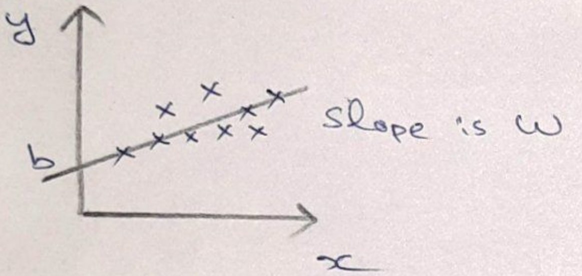
Simple linear Regression.

$$\boxed{f_{w,b}(x) = wx + b}$$

or

$$\hat{Y}_i = wx_i + b$$

$\hat{Y}_i$ = estimated Y_i



Cost function - MSE

$$\boxed{J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{Y}_i - Y_i)^2}$$

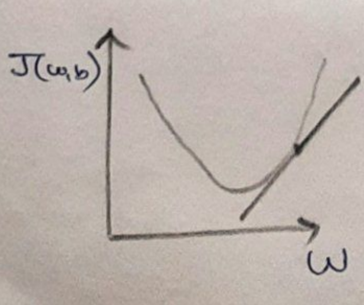
m = number of Samples.

Gradient Desent

Minimize $J(w,b)$

$$\boxed{\begin{aligned} W &= W - \alpha \frac{d}{dw} J(w,b) \\ b &= b - \alpha \frac{d}{db} J(w,b) \end{aligned}}$$

α is learning rate.



$\frac{d}{dw} J(w,b)$ is derivative. ie slope of tangent.

$$\frac{d}{dw} J(w,b) = \frac{d}{dw} \frac{1}{2m} \sum_{i=1}^{m} (\hat{Y}_i - Y_i)^2$$

$$= \frac{d}{dw} \frac{1}{2m} \sum_{i=1}^{m} (wx_i + b - Y_i)^2$$

$$= \frac{1}{2m} \times \sum_{i=1}^{m} (wx_i + b - Y_i) \times 2 \times x_i$$

$$= \frac{1}{m} \sum_{i=1}^{m} (wx_i + b - Y_i)\, x_i$$

$$= \frac{1}{m} \sum_{i=1}^{m} (\hat{Y}_i - Y_i)\, x_i$$

$$\frac{d}{db} J(w,b) = \frac{d}{db} \frac{1}{2m} \sum_{i=1}^{m} (\hat{Y}_i - Y_i)^2$$

$$= \frac{d}{db} \frac{1}{2m} \sum_{i=1}^{m} (wx_i + b - Y_i)^2$$

$$= \frac{1}{2m} \times \sum_{i=1}^{m} (wx_i + b - Y_i) \times 2$$

$$= \frac{1}{m} \sum_{i=1}^{m} (\hat{Y}_i - Y_i)$$

Since $\hat{Y}_i$ is $f(x_i)$

$$w = w - \alpha \frac{d}{dw} J(w,b)$$

$$\boxed{w = w - \alpha \frac{1}{m} \sum_{i=1}^{m} (f(x_i) - Y_i)\, x_i}$$

$$b = b - \alpha \frac{d}{db} J(w,b)$$

$$\boxed{b = b - \alpha \frac{1}{m} \sum_{i=1}^{m} (f(x_i) - Y_i)}$$

Multiple linear Regression

$$f(x) = w_1 x_1 + w_2 x_2 + \cdots\cdots + w_n x_n + b$$

$$\vec{w} = [w_1, w_2, w_3, \cdots\cdots w_n]$$

$$\vec{x} = [x_1, x_2, \cdots\cdots\cdots x_n]$$

$$\boxed{f(x) = \vec{w} \cdot \vec{x} + b}$$

numpy dot Product - Vectorization.

Cost function - MSE

$$\boxed{J(w_1, w_2, \cdots b) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{Y}_i - Y_i)^2}$$

Gradient Descent

$$\boxed{\begin{aligned} w_i &= w_i - \alpha \frac{d}{dw_i} J(w_1, w_2, \cdots b) \\ b &= b - \alpha \frac{d}{db} J(w_1, w_2, \cdots b) \end{aligned}}$$

$$\frac{d}{dw_i} J(w_1 w_2, \cdots b) = \frac{1}{m} \sum_{i=1}^{m} (\hat{Y}_i - Y_i) x_i$$

$$\frac{d}{db} J(w_1 w_2 \cdots b) = \frac{1}{m} \sum_{i=1}^{m} (\hat{Y}_i - Y_i)$$

$$w_1 = w_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (f(x_1) - Y_1) x_1$$

$$w_2 = w_2 - \alpha \frac{1}{m} \sum_{i=1}^{m} (f(x_2) - Y_2) x_2$$

$$\vdots$$

$$w_n = w_n - \alpha \frac{1}{m} \sum_{i=1}^{m} (f(x_n) - Y_n) x_n$$

$$b = b - \alpha \frac{1}{m} \sum_{i=1}^{m} (f(x_i) - Y_i)$$

$$\boxed{\begin{aligned} w_i &= w_i - \alpha \frac{1}{m} \sum_{i=1}^{m} (f(x_i) - Y_i) x_i \\ b &= b - \alpha \frac{1}{m} \sum_{i=1}^{m} (f(x_i) - Y_i) \end{aligned}}$$

