Chapter 2

Conceptual

Q1. For each of parts (a) through (d), indicate whether I. or ii. is correct and explain your answer. In general, do we expect the performance of a flexible statistical learning method to perform better or worse than an inflexible method when:

a) The sample size n is extremely large, and the number of predictors pp is small?

   A flexible method will fit the data better and perform better with the large sample size than an inflexible method.

b) The number of predictors pp is extremely large, and the number of observations n is small?

   The small number of observations would be overfit by a flexible method.

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

   A flexible method would fit better with more degrees of freedom than an inflexible one.

d) The variance of the error terms, i.e., $\sigma2=Var(\varepsilon)\sigma2=Var(\varepsilon)$, is extremely high?

   The noise in the error terms would be fit by a flexible method, which would increase variance.

Q2. Explain whether each scenario is a classification or regression problem and indicate whether we are most interested in inference or prediction. Finally, provide n and pp.

a) We collect a set of data on the top 500 firms in the US. For each firm we record profit, number of employees, industry, and the CEO salary. We are interested in understanding which factors affect CEO salary.

   Regression and inference with n=500n=500 *and* p=3

b) We are considering launching a new product and wish to know whether it will be a success or a failure. We collect data on 20 similar products that were previously launched. For each product we have recorded whether it was a success or failure, price charged for the product, marketing budget, competition price, and ten other variables.

   Classification and prediction with n=20n=20 *and* p=13

c) We are interested in predicting the % change in the US dollar in relation to the weekly changes in the world stock markets. Hence, we collect weekly data for all of 2012. For each week we record the % change in the dollar, the % change in the US market, the % change in the British market, and the % change in the German market.

   Regression and prediction with n=52n=52 *and* p=3

Q7. The table below provides a training data set containing 6 observations, 3 predictors, and 1 qualitative response variable. Suppose we wish to use this data set to make a prediction for Y when X1 = X2 = X3 = 0 using K-nearest neighbors.

Obs.	X_1	X_2	X_3	Y
1	0	3	0	Red
2	2	0	0	Red
3	0	1	3	Red
4	0	1	2	Green
5	−1	0	1	Green
6	1	1	1	Red

a) Compute the Euclidean distance between each observation and the test point, X1 = X2 = X3 = 0.

The Euclidean distance between points p and q is the length of the line segment connecting them pq¯.

1. sqrt (3^2) = 3
2. sqrt (2^2) = 2
3. sqrt (1^2 + 3^2) = sqrt (10) = 3.16 4, sqrt (1^2 + 2^2) = sqrt (5) = 2.24
4. sqrt (1^2 + 1^2) = sqrt (2) = 1.41
5. sqrt (1^2 + 1^2 + 1^2) = sqrt (3) = 1.73

b) What is our prediction with K=1? Why?

if $K=1$ Then $x_5 \in \mathcal{N}_0$ and we have

$$P(Y = \text{Red} \mid X = x_0) = \frac{1}{1} \sum_{i \in \mathcal{N}_0} I(y_i = \text{Red}) = I(y_5 = \text{Red}) = 0$$

and

$$P(Y = \text{Green} \mid X = x_0) = \frac{1}{1} \sum_{i \in \mathcal{N}_0} I(y_i = \text{Green}) = I(y_5 = \text{Green}) = 1$$

It is the class from the nearest neighbor - the observation 5 - Green.

c) What is our prediction with K=3? Why?

If $K=3$ then $x_2, x_5, x_6 \in \mathcal{N}_0$ and we have

$$P(Y = \text{Red} \mid X = x_0) = \frac{1}{3} \sum_{i \in \mathcal{N}_0} I(y_i = \text{Red}) = \frac{1}{3}(1+0+1) = \frac{2}{3}$$

and

$$P(Y = \text{Green} \mid X = x_0) = \frac{1}{3} \sum_{i \in \mathcal{N}_0} I(y_i = \text{Green}) = \frac{1}{3}(0+1+0) = \frac{1}{3}$$

so our prediction is Red

d) If the Bayes decision boundary in this problem is highly nonlinear, then would we expect the best value for KK to be large or small? Why?

