$$J(\Theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \log p_k^{(i)}$$

$$\Rightarrow \nabla_{\theta^{(\tilde{k})}} J(\Theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \frac{1}{p_k^{(i)}} \underbrace{\nabla_{\theta^{(\tilde{k})}} p_k^{(i)}} =$$

$$= \nabla_{\theta^{(\tilde{k})}} \frac{e^{\theta_k \cdot x^{(i)}}}{\sum_{j=1}^{K} e^{\theta_j \cdot x^{(i)}}} = x^{(i)} \delta_{k,\tilde{k}} p_k - \underbrace{\frac{e^{\theta_k \cdot x}}{(\Sigma \ldots)^2} x^{(i)} e^{\theta_{\tilde{k}} \cdot x^{(i)}}}_{= p_k^{(i)} x^{(i)} p_{\tilde{k}}^{(i)}}$$

$$= p_k^{(i)} x^{(i)} (\delta_{k,\tilde{k}} - p_{\tilde{k}}^{(i)})$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} x^{(i)} (\delta_{k,\tilde{k}} - p_{\tilde{k}}^{(i)}) =$$

$$-\frac{1}{m}\sum_{i=1}^{m} x^{(i)} \Big(\underbrace{\sum_{k=1}^{K} y_k^{(i)} p_{\tilde{k}}^{(i)}}_{= p_{\tilde{k}}^{(i)}} - \underbrace{\sum_{k=1}^{K} y_k^{(i)} \delta_{k,\tilde{k}}}_{= y_{\tilde{k}}^{(i)}}\Big) = \frac{1}{m}\sum_{i=1}^{m} x^{(i)} (p_{\tilde{k}}^{(i)} - y_{\tilde{k}}^{(i)})$$

Equation 4-23

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} \log p_i + (1-y^{(i)}) \log(1-p_i)\right)$$

$$\Rightarrow \partial_{\theta_j} J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left(\frac{y^{(i)}}{\sigma(\theta^T x^{(i)})} - \frac{(1-y^{(i)})}{1-\sigma(\theta^T x^{(i)})}\right) \partial_{\theta_j} p_i$$

$$\left\{ \partial_{\theta_j} p_i = \partial_{\theta_j} \frac{1}{1+e^{-\theta^T x}} = \frac{-1}{(1+e^{-\theta^T x})^2} (-x_j) e^{-\theta^T x} = x_j^{(i)} p_i^2 e^{-\theta^T x^{(i)}} \right\}$$

$$= -\frac{1}{m}\sum_{i=1}^{m} \frac{y^{(i)}(1-p_i) - (1-y^{(i)}) p_i}{p_i(1-p_i)} x_j^{(i)} p_i^2 e^{-\theta^T x^{(i)}} =$$

$$= -\frac{1}{m}\sum_{i=1}^{m} \frac{y^{(i)} - p_i}{p_i(1-p_i)} x_j^{(i)} p_i (1-p_i) =$$

$$p_i = \frac{1}{1+e^{-\theta^T x}} \Rightarrow 1+e^{-\theta^T x} = \frac{1}{p_i} \Rightarrow e^{-\theta^T x} = \frac{1-p_i}{p_i}$$

$$= -\frac{1}{m}\sum_{i=1}^{m} (y^{(i)} - p_i) x_j^{(i)} = \frac{1}{m}\sum_{i=1}^{m} (\underset{\sigma(\theta^T x^{(i)})}{p_i} - y^{(i)}) x_j^{(i)}$$

Wir bedanken uns für Ihren Aufenthalt und auf Wiedersehen

Equation 4-18

Quadratic Programming (hard margin: Equation 5-5, page 159; soft margin: Exercise 7, page 165 & page 480)

hard margin

bias term

$n'_p = n + 1$
↳ features

$n'_c = m$ (instances)

"bias term"

$$H' = \underbrace{\begin{pmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}}_{n'_p} \Big\} n'_p$$

$\underline{f}' = \underline{0}$ (n'_p - dimensional vector)

$\underline{b}' = 1$ (n'_c - dimensional vector)

$\underline{a}'^{(i)} = -t^{(i)} \underline{x}^{(i)}$ ✓
↳ $\overbrace{(1, x^{(i)})}^{n'_p}$

bias feature, n features

i-th row of A' ($n'_c \times n'_p$)

$$\boxed{\underline{p}' = \underbrace{(\underset{\text{bias}}{b}, \underset{\text{weights}}{\underline{w}})}_{n'_p}}$$

soft margin

$$H = \underbrace{\begin{pmatrix} H' & 0 & \cdots \\ 0 & 0 & \\ \vdots & & \ddots \end{pmatrix}}_{n_p = n'_p + m} \Big\} n_p$$

$$\underline{f} = \begin{pmatrix} \underline{f}' \\ C \\ \vdots \end{pmatrix} \begin{matrix} \} n'_p \\ \} m \end{matrix} \Big\} n_p \quad , \quad \underline{b} = \begin{pmatrix} \underline{b}' \\ 0 \\ \vdots \end{pmatrix} \begin{matrix} \} n'_c = m \\ \} m \end{matrix} \Big\} n_c = 2m$$

$$A = \begin{pmatrix} \underbrace{A'}_{n'_p} & \underbrace{I_m}_{n'_c = m} \\ 0 & -I_m \end{pmatrix} \begin{matrix} \} n'_c = m \\ \} n'_c = m \end{matrix} \Big\} n_c = 2m$$

$n_p = n'_p + m$

$$p = \begin{pmatrix} p' \\ \zeta^{(1)} \\ \zeta^{(2)} \\ \vdots \\ \zeta^{(m)} \end{pmatrix} \begin{matrix} \} n'_p \\ \\ \} m \\ \\ \end{matrix} \Big\} n_p$$

$$\Rightarrow \frac{1}{2} p^T H p + f^T p = \underbrace{\frac{1}{2} p'^T H' p'}_{= \frac{1}{2} \underline{w}^T \underline{w}} + C \sum_{i=1}^{m} \zeta^{(i)} = \frac{1}{2} \underline{w}^T \underline{w} + C \sum_{i=1}^{m} \zeta^{(i)}$$ ✓

$$(Ap)^{(i)} = -t^{(i)} (b + \underline{w}^T x^{(i)}) + \zeta^{(i)} \le 1 = b^{(i)}$$

$$(Ap)^{(i+m)} = -\zeta^{(i)} \le 0 = b^{(i+m)}$$ ✓

$(i = 1, \ldots, m)$
('')

⇒ Equation 5-5. Quadratic Programming problem (book, page 159) ✓