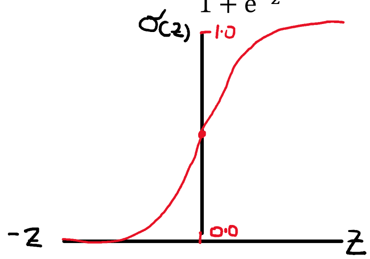
Gradient Descent - Logistic Regression

Wednesday, February 28, 2024 4:35 PM

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



$z = w^T x + b$

$\hat{y} = a = \sigma(z)$

Binary Classification:

$$L(a, y) = -[y \log(a) + (1-y)\log(1-a)]$$

We need to minimize loss

$x = [x_1 \ \ x_2]$

$W = [w_1 \ \ w_2]$

From R to L

$W^T, x \rightarrow [w_1 x_1 + w_2 x_2] \rightarrow a = \sigma(z) \rightarrow \boxed{L(a, y)}$

$dw = \frac{\partial L}{\partial w}$ $\quad da = \frac{\partial L}{\partial a}$

$dx = \frac{\partial L}{\partial x}$

Chain rule:

$$dw = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial z}{\partial w} \qquad dx = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial z}{\partial x}$$

Sigmoid derivative:

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

$$\frac{\partial a}{\partial z} = (1+e^{-z})^{-1} = -[1+e^{-z}]^{-2} \cdot (-1)e^{-z}$$

$$\frac{\partial a}{\partial z} = \frac{e^{-z}}{[1+e^{-z}]^2} = \underbrace{\frac{e^{-z}}{1+e^{-z}}}_{1-\sigma(z)} \cdot \underbrace{\frac{1}{1+e^{-z}}}_{\sigma(z)}$$

$$\frac{\partial a}{\partial z} = [1-\sigma(z)]\,\sigma(z)$$

Loss derivative:

$$L = -[y \log a + (1-y)\log(1-a)]$$

$$\frac{\partial L}{\partial a} = -\frac{\partial}{\partial a}(y \log a) - \frac{\partial}{\partial a}(1-y)\log(1-a)$$

$$\frac{\partial L}{\partial a} = \frac{-y}{a} + \frac{1-y}{1-a}$$

Derivative ∂L

Gradients.

Derivative $\frac{\partial L}{\partial z}$

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z}$$

$$\frac{\partial L}{\partial z} = \left[\frac{-y}{a} + \frac{1-y}{1-a}\right] \cdot \left[1-a\right] a$$

$$\frac{\partial L}{\partial z} = \left[-y(1-a) + (1-y)a\right] = \left[-y + \cancel{ya} + a - \cancel{ya}\right]$$

$$\frac{\partial L}{\partial z} = a - y$$

Gradients:

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z_1} \cdot \cancel{\frac{\partial z_1}{\partial w_1}}^{x_1} = [a-y] x_1$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial b_1} = [a-y]$$