Cha. 4

1.

Proof:

From (4.2)

We can get

So, we have proofed that equation 4.3 (odds) can be deduced from equation 4.2 (logistic function).

2.

(a) If the Bayes decision boundary is linear, we expect QDA to perform better on training set because of its higher flexibility; while LDA perform better on test set than QLA does, because QDA could overfit the linearity on the Bayes decision boundary.

(b) If the Bayes decision boundary is non-linear, we expect QDA better on both.

(c) General speaking, LDA tends to be a better bet than QDA if there are relatively few training observations. In contrast, when sample size n (training set) increases, QDA is recommended, so that the variance of the classifier is not a major concern, or if the assumption of a common covariance matrix for the K classes is clearly.

(d) False. It depends on different situation, sometimes it is an advantage to apply a method with more flexibility such as QDA which may lead a superior test error rate, but it can turn to a disadvantage when too much flexible may lead to overfit, turning to an inferior test error rate.

3.

In this case, we should choose logistic regression because of its lower test error rate. When KNN with K=1, we can get a 0% training error rate, which means we do not make any error on the training data within this data set, but we get an average error rate (averaged over both test and training data sets) of 18% which implies that the tests error rate is 36%; on the opposite, by using logistic regression, we can get a training error rate of 20% and the a test error rate of 30%, which is smaller than the test error rate of KNN when K=1. So, it is more appropriate to choose logistic regression if we expect a lower test error rate.

Cha. 5

6.

Proof:

Take the first derivative of Var related to , we can get:

To check whether the result is minimum, we can take the second derivative of , checking whether it greater than or equal to zero:

7.

(a) 1-1/n.

Because bootstrap sampling draws items with replacement, we are sampling from the same pool with the same probability for each time, so there are (n-1) items in the n that are not j and with a probability of 1-1/n that the first item is not j.

(b) 1-1/n

Still because the bootstrap sampling pattern is with replacement, so the probability would not change in this case.

(c) production nth observation is not in the bootstrap where n!=j.

In previous situation, we can get the probability that the jth observation is not in the bootstrap sample is (1-1/n) for each time sampling, when the bootstrap where n!=j, the same situation and probability for j have to repeated for n times, so the n times (1-1/n) multiplied, then the result is .

(d)

when N=5,

(e)

When N=100,

(f)

When N=10000,

(g)

Coding: x=1:100000

Plot(x,1-(1/x)^x)