ISLR Exercises.

~~3.7.4~~ 3.7.4

I collect a set of data ($n=100$) ~~wa~~ containing a single predictor and a quantitative response. I then fit a linear regression model to the data, as well as a separate cubic regression, i.e $Y=\beta_0+\beta_1 X+\beta_2 X^2+\beta_3 X^3+\epsilon$

a) Suppose that true relationship between X and Y is linear, i.e $Y=\beta_0+\beta_1 X+\epsilon$. Consider the training RSS for linear reg and training RSS for cubic reg. Would be expect one to be lower than the other, would be expect them to be same, or is there not enough info to tell? Justify your answer

Ans. I expect training RSS for the cubic regression to be better as it would provide a better fit to the train data

b) Answer a) using test rather than train RSS

Ans. Test RSS for linear regression would be better, as cubic regression would fit a lot of noise, thereby worsening it's ability to generalize well. Hence, ~~train~~ linear regression would have a lower RSS

c) Suppose the true relationship between X and Y is not linear, but we don't know how far it is from linear. Consider training RSS for linear reg and also training RSS for cubic regression. Which one would have a lower RSS?

Ans Again cubic regression would have a lower RSS owing to it's flexibility

d) Answer (c) using test rather than training RSS

Ans In this case, there is not information to tell since we do not know the underlying ground truth. If the true relationship is linear, linear regression would have a lower RSS, if it is closer to cubic then cubic regression would be better

4.7.7

Suppose that we wish to predict whether a stock company will issue a dividend this year ("Yes or "No") based on X, last year's profit. We examine $\bar{X} = 10$ for companies that issued dividend, and $\bar{X} = 0$ for those who didn't. In addition, the variance of X for these 2 companies was $\hat{\sigma}^2 = 36$. Finally 80% of companies issued dividends. Assume X follows a normal distribution, predict the probability that a company will issue a dividend given its % profit was $X=4$ last year.

~~PA|X=4~~

Ans $$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

X = last year's percent profit

Y = A binary variable (Yes or No) for indicating whether stock companies would issue dividend or not

$$P(Y=1) = 0.8 \quad , \quad P(Y=0) = 0.2 \qquad \sigma^2 = 36$$

We have to predict $P(Y=1 | X=4)$

$$P(Y=1|X=4) = \frac{P_r(X=4|Y=1) \cdot P_r(Y=1)}{P_r(X=4)}$$

$$= \frac{f_k(x)\ \pi_k(x)}{\sum_{l=1}^{k} \pi_l f_l(x)}$$

$$P_r(X=4|Y=1) = f_1(x)$$

$$= \frac{1}{\sqrt{2\pi \times 36}} e^{-[(4-10)^2/2\times 36]} = \frac{e^{-0.5}}{\sqrt{72\pi}}$$

$$= \cancel{0.127}\ 0.0403$$

$$f_2(x) = \frac{1}{\sqrt{2\pi \times 36}} e^{-[(4)^2/2\times 36]}$$

$$= \frac{1}{\sqrt{2\pi \times 36}} e^{-[16/72]} = \frac{e^{-2/9}}{\sqrt{72\pi}} = 0.053$$

$$P(Y=1|X=4) = \frac{0.8 \times 0.0403}{0.8 \times 0.0403 + 0.053 \times 0.2}$$

$$P(Y=1|X=4) = 0.752$$

ISLR 4.7.3

Q. This problem relates to QDA model, in which obs within each class are drawn from normal distribution with a class specific mean and covariance matrix. We consider simple case where $p=1$.

Suppose we have K classes, and if obs belongs to K^{th} class then x comes from 1D Normal dist, $X \sim N(\mu_k, \sigma_k^2)$. Prove that in this case, the Bayes' classifier is not linear. Argue that it is in fact quadratic.

Ans. We know that for $k \in (1, \ldots K)$

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2}\left(\frac{(x-\mu_k)^2}{\sigma_k^2}\right)}$$

The posterior probability is given by.

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma_k} e\left(-\frac{1}{2\sigma_k^2}(x-\mu_k)^2\right)}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma_l} e\left(-\frac{1}{2\sigma_l^2}(x-\mu_l)^2\right)}$$

The naive Baye's classifier will assign a ~~f~~ feature to a class k, whichever k makes $p_k(x)$ the largest.

$$\underset{k}{\operatorname{argmax}} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2}\left(\frac{(x-\mu_k)^2}{\sigma_k^2}\right)}.$$

Taking log

$$= -\log\sqrt{2\pi} - \log\sigma_k - \frac{1}{2} \cancel{\log}\left(\frac{(x-\mu_k)^2}{\sigma_k^2}\right)$$

Ignore $+ \log\pi_k$

On rearranging we get

$$\delta_k(x) = \log\pi_k - \log\sigma_k \cancel{+} - \frac{1}{2} \cancel{\log}\left(\frac{(x-\mu_k)^2}{\sigma_k^2}\right)$$

$$\delta_k(x) = -\frac{1}{2\sigma_k^2}(x-\mu_k)^2 + \log\pi_k - \log\sigma_k$$

$\therefore$ The Naive Baye's classifier is a quadratic function of x.