Logistic Regression

Want $\quad 0 \leq h_\theta(\vec{x}) \leq 1$

- a Logistic Regression model computes a weighted sum of the input features (plus a bias term), it outputs the logistic of this result

$$h_\theta(\boldsymbol{x}) = g(\boldsymbol{\theta}^T\boldsymbol{x}) = \frac{1}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}}},$$

$g(z) = \dfrac{1}{1+e^{-z}}$ is called the logistic function or the **sigmoid** function.

$$\begin{aligned} g'(z) &= \frac{d}{dz}\frac{1}{1+e^{-z}} \\ &= \frac{1}{(1+e^{-z})^2}(e^{-z}) \\ &= \frac{1}{(1+e^{-z})}\cdot\left(1-\frac{1}{(1+e^{-z})}\right) \\ &= g(z)(1-g(z)). \end{aligned}$$

Logistic regression cost function

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\mathrm{Cost}(h_\theta(x^{(i)}, y^{(i)})$$

$$\mathrm{Cost}(h_0(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{If } y = 1 \\ -\log(1-h_\theta(x)) & \text{If } y = 0 \end{cases}$$

Note: $y = 0$ or 1 always

$$\mathrm{Cost}(h_0(x), y) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$$

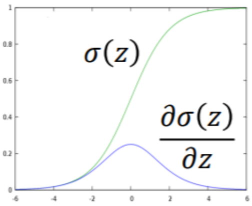
If y=1: $\quad \mathrm{Cost}(h_0(x), y) = -\log h_\theta(x)$

If y=0: $\quad \mathrm{Cost}(h_0(x), y) = -\log(1-h_\theta\,(x))$

$$\frac{-lnL(w,b)}{\partial w_i} = \sum_n -\left[\hat{y}^n \overbrace{\boxed{\frac{lnf_{w,b}(x^n)}{\partial w_i}}}^{\left(1-f_{w,b}(x^n)\right)x_i^n} + (1-\hat{y}^n)\frac{ln\left(1-f_{w,b}(x^n)\right)}{\partial w_i}\right]$$

$$\frac{\partial lnf_{w,b}(x)}{\partial w_i} = \frac{\partial lnf_{w,b}(x)}{\partial z}\frac{\partial z}{\partial w_i} \qquad \frac{\partial z}{\partial w_i} = x_i$$

$$\frac{\partial ln\sigma(z)}{\partial z} = \frac{1}{\sigma(z)}\frac{\partial\sigma(z)}{\partial z} = \frac{1}{\cancel{\sigma(z)}}\cancel{\sigma(z)}\left(1-\sigma(z)\right)$$



$$f_{w,b}(x) = \sigma(z) \qquad z = w \cdot x + b = \sum_i w_i x_i + b$$
$$= 1/1 + exp(-z)$$

$$\frac{-lnL(w,b)}{\partial w_i} = \sum_n -\left[\hat{y}^n \overbrace{\boxed{\frac{lnf_{w,b}(x^n)}{\partial w_i}}}^{\left(1-f_{w,b}(x^n)\right)x_i^n} + (1-\hat{y}^n)\overbrace{\boxed{\frac{ln\left(1-f_{w,b}(x^n)\right)}{\partial w_i}}}^{f_{w,b}(x^n)x_i^n}\right]$$

$$\frac{\partial ln\left(1-f_{w,b}(x)\right)}{\partial w_i} = \frac{\partial ln\left(1-f_{w,b}(x)\right)}{\partial z}\frac{\partial z}{\partial w_i} \qquad \frac{\partial z}{\partial w_i} = x_i$$

$$\frac{\partial ln\left(1-\sigma(z)\right)}{\partial z} = -\frac{1}{1-\sigma(z)}\frac{\partial\sigma(z)}{\partial z} = -\frac{1}{\cancel{1-\sigma(z)}}\sigma(z)\cancel{\left(1-\sigma(z)\right)}$$

$$f_{w,b}(x) = \sigma(z) \qquad z = w \cdot x + b = \sum_i w_i x_i + b$$
$$= 1/1 + exp(-z)$$

$$\frac{-lnL(w,b)}{\partial w_i} = \sum_n -\left[\hat{y}^n \overbrace{\boxed{\frac{lnf_{w,b}(x^n)}{\partial w_i}}}^{\left(1-f_{w,b}(x^n)\right)x_i^n} + (1-\hat{y}^n)\overbrace{\boxed{\frac{ln\left(1-f_{w,b}(x^n)\right)}{\partial w_i}}}^{f_{w,b}(x^n)x_i^n}\right]$$

$$= \sum_n -\left[\hat{y}^n\left(1-f_{w,b}(x^n)\right)x_i^n - (1-\hat{y}^n)f_{w,b}(x^n)x_i^n\right]$$

$$= \sum_n -\left[\hat{y}^n - \cancel{\hat{y}^n f_{w,b}(x^n)} - f_{w,b}(x^n) + \cancel{\hat{y}^n f_{w,b}(x^n)}\right]x_i^n$$

$$= \sum_n -\left(\hat{y}^n - f_{w,b}(x^n)\right)x_i^n$$

Larger difference, larger update

$$w_i \leftarrow w_i - \eta\sum_n -\left(\hat{y}^n - f_{w,b}(x^n)\right)x_i^n$$