

Statistical learning assignment 7- chapter 4

孫浩哲 M072040002

November 1, 2018

4.4 (a)

10%, because we choose 10% of the range of X .

(b)

$$10\% \times 10\% = 1\%$$

(c)

$$(0.1)^{100} \times 100\% = 10^{-98}\%$$

(d)

As p increases, the training observations "near" may decrease in power.

(e)

$$p = 1, l = 10\%^1 = 0.1$$

$$p = 2, l = \sqrt{10\%} \approx 0.3162$$

$$p = 100, l = \sqrt[100]{10\%} \approx 1$$

4.5 (a)

We expect QDA is better on training set because QDA is more flexible, and we expect LDA is better on test set because we want to avoid overfitting.

(b)

Because it is non-linear, so we prefer QDA on both sets.

(c)

Improve, because as the sample size increases, the model may be more complicated, so we need a more flexible model to make a better fit.

(d)

False, the QDA may be overfitting if the sample size is small.

4.6 (a)

$$P(X) = \frac{e^{(-6+0.05 \times 40+3.5)}}{1 + e^{(-6+0.05 \times 40+3.5)}} \approx 37.75\%$$

(b)

$$-6 + 3.5 + 0.05 \times \text{hours} = 0 \rightarrow \text{hours} = 50$$

$$(\because \frac{e^0}{1 + e^0} = 0.5)$$