Problem 1

$$J = -\sum_{m=1}^{M} \sum_{k=1}^{K} t_k^{(m)} \log y_k^{(m)}$$

$$= -\sum_{m=1}^{M} \sum_{k=1}^{K} t_k^{(m)} \log \frac{\exp(z_k)}{\sum_{k'} \exp(z_{k'})}$$

$$= -\sum_{m=1}^{M} \sum_{k=1}^{K} t_k^{(m)} \left(\log \exp(z_k) - \log \sum_{k'} \exp(z_{k'}) \right)$$

$$= -\sum_{m=1}^{M} \sum_{k=1}^{K} t_k^{(m)} \left(z_k - \log \sum_{k'} \exp(z_{k'}) \right)$$

$$= -\sum_{m=1}^{M} \sum_{k=1}^{K} t_k^{(m)} z_k + \sum_{m=1}^{M} \log \sum_{k'=1}^{K'} \exp(z_{k'})$$

$$\frac{\partial J}{\partial z_k} = -\sum_{m=1}^{M} t_k^{(m)} + \sum_{m=1}^{M} \frac{1}{\sum_{k'=1}^{K'} \exp(z_{k'})} \times \exp(z_{k'})$$

$$= \sum_{m=1}^{M} \left(-t_k^{(m)} + y_k^{(m)} \right)$$

$$\frac{\partial J}{\partial w_{k,i}} = \frac{\partial J}{\partial z_k} \cdot \frac{\partial z_k}{\partial w_{k,i}}$$

$$= \sum_{m=1}^{M} \left(-t_k^{(m)} + y_k^{(m)} \right) \cdot x_k^{(m)}$$

$$\frac{\partial J}{\partial b_k} = \frac{\partial J}{\partial z_k} \cdot \frac{\partial z_k}{\partial b_k} = \sum_{m=1}^{M} \left(-t_k^{(m)} + y_k^{(m)} \right)$$

Problem 2.

$$y = \sigma(w^T x + b)$$

$$= \frac{1}{1 + \exp(-(w_1^T x + b_1))}$$

$$= \frac{\exp(w_1^T x + b_1)}{\exp(w_1^T x + b_1) + 1}$$

$$= \frac{\exp(w_1^T x + b_1)}{\exp(w_1^T x + b_1) + \exp(w_0^T x + b_0)} \longrightarrow \text{where } w_0^T = 0,\ b_0 = 0$$

Thus, the logistic regression can also be reduced to 2-way softmax.

Problem 3.

$$\mathbb{E}_{t\sim y}[u] = y_1 u_1 \mathbb{1}\{\hat{t}=1\} + y_2 u_2 \mathbb{1}\{\hat{t}=2\} + \cdots + y_k u_k \mathbb{1}\{\hat{t}=k\}$$

$$= \sum_{k=1}^{k} y_k u_k \mathbb{1}\{\hat{t}=k\}$$

Choosing $\hat{t} = \arg\max_t y_t u_t$,

lead to the max total expected utility