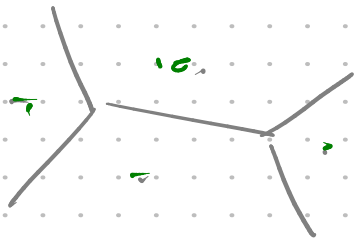
$$p(C_k|\bar{x}) = \frac{p(\bar{x}|C_k)\,p(C_k)}{\sum_s p(\bar{x}|C_s)\,p(C_s)} = \frac{e^{\bar{w}_k^T \bar{x}}}{e^{\bar{w}_s^T \bar{x}}}$$

$$\text{softmax}(a_1, \dots, a_K) = \left(\dots, \frac{e^{a_k}}{\sum e^{a_s}}, \dots \right)$$

$\sigma(a)$

t_{nk}

$W \quad \bar{w}_1, \bar{w}_2, \dots, \bar{w}_K \quad = (0 \dots 1 \dots 0)$



$$p(D|W) = \prod_n p(\bar{t}_n | W, \bar{x}_n) =$$

$$= \prod_n \prod_k \underbrace{p(t_{nk}=1 | W, \bar{x}_n)}_{= y_{nk}}$$

t_{nk}

$$\log p(D|W) = \sum_n \sum_k t_{nk} \log y_{nk}$$

$$y_{nk} = \frac{e^{\bar{w}_k^T \bar{x}_n}}{\sum_s e^{\bar{w}_s^T \bar{x}_n}}$$

$$\nabla_{\bar{w}_l} y_{nk} = \frac{\partial \left(\frac{e^{a_k}}{\sum e^{a_s}} \right)}{\partial a_l} \cdot \bar{x}_n$$

$$\frac{\partial y_{nk}}{\partial a_k} = \frac{e^{a_k}(\sum e^{a_s}) - e^{a_k} \cdot e^{a_k}}{(\sum e^{a_s})^2} = \underbrace{\frac{e^{a_k}}{\sum e^{a_s}}}_{y_{nk}} \cdot \underbrace{\frac{\sum e^{a_s} - e^{a_k}}{\sum e^{a_s}}}_{1 - y_{nk}}$$

$$\frac{\partial y_{nk}}{\partial a_l} = \frac{-e^{a_k} \cdot e^{a_l}}{(\sum e^{a_s})^2} = -y_{nk} \cdot y_{nl}$$

$l \neq k$

$t_{nk} = 1 \Rightarrow t_{nl}$

$$\nabla_{\bar{w}_l} \ln p(D|W) = \sum_n \sum_k t_{nk} \cdot y_{nk} \cdot (\underbrace{[l=k]}_{y_{nk}} - y_{nl}) \bar{x}_n =$$

$$= \sum_n \left(\underbrace{[t = \text{"prob. of } \bar{x}_n\text{"}]}_{t_{nl}} - y_{nl} \right) \bar{x}_n = \sum_n (t_{nl} - y_{nl}) \bar{x}_n$$