Gradient Descent in Logistic Regression

Friday, December 23, 2022 5:52 PM

$$w, b = \underset{w, b}{\arg\min}\ L(w, b)$$

$$L(w,b) = \frac{1}{n}\sum_{i=1}^{n}\left[-y^{(i)}\log(\sigma(z^{(i)})) - (1-y^{(i)})\log(1-\sigma(z^{(i)}))\right]$$

$$z^{(i)} = w\cdot x^{(i)} + b$$

Let $L(w,b) = -y\log(\sigma(z)) - (1-y)\log(1-\sigma(z))$ $\quad z = wx+b$

$$\frac{\partial L}{\partial w_j} = -y\frac{1}{\sigma(z)}\frac{\partial\sigma}{\partial z}\frac{\partial z}{\partial w_j} - (1-y)\frac{1}{1-\sigma(z)}\frac{\partial(1-\sigma(z))}{\partial z}\frac{\partial z}{\partial w_j}$$

$$\frac{\partial\sigma}{\partial z} = \frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right) = \frac{d}{dz}\left(\frac{e^{z}}{1+e^{z}}\right) = \frac{d}{dz}\left(1-\frac{1}{1+e^{z}}\right) = -\frac{d}{dz}\left(\frac{1}{1+e^{z}}\right)$$

$$= -(-1)(1+e^{z})^{-2}e^{z} = \frac{e^{z}}{(1+e^{z})^{2}}$$

$$\frac{1}{\sigma(z)}\frac{\partial\sigma}{\partial z} = \frac{1}{\frac{1}{1+e^{-z}}}\frac{e^{z}}{(1+e^{z})^{2}} = (1+e^{-z})\frac{e^{z}}{(1+e^{z})^{2}} = \frac{e^{z}+1}{(1+e^{z})^{2}} = \frac{1}{1+e^{z}}$$

$$\frac{1}{1-\sigma(z)}\frac{\partial(1-\sigma(z))}{\partial z} = \frac{1}{1-\frac{1}{1+e^{-z}}}\left(-\frac{e^{z}}{(1+e^{z})^{2}}\right) = \frac{1+e^{-z}}{e^{-z}}\cdot\left(-\frac{e^{z}}{(1+e^{z})^{2}}\right)$$

$$= (e^{z}+1)\left(-\frac{e^{z}}{(1+e^{z})^{2}}\right) = -\frac{e^{z}}{1+e^{z}}$$

$$-y\frac{1}{\sigma(z)}\frac{\partial\sigma}{\partial z} - (1-y)\frac{1}{1-\sigma(z)}\frac{\partial(1-\sigma(z))}{\partial z} = -y\cdot\frac{1}{1+e^{z}} - (1-y)\left(-\frac{e^{z}}{1+e^{z}}\right)$$

$$= -y\frac{1}{1+e^{z}} + (1-y)\frac{e^{z}}{1+e^{z}} = \frac{e^{z}}{1+e^{z}} - y = \frac{1}{1+e^{-z}} - y = \sigma(z) - y$$

Thus, $\dfrac{\partial L}{\partial w_j} = [\sigma(z)-y]\dfrac{\partial z}{\partial w_j} = [\sigma(z)-y]x_j$

Similarly, $\dfrac{\partial L}{\partial b} = [\sigma(z)-y]\dfrac{\partial z}{\partial b} = \sigma(z)-y$

$$\Rightarrow \frac{\partial L}{\partial w_j} = \frac{1}{n}\sum_{i=1}^{n}[\sigma(z^{(i)})-y^{(i)}]x_j^{(i)}$$

$$\frac{\partial L}{\partial b} = \frac{1}{n}\sum_{i=1}^{n}(\sigma(z^{(i)})-y^{(i)})$$

$$w_j \leftarrow w_j - \eta\frac{\partial L}{\partial w_j}$$

$$b \leftarrow b - \eta\frac{\partial L}{\partial b}$$

Vectorization:

$$\left(\frac{\partial L}{\partial w_1}\cdots\frac{\partial L}{\partial w_m}\right) = \left(\frac{1}{n}\sum_{i=1}^{n}(\sigma(z^{(i)})-y^{(i)})x_1^{(i)}, \ldots, \frac{1}{n}\sum_{i=1}^{n}(\sigma(z^{(i)})-y^{(i)})x_m^{(i)}\right)$$

$$= \frac{1}{n}\left(\sigma(z^{(1)})-y^{(1)}, \cdots, \sigma(z^{(n)})-y^{(n)}\right)\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_m^{(1)} \\ \vdots & & & \\ x_1^{(n)} & x_2^{(n)} & & x_m^{(n)} \end{bmatrix}$$

$$= \frac{1}{n}(\sigma(z) - y)X$$

$y = (y^{(1)}\cdots y^{(n)})$ $\quad \sigma(z) = (\sigma(z^{(1)})\cdots\sigma(z^{(n)})) = \sigma(z^{(1)}\cdots z^{(n)})$

$z = (z^{(1)}\cdots z^{(n)}) = (w\cdot x^{(1)}+b, \cdots wx^{(n)}+b) = (w\cdot x^{(1)}\cdots w\cdot x^{(n)}) + b$