$$\nabla_{w_k} L(W) \tag{1}$$

$$= -\frac{d}{d\boldsymbol{w}_k} \sum_{i=1}^{N} \left[\left(\sum_{k=1}^{K} y_{ik} \boldsymbol{w}_k^T \boldsymbol{z}_i) - \log(\sum_{k'=1}^{K} \exp(\boldsymbol{w}_{k'}^T \boldsymbol{z}_i) \right) \right] \tag{2}$$

$$= -\frac{d}{d\boldsymbol{w}_k} \left[\left(\sum_{k=1}^{K} y_k \boldsymbol{w}_k^T \boldsymbol{z} \right) - \log \left(\sum_{k'=1}^{K} \exp(\boldsymbol{w}_{k'}^T \boldsymbol{z} \right) \right] \tag{3}$$

$$= -\frac{d}{d\boldsymbol{w}_k} \left(\sum_{k=1}^{K} y_k \boldsymbol{w}_k^T \boldsymbol{z} \right) + \frac{d}{d\boldsymbol{w}_k} \log \left(\sum_{k'=1}^{K} \exp(\boldsymbol{w}_{k'}^T \boldsymbol{z}) \right) \tag{4}$$

$$= -y_k \boldsymbol{z} + \hat{y}_k \boldsymbol{z} \tag{5}$$

$$= \left(\hat{y}_k - y_k \right) \boldsymbol{z} \tag{6}$$

See below for further derivations of (4).

From first equation of (4):

$$\frac{d}{d\boldsymbol{w}_k}\left(\sum_{k=1}^{K} y_k \boldsymbol{w}_k^T \boldsymbol{z}\right)$$

$$= \underbrace{\frac{d}{d\boldsymbol{w}_k} y_1 \boldsymbol{w}_1^T \boldsymbol{z} + \frac{d}{d\boldsymbol{w}_k} y_2 \boldsymbol{w}_2^T \boldsymbol{z} + \ldots}_{=0} + \frac{d}{d\boldsymbol{w}_k} y_k \boldsymbol{w}_k^T \boldsymbol{z}$$

$$= y_k \boldsymbol{z}$$

From second equation of (4):

$$\frac{d}{d\boldsymbol{w}_k} \log\left(\sum_{k'=1}^{K} \exp(\boldsymbol{w}_{k'}^T \boldsymbol{z})\right)$$

$$= \frac{1}{\sum_{k'=1}^{K} \exp(\boldsymbol{w}_{k'}^T \boldsymbol{z})}\left[\frac{d}{d\boldsymbol{w}_k} \sum_{k'=1}^{K} \exp(\boldsymbol{w}_{k'}^T \boldsymbol{z})\right]$$

$$= \frac{1}{\sum_{k'=1}^{K} \exp(\boldsymbol{w}_{k'}^T \boldsymbol{z})}\left[\underbrace{\frac{d}{d\boldsymbol{w}_k} \exp(\boldsymbol{w}_1^T \boldsymbol{z}) + \frac{d}{d\boldsymbol{w}_k} \exp(\boldsymbol{w}_2^T \boldsymbol{z}) + \ldots}_{=0} + \frac{d}{d\boldsymbol{w}_k} \exp(\boldsymbol{w}_k^T \boldsymbol{z})\right]$$

$$= \frac{1}{\sum_{k'=1}^{K} \exp(\boldsymbol{w}_{k'}^T \boldsymbol{z})}\left[\exp(\boldsymbol{w}_k^T \boldsymbol{z}) \frac{d}{d\boldsymbol{w}_k} \boldsymbol{w}_k^T \boldsymbol{z}\right]$$

$$= \frac{\exp(\boldsymbol{w}_k^T \boldsymbol{z})}{\sum_{k'=1}^{K} \exp(\boldsymbol{w}_{k'}^T \boldsymbol{z})} \boldsymbol{z}$$

$$= \hat{y}_k \boldsymbol{z}$$