Chapter 11

Neural Networks

Problem 11.3

With logit function as the estimated probability distribution and cross-entropy as the cost function, we can write the problem as the minimization of

$$R(\theta) = \sum_{i=1}^{N} R_i(\theta) = \sum_{i=1}^{N} \sum_{k=1}^{K} y_{ik} \log \frac{1}{f_k(x_i)}.$$

where

$$f_k(x_i) = \frac{\exp\left(g_k(\beta_k^T z_i)\right)}{\sum_{l=1}^{K} \exp\left(g_l(\beta_l^T z_i)\right)} \text{ and}$$

$$R_i(\theta) = \sum_{k=1}^{K} y_{ik} \log\left(\sum_{l=1}^{K} \exp\left(g_l(\beta_l^T z_i) - g_k(\beta_k^T z_i)\right)\right).$$

The partial derivative of $R_i(\theta)$ with respect to β_{qm} can be expressed as

$$\begin{aligned}
\frac{\partial R_i(\theta)}{\partial \beta_{qm}} &= \sum_{k=1}^{K} y_{ik} \frac{\frac{\partial}{\partial \beta_{qm}} \left[\sum_{l=1}^{K} \exp\left(g_l(\beta_l^T z_i) - g_k(\beta_k^T z_i)\right)\right]}{\sum_{l=1}^{K} \exp\left(g_l(\beta_l^T z_i) - g_k(\beta_k^T z_i)\right)} \\
&= \frac{\left(\sum_{k=1,k\neq q}^{K} y_{ik} - y_{iq}\right) \exp\left(g_q'(\beta_q^T z_i)\right) z_{mi}}{\sum_{l=1}^{K} \exp\left(g_l(\beta_l^T z_i) - g_k(\beta_k^T z_i)\right)} \\
&= \delta_{qi} z_{mi}
\end{aligned}$$

where

$$\delta_{qi} = \frac{\partial R_i(\theta)}{\partial(\beta_{qm} z_{mi})}.$$

Then we can express the partial derivative of $R_i(\theta)$ with respect to α_{mp} can be expressed as

$$\begin{aligned}
\frac{\partial R_i(\theta)}{\partial \alpha_{mp}} &= \sum_{k=1}^{K} \frac{\partial R_i(\theta)}{\partial(\beta_{km} z_{mi})} \frac{\partial(\beta_{km} z_{mi})}{\partial z_{mi}} \frac{\partial z_{mi}}{\partial \alpha_{mp}} \\
&= \left(\sum_{k=1}^{K} \delta_{qi} \beta_{km}\right) \sigma'(\alpha_m^T x) x_{pi} = \frac{\partial R_i(\theta)}{\partial(\alpha_{mp} x_{pi})} x_{pi},
\end{aligned}$$

which is the *back-propagation equation*.

Problem 11.4

Multi-nomail logistic model itself is defined with normalized exponential of linear function as probability and the cross-entropy as cost, which is exactly the neural network for classification without hidden layer.