Stat Learning PS4

Jay Lee

September 30, 2017

4.4

(a) On average, we will use 10% (1/10) of the available observations to predict our test observation's response, because we use the surrounding range of 10% in one variable.

(b) In this case we will use 1% (1/100) of the available observations, because we use the surrounding 10% of two variables.

(c) Informally "inducting" from the first two cases, we will use $1/10^{100}$ of the available observations to make the prediction.

(d) When p is very large, the probability of finding a "neighbor" in our range grows ever smaller. In the case of part (c), we would need to have 10^{100} observations to expect even one point to be within our range for comparison.

(e) Because we exponentiated for the earlier parts, we can take roots for this part. When $p = 1$, the length of each side is $\sqrt[1]{.1} = .1$; when $p = 2$ the length is $\sqrt[2]{.1} \approx .31622$; when $p = 100$ the length is $\sqrt[100]{.1} \approx .97724$.

4.6

(a)

$$\begin{aligned} P(Y) &= \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2}} \\ &= \frac{e^{-6+.05*40+1*3.5}}{1+e^{-6+.05*40+1*3.5}} \\ &\approx .3775 \end{aligned}$$

(b)

$$\begin{aligned} &\log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \\ \implies &\log(.5/.5) = -6 + .05 * x_1 + 1 * x_2 \\ \implies &\quad 0 = -6 + .05 * x_1 + 1 * 3.5 \\ \implies &x_1 = 50 \text{ hours.} \end{aligned}$$

4.7

Let $Y = 1$ represent the event a company issues a dividend in a given year, and $Y = 0$ the event that company does not issue a dividend that year. Then, by Bayes' Theorem,

$$P(Y = 1|X = 4) = \frac{f_1(X = 4)\pi_1}{f_1(X = 4)\pi_1 + f_0(X = 4)\pi_0},$$

where $\pi_i = P(Y = i)$, and $f_i(X = 4) = P(X = 4|Y = i)$. Because both f_i are distributed normally, $f_i(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu_i)^2/2\sigma^2}$. Substituting $\mu_0 = 0$, $\mu_1 = 10$, $\sigma^2 = 36$, $\pi_0 = .2$, and $\pi_1 = .8$,

$$
\begin{aligned}
P(Y = 1|X = 4) &= \frac{f_1(X = 4)\pi_1}{f_1(X = 4)\pi_1 + f_0(X = 4)\pi_0} \\
&= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu_1)^2/2\sigma^2}\pi_1}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu_1)^2/2\sigma^2}\pi_1 + \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu_0)^2/2\sigma^2}\pi_0} \\
&= \frac{\frac{1}{\sqrt{72\pi}} e^{-(4-10)^2/72} * .8}{\frac{1}{\sqrt{72\pi}} e^{-(4-10)^2/72} * .8 + \frac{1}{\sqrt{72\pi}} e^{-(4-0)^2/72} * .2} \\
&= \frac{e^{-(4-10)^2/72} * 4}{e^{-(4-10)^2/72} * 4 + e^{-(4-0)^2/72}} \\
&\approx 0.7519.
\end{aligned}
$$

4.2

Let K be the set of all classes an observation can be classified into. Consider x, our example observation. Let k be the class that maximizes the Bayes classifier for x (so the Bayes classifier assigns observation x to class k). Then, $\forall j \in K$ s.t. $j \neq k$ (I drop the denominator of the Bayes classifier here to save some LaTeX time, because it's just the sum over K on both sides),

$$
\begin{aligned}
p_k(x) &> p_j(x) \\
\implies \pi_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right) &> \pi_j \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_j)^2\right) \\
\implies \log\left(\pi_k \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)\right) &> \log\left(\pi_j \exp\left(-\frac{1}{2\sigma^2}(x-\mu_j)^2\right)\right) \\
\implies \log(\pi_k) - \frac{1}{2\sigma^2}(x-\mu_k)^2 &> \log(\pi_j) - \frac{1}{2\sigma^2}(x-\mu_j)^2 \\
\implies \log(\pi_k) - \frac{x}{2\sigma^2} + x\frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} &> \log(\pi_j) - \frac{x}{2\sigma^2} + x\frac{\mu_j}{\sigma^2} - \frac{\mu_j^2}{2\sigma^2} \\
\implies \log(\pi_k) + x\frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} &> \log(\pi_j) + x\frac{\mu_j}{\sigma^2} - \frac{\mu_j^2}{2\sigma^2}.
\end{aligned}
$$

So, the discriminant function evaluated for k is greater than when it is evaluated for any other class j, and classifying x into k maximizes 4.13, the discriminant function.