MULTICLASS LOGISTIC REGRESSION

$$\nabla_{W_k} L(W) = \frac{d}{dW_k} - \sum_{i=1}^{N} \left[\left(\sum_{j=1}^{K} y_{ik} W_k^T z_i \right) - \log \left(\sum_{k'=1}^{K} \exp(W_{k'}^T z_i) \right) \right]$$

$$= \frac{d}{dW_k} - \left[\left(\sum_{k=1}^{K} y_k W_k^T z \right) - \log \left(\sum_{k'=1}^{K} \exp(W_{k'}^T z) \right) \right]$$

$$= \underbrace{-\frac{d}{dW_k} \left(\sum_{k=1}^{K} y_k W_k^T z \right)} + \underbrace{\frac{d}{dW_k} \log \left(\sum_{k'=1}^{K} \exp(W_{k'}^T z) \right)} = -y_k z + \hat{y}_k z$$

$$= (\hat{y}_k - y_k) z \quad \#$$

$$\frac{d}{dW_k} \left(\sum_{k=1}^{K} y_k W_k^T z \right)$$

$$= \underbrace{\frac{d}{dW_k} y_1 W_1^T z}_{0} + \underbrace{\frac{d}{dW} y_2 W_2^T z}_{0} + \underbrace{\ldots}_{0} + \underbrace{\frac{d}{dW_k} y_k W_k^T z}_{y_k z}$$

$$= y_k z$$

$$\frac{d}{dW_k} \log \left(\sum_{k'=1}^{K} \exp(W_{k'}^T z) \right)$$

$$= \frac{1}{\sum_{k'=1}^{K} \exp(W_{k'}^T z)} \frac{d}{dW_k} \sum_{k'=1}^{K} \exp(W_{k'}^T z)$$

$$= \frac{1}{\sum_{k'=1}^{K} \exp(W_{k'}^T z)} \left(\underbrace{\frac{d}{dW_k} \exp(W_1^T z)}_{0} + \underbrace{\frac{d}{dW_k} \exp(W_2^T z)}_{0} + \underbrace{\ldots}_{0} + \frac{d}{dW_k} \exp(W_k^T z) \right)$$

$$= \frac{1}{\sum_{k'=1}^{K} \exp(W_{k'}^T z)} \exp(W_k^T z) \frac{d}{dW_k} \underbrace{\overset{W_k^T z}{\cancel{\exp(W_k^T z)}}}_{z}$$

$$= \underbrace{\boxed{\frac{\exp(W_k^T z)}{\sum_{k'=1}^{K} \exp(W_{k'}^T z)}}}_{= \hat{y}_k} z$$

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