1 (b). This observation suggests that when we perform logistic regression to learn the function $a(x)$, we actually indirectly learn the class distributions $p(x|C_1)$ and $p(x|C_2)$. It is interesting to consider how different regression models for $a(x)$ relate to different assumptions about class distributions. Assume that both classes are draws from the same exponential family, with different parameters.

$$p(x|C_k) = h(x) \exp(\phi(x)^T \omega_k - \log Z(\omega_k))$$

Show that this implies a <u>linear</u> model for $a(x)$

$$a(x) = \phi(x)^T \vec{\theta} + \theta_0$$

What are the parameters $\vec{\theta}$ and θ_0 of this model in terms of the parameters ω of the class distributions?

Answer:

① Let $p(x|C_1) = h(x) \exp(\phi(x)^T \omega_1 - \log Z(\omega_1))$
and $p(x|C_2) = h(x) \exp(\phi(x)^T \omega_2 - \log Z(\omega_2))$

then we have $\dfrac{p(C_1|x)}{p(C_2|x)} = \dfrac{p(x|C_1)\, p(C_1) / p(x)}{p(x|C_2)\, p(C_2) / p(x)} = \dfrac{p(x|C_1)}{p(x|C_2)} \dfrac{p(C_1)}{p(C_2)}$

$$\Rightarrow \frac{\cancel{h(x)} \exp(\phi^T(x)\, \omega_1 - \log Z(\omega_1))}{\cancel{h(x)} \exp(\phi^T(x)\, \omega_2 - \log Z(\omega_2)} \quad \frac{p(C_1)}{p(C_2)}$$

$$\Rightarrow \exp\left[\phi^T(x)(\omega_1 - \omega_2) - \log Z(\omega_1) + \log Z(\omega_2)\right] \quad \frac{p(C_1)}{p(C_2)}$$

$$\Rightarrow \exp\left[\phi^T(x)(\omega_1 - \omega_2) - \ln \frac{Z(\omega_1)}{Z(\omega_2)} + \ln \frac{p(C_1)}{p(C_2)}\right]$$

We have $a(x) = \ln \dfrac{p(C_1|x)}{p(C_2|x)}$

Thus

$$a(x) = \ln \left[\exp \left[\phi^T(x)(\omega_1 - \omega_2) - \ln \frac{Z(\omega_1)}{Z(\omega_2)} + \ln \frac{P(C_1)}{P(C_2)} \right] \right]$$

$$a(x) = \phi^T(x)(\omega_1 - \omega_2) + \left[\ln \frac{P(C_1)}{P(C_2)} - \ln \frac{Z(\omega_1)}{Z(\omega_2)} \right]$$

thus $\vec{\theta} = \omega_1 - \omega_2$

$$\theta_0 = \ln \frac{P(C_1)}{P(C_2)} - \ln \frac{Z(\omega_1)}{Z(\omega_2)}$$

and $a(x) = \phi^T(x)\,\vec{\theta} + \theta_0$ is a linear model