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(μk,σ2) distribution, the Bayes classifier assigns an observation to the class for which the discriminant function is maximized.

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

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

π_k：類別 k 的先驗机率

$p_k(x)$：類別 k 的後驗机率

By Bayes classifier, 將 x 分類到使 $p_k(x)$ 最大的類別 k

等價於分類到使 $\log(p_k(x))$ 最大的類別 k

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

Ignore the second term since it is independent of k.

There are some constants we can also ignore in the first term.

Expanding $\log(\pi_k) - \frac{(x-\mu_k)^2}{2\sigma^2}$

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

$$= x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k) - \cancel{\frac{x^2}{2\sigma^2}} \quad \text{ignore}$$

$$= \delta_k(x) \quad \#$$