is $\sqrt{(0-0)^2 + (3-0)^2 + (0-0)^2} = 3$. Repeating this for all six observations we get,

   (a)

Observation	Distance
1	3
2	2
3	3.16
4	2.24
5	1.41
6	1.73

   (b) The single closest observation is number 5, so our estimate is the colore of observation 5 or Green.

   (c) For K=3 the three closest neighbors are numbers 5, 6, and 2, which are colors green, red and red. Thus, the probability of a green neighbor is 1/3 and the probability of a red neighbor is 2/3. So our prediction is red since that color has the highest probability.

   (d) A large K will have reduced variance but very large bias making it difficult to follow a non-linear boundary. Thus, we would expect in general a smaller K to do better.
2. (a) even though the cubic model is wrong, the RSS of the training data will be at worst equal to the RSS of the linear equation and in all likelihood it will be less. Since the training data is being used to both estimate the regression coefficients and estimate the RSS it will be biased downwards.

   (b) If we use the test data for our estimation of the RSS we would expect that the RSS of the linear equation would be less than the cubic equation. By using test data we eliminate the bias mentioned above and thus the model with superfluous parameters is expected to do more poorly prediction new observations.

   (c) The answer would be the same as (a). The more complicated model will always be able to do at least as well as the linear model which is a subset of the cubic equation when it is tested with training data.

   (d) Now we really can tell which RSS will be better since it depends on the whether the departure from linearity is small and thus a linear approximation might be best or very large in which case the cubic equation might be best.
3. Using the equation $y = \hat{\beta}_0 + \hat{\beta}_1 x$, then let $x = \bar{x}$ and solve for y. Thus, $y = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} = \bar{y}$ since from eq. (3.4) we know that $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.