Chapter 5 - Ex 2

证明：假设 x_1, x_2 都属于 w_i 类

则 $\max\limits_j g_j(x_1) = g_i(x_1) = w_i^T x_1 + w_{i0}$ ①

$\max\limits_j g_j(x_2) = g_i(x_2) = w_i^T x_2 + w_{i0}$ ②

而对任意的 j，$0 \le \lambda \le 1$，

$$w_j^T[\lambda x_1 + (1-\lambda)x_2] + w_{j0}$$

$$= \lambda(w_j^T x_1 + w_{j0}) + (1-\lambda)(w_j^T x_2 + w_{j0})$$

$$\le \lambda(w_i^T x_1 + w_{i0}) + (1-\lambda)(w_i^T x_2 + w_{i0}) \quad ③$$

又对于 x_1 与 x_2 之间的一点 $\lambda x_1 + (1-\lambda)x_2$

$$\max\{w_j^T[\lambda x_1 + (1-\lambda)x_2] + w_{j0}\}$$

$$= \alpha \max(w_j^T x_1 + w_{j0}) + (1-\lambda)\cdot\max(w_j^T x_2 + w_{j0})$$

$$\overset{由①,②}{=} \lambda(w_i^T x_1 + w_{i0}) + (1-\lambda)(w_i^T x_2 + w_{i0})$$

$$= w_i^T[\lambda x_1 + (1-\lambda)x_2] + w_{i0} \quad ④$$

结合 ③，④，则点 $\lambda x_1 + (1-\lambda)x_2$ 也属于 w_i 类.

得证

Chapter 6 . Ex 21

解: a)

$$\text{softmax}(x) = \frac{e^{x_j}}{\sum_j e^{x_j}} \xlongequal{\text{记}} z(x)$$

则 $\frac{\partial z}{\partial x} = z \cdot (1-z)$

易知三层神经网络向前计算为

$$a_j = \sum_{i=1}^{d} w_{ji} x_i$$

$$y_j = f(a_j)$$

$$o_k = \sum_{j=1}^{c} w_{kj} y_j$$

$$z_k = \frac{e^{o_k}}{\sum_{m=1}^{c} e^{o_m}}$$

目标函数为 $J = \frac{1}{2} \sum_{k=1}^{c} (t_k - z_k)^2$

现在要求 $\frac{\partial J}{\partial w_{kj}}$ 与 $\frac{\partial J}{\partial w_{ji}}$

$$\frac{\partial J}{\partial w_{kj}} = \frac{\partial J}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_{kj}} = \left(\sum_{s=1}^{c} \frac{\partial J}{\partial z_s} \cdot \frac{\partial z_s}{\partial o_k} \right) \cdot \frac{\partial o_k}{\partial w_{kj}}$$

因 $$\frac{\partial z_s}{\partial O_k} = \begin{cases} \dfrac{e^{O_s}\cdot(-1)\cdot e^{O_k}}{\left(\sum_{m=1}^{c} e^{O_m}\right)^2} = -z_s z_k, & 当\ s \neq k \\[2ex] \dfrac{e^{O_s}}{\left(\sum_{m=1}^{c} e^{O_m}\right)} + (-1)\dfrac{e^{O_s}\cdot e^{O_s}}{\left(\sum_{m=1}^{c} e^{O_m}\right)^2} = z_k - z_k^2, & 当\ s = k \end{cases}$$

$$\therefore \frac{\partial J}{\partial z_s} = (-1)(t_s - z_s)$$

$$\therefore \frac{\partial J}{\partial O_k} = \sum_{s\neq k}^{c} (-1)(t_s - z_s)(-z_s z_k) + (-1)(t_k - z_k)(z_k - z_k^2)$$

又 $$\frac{\partial O_k}{\partial w_{kj}} = y_j$$

$$\boxed{\therefore \frac{\partial J}{\partial w_{kj}} = y_j \sum_{s\neq k}^{c} (t_s - z_s)(z_s z_k) - y_j (t_k - z_k)(z_k - z_k^2)}$$

再 $$\frac{\partial J}{\partial w_{ji}} = \frac{\partial J}{\partial y_j} \cdot \frac{\partial y_j}{\partial a_j} \cdot \frac{\partial a_j}{\partial w_{ji}}$$

其中 $$\frac{\partial J}{\partial y_j} = \sum_{s=1}^{c} \frac{\partial J}{\partial z_s} \frac{\partial z_s}{\partial y_j}$$

$$= -\sum_{s=1}^{c} (t_s - z_s) \frac{\partial z_s}{\partial y_j}$$

$$= -\sum_{s=1}^{c} (t_s - z_s) \left[\frac{\partial z_s}{\partial O_s} \frac{\partial O_s}{\partial y_j} + \sum_{r\neq s}^{c} \frac{\partial z_s}{\partial O_r} \cdot \frac{\partial O_r}{\partial y_j} \right]$$

$$= -\sum_{s=1}^{c}(t_s - z_s)(z_s - z_s^2)w_{sj}$$

$$+ \sum_{s=1}^{c}\sum_{r\neq s}^{c} z_s z_r w_{rj}(t_s - z_s)$$

又易知 $\dfrac{\partial a_j}{\partial w_{ji}} = x_i$

$$\boxed{\therefore \frac{\partial J}{\partial w_{ji}} = -x_i \cdot f'(a_j)\sum_{s=1}^{c}(t_s - z_s)w_{sj}(z_s - z_s^2)}$$

综上学习规则为：

$$\Delta w_{ji} = -\eta\frac{\partial J}{\partial w_{ji}}$$

$$\Delta w_{kj} = -\eta\frac{\partial J}{\partial w_{kj}}$$

其中 η 为学习率