

# Gradient Derivations

$$\arg \max_{\theta} \sum_{m=1}^M \sum_{i \in S} \left[ \log(\sigma(x_i \vec{w}^\top \vec{h}_i(\vec{y}))) + \sum_{j \in \mathcal{N}_i} x_i x_j \vec{v}^\top \vec{\mu}_{ij}(\vec{y}) - \log(z_i) \right] - \frac{1}{2\tau^2} \vec{v}^\top \vec{v}.$$

$$\frac{\partial}{\partial w_k} = \sum_{m=1}^M \sum_{i \in S} \left[ x_i (\vec{h}_i(\vec{y}))_k (1 - \sigma(x_i \vec{w}^\top \vec{h}_i(\vec{y}))) - \frac{\frac{\partial}{\partial w_k} z_i}{z_i} \right]$$

$$\frac{\partial}{\partial w_k} z_i = \sum_{x_i \in \{-1, 1\}} \exp \left( \log(\sigma(x_i \vec{w}^\top \vec{h}_i(\vec{y}))) + \sum_{j \in \mathcal{N}_i} x_i x_j \vec{v}^\top \vec{\mu}_{ij}(\vec{y}) \right) (1 - \sigma(x_i \vec{w}^\top \vec{h}_i(\vec{y}))) (x_i (\vec{h}_i(\vec{y}))_k)$$

$$\frac{\partial}{\partial v_k} = \sum_{m=1}^M \sum_{i \in S} \left[ \sum_{j \in \mathcal{N}_i} x_i x_j (\vec{\mu}_{ij}(\vec{y}))_k - \frac{\frac{\partial}{\partial v_k} z_i}{z_i} \right] - \frac{v_k}{\tau^2}$$

$$\frac{\partial}{\partial v_k} z_i = \sum_{x_i \in \{-1, 1\}} \exp \left( \log(\sigma(x_i \vec{w}^\top \vec{h}_i(\vec{y}))) + \sum_{j \in \mathcal{N}_i} x_i x_j \vec{v}^\top \vec{\mu}_{ij}(\vec{y}) \right) \sum_{j \in \mathcal{N}_i} x_i x_j (\vec{\mu}_{ij}(\vec{y}))_k$$