20835281

BI Qiuyu

4.1

$$P(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} \quad (4.2)$$

$$1-P(X) = \frac{1}{1+e^{\beta_0+\beta_1 X}}$$

$$\frac{P(X)}{1-P(X)} = e^{\beta_0+\beta_1 X} \quad (4.3)$$

So, I proved that (4.2) is equivalent to (4.3)

4.2

$$P_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)} \quad (4.17)$$

The Bayes classifier assigns observation $X=x$ to class k if $P_k(x)$ is largest.

$$\log(P_k(x)) = \log \pi_k + \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2\sigma^2}(x-\mu_k)^2 - \log\left[\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)\right]$$

$$= \log \pi_k - \log(\sqrt{2\pi}\sigma) - \frac{x^2}{2\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \frac{\mu_k \cdot x}{\sigma^2} - \log\left[\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)\right]$$

$$= \left[x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log \pi_k\right] + \left[-\log(\sqrt{2\pi}\sigma) - \frac{x^2}{2\sigma^2} - \log\left(\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)\right)\right]$$

For a particular observation x,

$\left[-\log(\sqrt{2\pi}\sigma) - \frac{x^2}{2\sigma^2} - \log\left(\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)\right)\right]$ does not change, whichever class k the Bayes classifier assigns x to.

So $\arg\max_k P_k(x)$

is equivalent to $\arg\max_k \log(P_k(x)) = \arg\max_k \left(x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log \pi_k + C\right)$

is equivalent to $\arg\max_k \left(x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log \pi_k\right)$

$$x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log \pi_k \quad (4.18)$$

So, I proved classifying an observation to the class for which (4.17) is largest is equivalent to classifying an observation to the class for which (4.18) is largest.

43

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

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

Bayes classifier assigns observation $X=x$ to. $\arg\max_k P_k(x)$ class.

$$\log(P_k(x)) = \log \pi_k - \log(\sqrt{2\pi}\sigma_k) - \frac{1}{2\sigma_k^2}(x^2+\mu_k^2-2x\mu_k) + C$$

$$= -\frac{1}{2\sigma_k^2} x^2 + \frac{\mu_k}{\sigma_k^2} x + \left[\log \pi_k - \log(\sqrt{2\pi}\sigma_k) - \frac{\mu_k^2}{2\sigma_k^2}\right] + C$$

Bayes classifer assigns observation $X=x$ to

$$\arg\max_k \left(-\frac{1}{2\sigma_k^2} x^2 + \frac{\mu_k}{\sigma_k^2} x + \left[\log \pi_k - \log(\sqrt{2\pi}\sigma_k) - \frac{\mu_k^2}{2\sigma_k^2}\right]\right) \text{ class.}$$

We can see that, for different class k, $-\frac{1}{2\sigma_k^2} \underline{\underline{x^2}} + \frac{\mu_k}{\sigma_k^2} \underline{\underline{x}} + \left[\log \pi_k - \log(\sqrt{2\pi}\sigma_k) - \frac{\mu_k^2}{2\sigma_k^2}\right]$ is quadratic function of x,

So in QDA model. Bayes classifier is in fact quadratic.

5.1

$$Var(\alpha X + (1-\alpha)Y) = \alpha^2 Var(X) + (1-\alpha)^2 Var(Y) + 2\alpha(1-\alpha) Cov(X, Y)$$
$$= \alpha^2 \sigma_X^2 + (1+\alpha^2-2\alpha)\sigma_Y^2 + 2\alpha(1-\alpha)\sigma_{XY}$$
$$= \sigma_X^2 \alpha^2 + \sigma_Y^2\alpha^2 - 2\sigma_Y^2\alpha + \sigma_Y^2 - 2\sigma_{XY}\alpha^2 + 2\sigma_{XY}\alpha$$
$$= (\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})\alpha^2 + (2\sigma_{XY} - 2\sigma_Y^2)\alpha + \sigma_Y^2$$

It's a quadratic function of α, and $\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY} = Var(X-Y) > 0$

so.

$Var(\alpha X + (1-\alpha)Y)$ gets minimum when $\alpha = -\dfrac{2\sigma_{XY} - 2\sigma_Y^2}{2(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})} = \dfrac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}$

5.2

(a). the probability is $(1-\frac{1}{n})$.

According to the bootstrap procedure, bootstrap observation is selected randomly from original samples. So the probability that first bootstrap observation is jth observation from original sample is $\frac{1}{n}$, that's to say, the probability that first bootstrap observation is jth observation from original sample is $1-\frac{1}{n}$.

(b). the probability is also $(1-\frac{1}{n})$

because bootstrap is sampling with replacement.

(c). For ith bootstrap observation, the probability that it's not jth observation from original sample is $(1-\frac{1}{n})$.

If jth observation is not in the bootstrap sample, then for each ith bootstrap observation it's not jth observation from original sample. Bootstrap sample has n observations.

The probability is $(1-\frac{1}{n})^n$

(d). $n=5$.

The probability is $1-(1-\frac{1}{5})^5=0.67232$

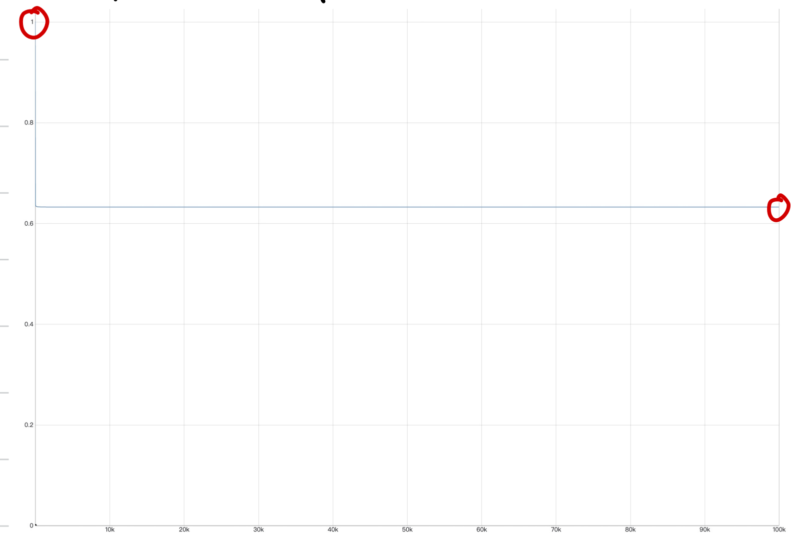
(e). $n=100$

The probability is $1-(1-\frac{1}{100})^{100}\approx 0.634$

(f). $n = 10{,}000$

The probability is $1-(1-\frac{1}{10000})^{10000} \approx 0.632$

(g). plot the function $1-(1-\frac{1}{n})^{n}$:



We can see it's a monotonously decreasing function from 1 to $1-\frac{1}{e}$

(h). the result mean(store) is very close to $1-\frac{1}{e}$.