The boundary becomes inflexible as K grows larger (linear). As a result, we should expect the best value for K to be small in this case.

Chapter 3

Conceptual

Q1. Describe the null hypotheses to which the p-values given in Table 3.4 correspond. Explain what conclusions you can draw based on these p-values. Your explanation should be phrased in terms of sales, TV, radio, and newspaper, rather than in terms of the coefficients of the linear model.

	Coefficient	Std. error	t-statistic	p-value
Intercept	2.939	0.3119	9.42	< 0.0001
TV	0.046	0.0014	32.81	< 0.0001
radio	0.189	0.0086	21.89	< 0.0001
newspaper	−0.001	0.0059	−0.18	0.8599

TABLE 3.4. *For the* Advertising *data, least squares coefficient estimates of the multiple linear regression of number of units sold on radio, TV, and newspaper advertising budgets.*

- The null hypothesis states that no amount of television, radio, or newspaper advertising will affect sales. Sales = TV0 + Radio0 + Newspaper*0 + Intercept, or Y = Intercept, is the formula.
- The fact that the Intercept has a p-value of 0.0001 indicates that TODO
- When we hold Radio and Newspaper advertising constant, the second p-value, p 0.0001, shows that we can reject the null hypothesis and conclude that TV has an impact on sales.
- When we hold TV an
- When we hold TV and newspaper advertising constant, the third p-value, the radio has p < 0.0001, shows that we can reject the null hypothesis and conclude that radio has an impact on sales.
- The fourth p-value, p < 0.0001, indicates that newspaper advertising has no effect on sales when TV and radio advertising are held constant.

Q3. Suppose we have a data set with five predictors, X[1] = GPA, X[2] = IQ, X[3] = Gender (1 for Female and 0 for Male), X[4] = Interaction between GPA and IQ, and X[5] = Interaction between GPA and Gender. The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model, and get $\hat{\beta}[0] = 50$, $\hat{\beta}[1] = 20$, $\hat{\beta}[2] = 0.07$, $\hat{\beta}[3] = 35$, $\hat{\beta}[4] = 0.01$, $\hat{\beta}[5] = -10$.

a. Which answer is correct, and why?

i. For a fixed value of IQ and GPA, males earn more on average than females.
ii. For a fixed value of IQ and GPA, females earn more on average than males.
iii. For a fixed value of IQ and GPA, males earn more on average than females provided that the GPA is high enough.
iv. For a fixed value of IQ and GPA, females earn more on average than males provided that the GPA is high enough.

The least square line is given by

$$\hat{y} = 50 + 20GPA + 0.07IQ + 35Gender + 0.01GPA \times IQ - 10GPA \times Gender$$

which becomes for the males

$$\hat{y} = 50 + 20GPA + 0.07IQ + 0.01GPA \times IQ$$

and for the males

$$\hat{y} = 85 + 10GPA + 0.07IQ + 0.01GPA \times IQ$$

so, the starting salary for males is higher than for females on average iff $50 + 20GPA \geq 85 + 10GPA$ which is equivalent to $GPA \geq 3.5$.

Therefore iii. is the right answer.

b) Salary of a female with IQ of 110 and GPA of 4.0

by estimating that women earn on an average of .

$$50 + 20GPA + 0.07IQ + 35 + 0.01(GPA \times IQ) - 10GPA$$

$$\Rightarrow 50 + 20(4) + 0.07(110) + 35 + 0.01(4)(110) - 10(4)$$

$$= 137.1$$

Since the unit was in 1000s of dollars, we can predict a post-grad salary of \$137,000.

(c) False, because the statistical significance of an interaction is different from the magnitude of the interaction. It's possible to have a lot of evidence for a small effect. Also, a small coefficient doesn't even mean the interaction effect is small, since it is very sensitive to the units of two variables