

## MULTICLASS LOGISTIC REGRESSION

$$\begin{aligned}\nabla_{W_k} L(W) &= \frac{d}{dW_k} - \sum_{i=1}^N \left[ \left( \sum_{k=1}^K y_{ik} W_k^T z_i \right) - \log \left( \sum_{k=1}^K \exp(W_k^T z_i) \right) \right] \\ &= \frac{d}{dW_k} - \left[ \left( \sum_{k=1}^K y_k W_k^T z \right) - \log \left( \sum_{k=1}^K \exp(W_k^T z) \right) \right] \\ &= - \underbrace{\frac{d}{dW_k} \left( \sum_{k=1}^K y_k W_k^T z \right)}_{\text{Term 1}} + \underbrace{\frac{d}{dW_k} \log \left( \sum_{k=1}^K \exp(W_k^T z) \right)}_{\text{Term 2}} = -y_k z + \hat{y}_k z \\ &= (\hat{y}_k - y_k) z \quad \# \end{aligned}$$

$$\begin{aligned}\frac{d}{dW_k} \left( \sum_{k=1}^K y_k W_k^T z \right) &= \frac{d}{dW_k} y_1 W_1^T z + \frac{d}{dW_k} y_2 W_2^T z + \dots + \frac{d}{dW_k} y_k W_k^T z \\ &\quad \underbrace{\phantom{\frac{d}{dW_k} y_1 W_1^T z}}_0 \quad \underbrace{\phantom{\frac{d}{dW_k} y_2 W_2^T z}}_0 \quad \underbrace{\phantom{\frac{d}{dW_k} y_k W_k^T z}}_0 \quad \underbrace{\phantom{\frac{d}{dW_k} y_k W_k^T z}}_{y_k z} \\ &= y_k z\end{aligned}$$

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

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

$$= \frac{1}{\sum_{k=1}^K \exp(W_k^T z)} \left( \underbrace{\frac{d}{dW_k} \exp(W_1^T z)}_0 + \underbrace{\frac{d}{dW_k} \exp(W_2^T z)}_0 + \dots + \underbrace{\frac{d}{dW_k} \exp(W_k^T z)}_0 \right)$$

$$\begin{aligned}&= \frac{1}{\sum_{k=1}^K \exp(W_k^T z)} \exp(W_k^T z) \frac{d}{dW_k} \underbrace{W_k^T z}_z \\ &= \frac{\exp(W_k^T z)}{\sum_{k=1}^K \exp(W_k^T z)} z \\ &\quad \underbrace{\phantom{\frac{\exp(W_k^T z)}{\sum_{k=1}^K \exp(W_k^T z)}}}_{=\hat{y}_k}\end{aligned}$$

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