Logistic Regression Cost Function and Gradient Derivation

Cost function:

$$J(\vec{\omega}, b) = \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(\text{sigmoid}\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right)\right)+\left(1-y^{(i)}\right)\log\left(1-\text{sigmoid}\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right)\right)\right]$$

$$J(\vec{\omega}, b) = \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(\frac{1}{1+e^{-\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right)}}\right)+\left(1-y^{(i)}\right)\log\left(1-\frac{1}{1+e^{-\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right)}}\right)\right]$$

Identity:

$$e^{-z}\cdot\text{sigmoid}(z) = e^{-z}\frac{1}{1+e^{-z}}$$

$$= \frac{e^{-z}}{1+e^{-z}}$$

$$= \frac{e^{-z}+1-1}{1+e^{-z}}$$

$$= \frac{1+e^{-z}}{1+e^{-z}}-\frac{1}{1+e^{-z}}$$

$$= 1-\text{sigmoid}\,(z)$$

Useful partial derivative:

$$\frac{\partial}{\partial\omega_n}\text{sigmoid}\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right) = \frac{\partial}{\partial\omega_n}\left(\frac{1}{1+e^{-\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right)}}\right)$$

$$= \frac{\partial}{\partial\omega_n}\left(1+e^{-\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right)}\right)^{-1}$$

$$= (-1)\left(1+e^{-\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right)}\right)^{-2}\left(e^{-\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right)}\right)(-X_n)$$

$$= \frac{1}{\left(1+e^{-\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right)}\right)^2}\left(e^{-\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right)}\right)(-1)(-X_n)$$

$$= \text{sigmoid}\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right)^2\left(e^{-\left(\vec{X}^{(i)}\cdot\vec{\omega}+b\right)}\right)X_n$$

Using the identity above:

$$\frac{\partial}{\partial \omega_n}\text{sigmoid}\left(\vec{X}^{(i)} \cdot \vec{\omega} + b\right) = \text{sigmoid}\left(\vec{X}^{(i)} \cdot \vec{\omega} + b\right)\left(1 - \text{sigmoid}\left(\vec{X}^{(i)} \cdot \vec{\omega} + b\right)\right)X_n$$

Gradient of $J(\vec{\omega}, b)$ w.r.t. ω_n. Let $z = \vec{X}^{(i)} \cdot \vec{\omega} + b$:

$$\frac{\partial J(\vec{\omega}, b)}{\partial \omega_n} = \frac{\partial}{\partial \omega_n}\frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(\text{sigmoid }(z)\right) + \left(1 - y^{(i)}\right)\log\left(1 - \text{sigmoid }(z)\right)\right]$$
$$= \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}\left(\frac{1}{\text{sigmoid }(z)}\right) \cdot \frac{\partial}{\partial \omega_n}\left(\text{sigmoid }(z)\right) + \left(1 - y^{(i)}\right)\left(\frac{1}{1 - \text{sigmoid }(z)}\right) \cdot \frac{\partial}{\partial \omega_n}\left(1 - \text{sigmoid }(z)\right)\right]$$

Using the partial derivative above:

$$\frac{\partial J(\vec{\omega}, b)}{\partial \omega_n} = \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}\left(\frac{1}{\text{sigmoid (z)}}\right) \cdot \text{sigmoid}(z)^2(e^{-z})X_n^{(i)} + \left(1 - y^{(i)}\right)\left(\frac{1}{1 - \text{sigmoid (z)}}\right) \cdot (\text{-}1)\ \text{sigmoid}(z)^2(e^{-z})X_n^{(i)}\right]$$
$$= \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}\left(\frac{1}{\cancel{\text{sigmoid (z)}}}\right) \cdot \text{sigmoid}(z)^{\cancel{2}}(e^{-z})X_n^{(i)} + \left(1 - y^{(i)}\right)\left(\frac{1}{1 - \text{sigmoid (z)}}\right) \cdot (\text{-}1)\ \text{sigmoid}(z)^2(e^{-z})X_n^{(i)}\right]$$
$$= \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}\ \text{sigmoid}(z)(e^{-z})X_n^{(i)} + \left(1 - y^{(i)}\right)\left(\frac{1}{1 - \text{sigmoid (z)}}\right) \cdot (\text{-}1)\ \text{sigmoid}(z)^2(e^{-z})X_n^{(i)}\right]$$

Using the identity to sub in $(1 - \text{sigmoid}(z))$ for $(e^{-z})\text{sigmoid }(z)$:

$$\frac{\partial J(\vec{\omega}, b)}{\partial \omega_n} = \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}\left(1 - \text{sigmoid }(z)\right)X_n^{(i)} + \left(1 - y^{(i)}\right)\left(\frac{1}{1 - \text{sigmoid (z)}}\right) \cdot (\text{-}1)\ \text{sigmoid}(z)(1 - \text{sigmoid }(z))X_n^{(i)}\right]$$
$$= \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}(1 - \text{sigmoid }(z))X_n^{(i)} + \left(1 - y^{(i)}\right)\left(\frac{1}{\cancel{1 - \text{sigmoid (z)}}}\right) \cdot (\text{-}1)\ \text{sigmoid}(z)\cancel{(1 - \text{sigmoid }(z))}X_n^{(i)}\right]$$
$$= \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}(1 - \text{sigmoid }(z))X_n^{(i)} + \left(1 - y^{(i)}\right) \cdot (\text{-}1)\ \text{sigmoid}(z)X_n^{(i)}\right]$$

Now bring the -1 into the sum:

$$\frac{\partial J(\vec{\omega}, b)}{\partial \omega_n} = \frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}(\text{sigmoid}(z) - 1)X_n^{(i)} + \left(1 - y^{(i)}\right) \cdot \text{sigmoid}(z)X_n^{(i)}\right]$$

Since $y^{(i)}$ can only ever be a 0 or a 1:

$$\frac{\partial J(\vec{\omega}, b)}{\partial \omega_n} = \frac{1}{m}\sum_{i=1}^{m}\left(\text{sigmoid}(z) - y^{(i)}\right)X_n^{(i)}$$