Logistic Regression

Sigmoid / logistic function.

$$g(z) = \frac{1}{1+e^{-z}} \qquad 0 < g(z) < 1$$

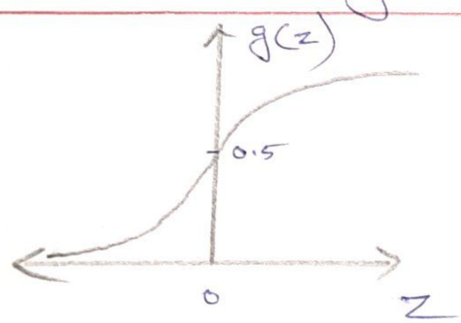
$$f(x) = g(\vec{w} \cdot \vec{x} + b)$$

$$\boxed{f(x) = \frac{1}{1+e^{-(\vec{w} \cdot \vec{x} + b)}}}$$

Decision boundary

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

Decision boundary at $Z=0$



$g(z) \geq 0.5$

$z \geq 0$

$\vec{w} \cdot \vec{x} + b \geq 0$

$\hat{Y} = 1$

when $g(z) < 0.5$ then $\hat{Y} = 0$

$\hat{Y}$ estimate is decided based on $f(x)$

$f(x) \geq$ decision boundary

$f(x) <$ decision boundary.

Cost function

MSE will not be Convex for logistic regression.

$$L(f(x), Y_i) = \begin{cases} -\log f(x) & Y_i = 1 \\ -\log(1-f(x)) & Y_i = 0 \end{cases}$$

$$J(w,b) = \frac{1}{m} \sum_{i=1}^{m} L(f(x), Y_i)$$

$$\boxed{J(w,b) = -\frac{1}{m} \sum_{i=1}^{m} \left(\log f(x) \times Y_i\right) + \left(\log[1-f(x)] \times (1-Y_i)\right)}$$

Gradient Descent

$$w = w - \alpha \frac{d}{dw} J(w,b)$$

$$b = b - \alpha \frac{d}{db} J(w,b)$$

$$w_i = w_i - \alpha \frac{1}{m} \sum_{i=1}^{m} (f(x_i) - Y_i)\, x_i$$

$$b = b - \alpha \frac{1}{m} \sum_{i=1}^{m} (f(x_i) - Y_i)$$

Where

$$f(x_i) = \frac{1}{1+e^{-(\vec{w} \cdot \vec{x} + b)}}$$

Regularization

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}(f(x_i) - Y_i)^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$$

λ = Regularization Parameter.

$\frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$ is Regularization term.

$\lambda = 10^{10}$ large the $w_i \approx 0$

Then it cause Underfitting

$\lambda \approx 0$ Small then Overfit.

Regularization Linear Regression

We want to minimize $J(w,b)$

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}(f(x_i) - Y_i)^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$$

$$w_j = w_j - \alpha \frac{d}{dw} J(\vec{w}, b)$$

$$b = b - \alpha \frac{d}{db} J(\vec{w}, b)$$

$$\frac{d}{dw_j} J(\vec{w}, b) = \frac{1}{m}\sum_{i=1}^{m}(f(x_i) - Y_i)x_j + \frac{\lambda}{m} w_j$$

$$\frac{d}{db} J(\vec{w}, b) = \frac{1}{m}\sum_{i=1}^{m}(f(x_i) - Y_i)$$

$$w_j = w_j - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}(f(x_i) - Y_i)x_j + \frac{\lambda}{m} w_j\right]$$

$$w_j = w_j\left(1 - \frac{\alpha\lambda}{m}\right) - \alpha\frac{1}{m}\sum_{i=1}^{m}(f(x_i) - Y_i)x_j$$

$m = 5$ $\alpha = 0.01$ $\lambda = 1$

$$\frac{\alpha\lambda}{m} = 0.01 \times \frac{1}{50} = 0.0002$$

$1 - 0.002 = 0.9998$

So w_j is decreasing at each step.

Regularized logistic Regression

$$J(w,b) = -\frac{1}{m}\sum_{i=1}^{m}\left[Y_i \log(f(x_i)) + (1 - Y_i)\log(1 - f(x_i))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$$

$$w_j = w_j - \alpha \frac{d}{dw} J(\vec{w}, b)$$

$$b = b - \alpha \frac{d}{db} J(\vec{w}, b)$$

$\frac{d}{dw} J(\vec{w}, b)$ and $\frac{d}{db} J(\vec{w}, b)$ and w_j and b equation is Same.

But $f(x) = \frac{1}{1 + e^{-(\vec{w}\cdot\vec{x} + b)}}$

No Change in b, Because in both cases we are not regularizing b.