2. It was stated in the text that 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. Prove that this is the case. In other words, under the assumption that the observations in the kth class are drawn from a $N(\mu_k, \sigma^2)$ distribution, the Bayes classifier assigns an observation to the class for which the discriminant function is maximized.

第k個先驗機率

第k個後驗機率

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

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

1. 先找到k使最大化 $\pi_k \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)$

2. 取ln: $\ln\left(\pi_k \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)\right)$

$= \ln(\pi_k) - \frac{1}{2\sigma^2}(x-\mu_k)^2 - \ln\sqrt{2\pi\sigma^2}$

3. 忽略 $\ln\sqrt{2\pi\sigma^2}$, 並展開

$\Rightarrow \ln(\pi_k) - \frac{1}{2\sigma^2}(x^2 - 2\mu_k + \mu_k^2)$

$= \ln(\pi_k) - \cancel{\frac{x^2}{2\sigma^2}} + \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} = \delta_k(x)$

常數項不計

$\therefore$ 最大化 $p_k(x)$ = 最大化 $\delta_k(x)$