An Introduction to Statistical Learning: Chapter 4

Due on January 5,2016 at 3:10pm

G. James et.al.

Tony Jiang

1. Using a little bit of algebra, prove that (4.2) is equivalent to (4.3). In other words, the logistic function representation and logit representation for the logistic regression model are equivalent.

Answer:

$$\begin{aligned} p(X) &= \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} \quad (4.2) \\ \frac{p(X)}{1-p(X)} &= \frac{\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}{1-\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}} \\ &= \frac{\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}{\frac{1+e^{\beta_0+\beta_1 X}-e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}} \\ &= e^{\beta_0+\beta_1 X} \quad (4.3) \end{aligned}$$

2. It was stated in the text that classifying an observation to the class for which (4.12) is largest is equivalent to classifying an observation to the class for which (4.13) 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.

Answer:

Recall 4.12 has the form:

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

and 4.13

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

Taking the log of 4.12 on both sides:

$$\begin{aligned} log\big(p_k(x)\big) &= log\Big(\frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} exp\big(-\frac{1}{2\sigma^2}(x-\mu_k)^2\big)}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma} exp\big(-\frac{1}{2\sigma^2}(x-\mu_l)^2\big)}\Big) \\ &= log\Big(\pi_k \frac{1}{\sqrt{2\pi}\sigma} exp\big(-\frac{1}{2\sigma^2}(x-\mu_k)^2\big)\Big) - log\Big(\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma} exp\big(-\frac{1}{2\sigma^2}(x-\mu_l)^2\big)\Big) \end{aligned}$$

To maximize this , we only need to maximize all items related to class k. that is to say, we only care about what is a function of k (the rest are the same to a given x).

$$
\begin{aligned}
\arg\max_x p_k(x) = \arg\max_x log(p_k(x)) \\
&= \arg\max_x \quad log\Big(\pi_k \frac{1}{\sqrt{2\pi}\sigma} exp\big(-\frac{1}{2\sigma^2}(x-\mu_k)^2\big)\Big) \\
&= \arg\max_x \quad \Big(log(\pi_k) + log(\frac{1}{\sqrt{2\pi}\sigma}) - [\frac{1}{2\sigma^2}(x^2 - 2\mu_k x + \mu_k^2)]\Big) \\
&= \arg\max_x \quad \Big(log(\pi_k) - \frac{x^2}{2\sigma^2} + \frac{2\mu_k x}{2\sigma^2} - \frac{\mu_k^2}{2\sigma^2}\Big) \\
&= \arg\max_x \quad \Big(log(\pi_k) - \frac{x^2}{2\sigma^2} + \frac{\mu_k x}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2}\Big) \\
&= \arg\max_x \quad \Big(x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + log(\pi_k)\Big)
\end{aligned}
$$

The last line bears the same form as 4.13.

3. This problem relates to the QDA model,in which the observations within each class are drawn from a normal distribution with a class-specific mean vector and a class specific covariance matrix. We consider the simple case where $p = 1$; i.e. there is only one feature.

Suppose that we have K classes, and that if an observation belongs to the kth class then X comes form a one-dimensional normal distribution, $X\ N(\mu_k, \sigma_k^2)$. Recall that the density function for the one-dimensional normal distribution is given in (4.11). Prove that in this case, the Bayes' classifier is *not* linear. Argue that it is in fact quadratic.

Hint: For this problem, you should follow the arguments laid out in Section 4.4.2, but without making the assumption that $\sigma_1^2 = ... = \sigma_K^2$.

Answer:

$$
\begin{aligned}
\arg\max_x p_k(x) &= \arg\max_x log(p_k(x)) \\
&= \arg\max_x \quad log\Big(\pi_k \frac{1}{\sqrt{2\pi}\sigma_k} exp\big(-\frac{1}{2\sigma_k^2}(x-\mu_k)^2\big)\Big) \\
&= \arg\max_x \quad \Big(log(\pi_k) + log(\frac{1}{\sqrt{2\pi}\sigma_k}) - [\frac{1}{2\sigma_k^2}(x^2 - 2\mu_k x + \mu_k^2)]\Big) \\
&= \arg\max_x \quad \Big(log(\pi_k) - \frac{x^2}{2\sigma_k^2} + \frac{2\mu_k x}{2\sigma_k^2} - \frac{\mu_k^2}{2\sigma_k^2}\Big) \\
&= \arg\max_x \quad \Big(log(\pi_k) - \frac{x^2}{2\sigma_k^2} + \frac{\mu_k x}{\sigma_k^2} - \frac{\mu_k^2}{2\sigma_k^2}\Big) \\
&= \arg\max_x \quad \Big(-x^2 \cdot \frac{1}{2\sigma_k^2} + x \cdot \frac{\mu_k}{\sigma_k^2} - \frac{\mu_k^2}{2\sigma_k^2} + log(\pi_k)\Big)
\end{aligned}
$$

It does contain the x^2 term.