Gradient Descent for linear regression:-

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

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

$$= \frac{1}{2m}\sum_{i=1}^{m}\left(f_{w,b}(x_i) - y_i\right)^2$$

Gradient Descent:-

Repeat until convergences

$$\{$$

$$w = w - \alpha \boxed{\frac{\partial}{\partial w} J(w,b)}$$

$$b = b - \alpha \boxed{\frac{\partial}{\partial b} J(w,b)}$$

$$\}$$

$$\frac{\partial}{\partial w} J(w,b) = \frac{\partial}{\partial w}\frac{1}{2m}\sum_{i=1}^{m}\left(f_{w,b}(x_i) - y_i\right)^2$$

$$= \frac{\partial}{\partial w}\frac{1}{2m}\sum_{i=1}^{m}(wx_i + b - y_i)^2$$

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

$$\boxed{\frac{\partial J(w,b)}{\partial w} = \frac{1}{m}\sum_{i=1}^{m}\left(f_{w,b}(x_i) - y_i\right)x_i}$$

$$\frac{\partial}{\partial b} J(w,b) = \frac{\partial}{\partial b}\frac{1}{2m}\sum_{i=1}^{m}\left(f_{w,b}(x_i) - y_i\right)^2$$

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

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

$$\boxed{\frac{\partial}{\partial b} J(w,b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{w,b}(x_i) - y_i\right)}$$

Logistic Regression:-

Binary Classification

Cat v/s Non-cat
(1) (0)

(x, y) $x \in \mathbb{R}^n$
$y \in \{0, 1\}$

m training examples :—

$$\{ (x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots \ldots (x^{(m)}, y^{(m)}) \}$$

$$X = \begin{bmatrix} | & | & | & & | \\ x^{(1)} & x^{(2)} & x^{(3)} & \cdots & x^{(m)} \\ | & | & | & & | \end{bmatrix} \Big\updownarrow n$$

$$X \in \mathbb{R}^{n \times m}$$

$$X.\text{shape} = (n, m)$$

$$Y = [y^{(1)} \; y^{(2)} \; y^{(3)} \ldots \ldots y^{(m)}]$$

$$Y \in \mathbb{R}^{1 \times m}$$

$$Y.\text{shape} = (1, m)$$

Given x, $\hat{y} = P(y = 1 / x)$

$$\hat{y} = \sigma(w^T b + x)$$

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

If z is large $\sigma(z) \simeq \frac{1}{1+0} \simeq 1$

If z is small $\sigma(z) \simeq \frac{1}{1+\infty} \simeq 0$

Cost Function:-

$$\mathcal{L}(\hat{y}, y) = -\left(y \log \hat{y} + (1-y) \log(1-\hat{y})\right)$$

$$J(w, b) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}(\hat{y}^{(i)}, y^{(i)})$$

Binary Cross-entropy Loss.