CSC 412 Machine Learning and Knowledge Discovery

Logistic Regression

1 $\ell'(\theta)$

$$\begin{aligned}
\ell(\theta) &= y \cdot \log p + (1-y) \cdot \log(1-p) \\
p &= \sigma(z) \\
z &= \theta^{\intercal} x \\
\frac{\mathrm{d}p}{\mathrm{d}z} &= \sigma(z) \cdot \big(1-\sigma(z)\big) \\
\frac{\mathrm{d}z}{\mathrm{d}\theta} &= x
\end{aligned}$$

$$\begin{aligned}
\ell'(\theta) &= \frac{\mathrm{d}\ell(\theta)}{\mathrm{d}\theta} \\
&= \frac{\mathrm{d}\ell(\theta)}{\mathrm{d}p} \cdot \frac{\mathrm{d}p}{\mathrm{d}z} \cdot \frac{\mathrm{d}z}{\mathrm{d}\theta} \\
&= \Big(y \cdot \frac{\mathrm{d}\log p}{\mathrm{d}p} + (1-y) \cdot \frac{\mathrm{d}\log(1-p)}{\mathrm{d}p}\Big) \cdot \Big(\sigma(z) \cdot \big(1-\sigma(z)\big)\Big) \cdot x \\
&= \Big(y \cdot \frac{1}{p} + (1-y) \cdot \frac{1}{1-p} \cdot (-1)\Big) \cdot \Big(\sigma(z) \cdot \big(1-\sigma(z)\big)\Big) \cdot x \\
&= \Big(y \cdot \frac{1}{\sigma(z)} - (1-y) \cdot \frac{1}{1-\sigma(z)}\Big) \cdot \Big(\sigma(z) \cdot \big(1-\sigma(z)\big)\Big) \cdot x \\
&= \left[y \cdot \frac{1}{\sigma(z)} \cdot \sigma(z) \cdot \Big(1-\sigma(z)\Big) - (1-y) \cdot \frac{1}{1-\sigma(z)} \cdot \sigma(z) \cdot \Big(1-\sigma(z)\Big)\right] \cdot x \\
&= \left[y \cdot \Big(1-\sigma(z)\Big) - (1-y) \cdot \sigma(z)\right] \cdot x \\
&= \Big(y - y \cdot \sigma(z) - \sigma(z) + y \cdot \sigma(z)\Big) \cdot x \\
&= \Big(y - \sigma(z)\Big) \cdot x \\
&= \Big(y - \sigma(\theta^{\intercal} x)\Big) \cdot x
\end{aligned}$$

So,

$$\ell'(\theta) = \Big(y - \sigma(\theta^{\intercal} x)\Big) \cdot x$$

So,

$$\boxed{\frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta}) = \frac{1}{m} \sum_{i=1}^{m} \big(\sigma(\boldsymbol{\theta}^{\intercal} \boldsymbol{x}^{(i)}) - y^{(i)}\big) x_j^{(i)}}$$