

$$P_1: y_k = \frac{e^{z_k}}{\sum_k e^{z_k}}$$

$$\begin{aligned} J &= -\sum_k t_k \log y_k \\ &= -\sum_k t_k (\log e^{z_k} - \log \sum_k e^{z_k}) \\ &= -\sum_k t_k [z_k - \log \sum_k e^{z_k}] \\ &= -\sum_k t_k z_k - \log \sum_k e^{z_k} \end{aligned}$$

$$\begin{aligned} \frac{\partial J}{\partial z_k} &= \frac{\partial}{\partial z_k} (t_k z_k - \log \sum_k e^{z_k}) \\ &= -t_k + \frac{1}{\sum_k e^{z_k}} e^{z_k} \\ &= -t_k + y_k \end{aligned}$$

$$\therefore \frac{\partial J}{\partial w_{kj}} = \frac{\partial J}{\partial z_k} \frac{\partial z_k}{\partial w_{kj}} = (y_k - t_k) x_j$$

$$\frac{\partial J}{\partial b_k} = \frac{\partial J}{\partial z_k} \cdot \frac{\partial z_k}{\partial b_k} = y_k - t_k$$

$$\begin{aligned} \text{total loss: } \frac{\partial J}{\partial w_{kj}} &= \sum_m (y_k^m - t_k^m) x_j^m \\ \frac{\partial J}{\partial b_k} &= \sum_m (y_k^m - t_k^m) \end{aligned}$$

$$\begin{aligned} P_2: \frac{\partial J}{\partial w} &= \|y - t\| \\ \frac{\partial J}{\partial b} &= \|y - t\| \end{aligned}$$

$$\begin{aligned} P_3: E[CW] &= \sum_k y_k u_k \mathbb{1}\{\hat{t} = k\} \\ \text{Max utility: } &\rightarrow \hat{t} = \underset{k}{\operatorname{argmax}} y_k u_k \end{aligned}$$

$$\begin{aligned} P_4: \log L(\theta) &= \log \prod_{m=1}^M P(x^m, t^m) \\ &= \log \prod_m P(x^m | t^m) P(t^m) \\ &= \sum_m \log P(x^m | t^m) + \sum_m \log P(t^m) \\ &= \sum_m \log P(x_i^m | t^m = k; \theta_{k,i}) + \sum_m \log P(t^m; \pi_1, \dots, \pi_K) \end{aligned}$$

$$\begin{aligned} \hat{\pi}_k &= \frac{\sum_{m=1}^M \mathbb{1}\{t^m = k\}}{M} & \hat{\theta}_{k,i} &= \frac{\sum_m \mathbb{1}\{t^m = k; x_i = 1\}}{\mathbb{1}\{t^m = k\}} \end{aligned}$$