Machine Learning

Logistic Regression Classifier

Dr. Mehran Safayani

safayani@iut.ac.ir

safayani.iut.ac.ir

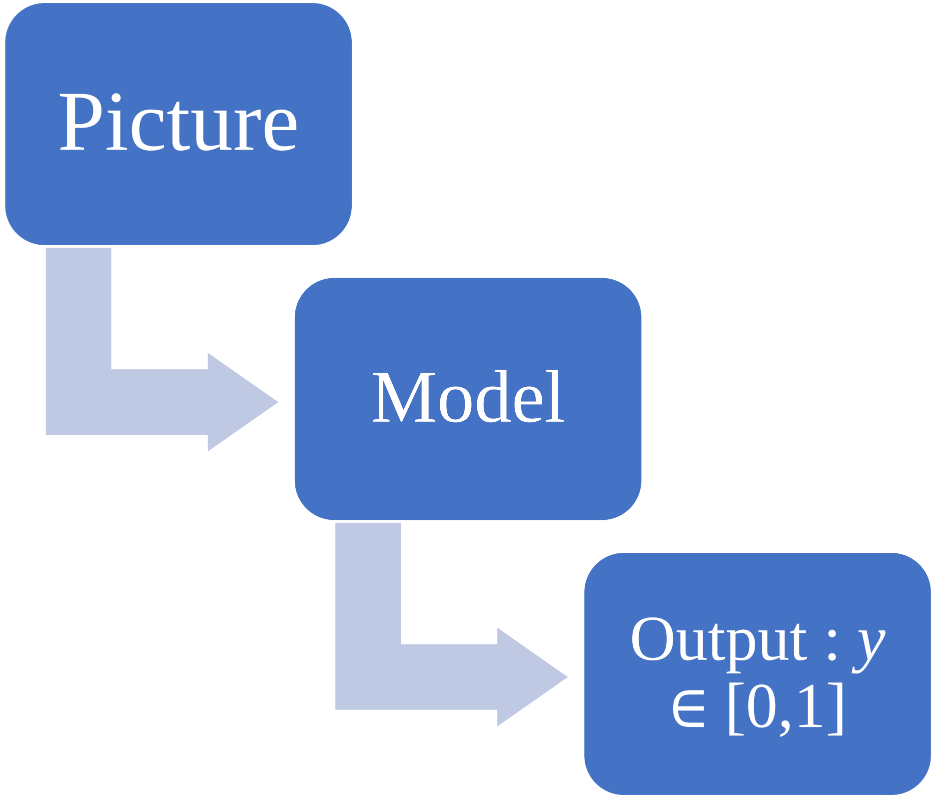
https://www.aparat.com/mehran.safayani

https://github.com/safayani/machine_learning_course

Department of Electrical and computer engineering, Isfahan university of technology, Isfahan, Iran

Classification

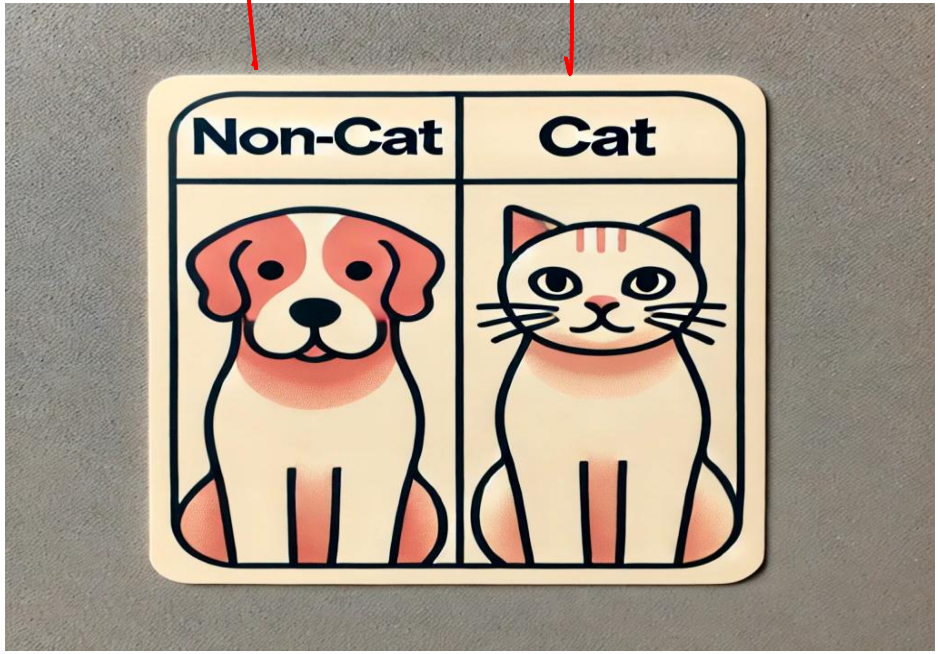
- Email: spam/not spam
- Animal: cat/non cat



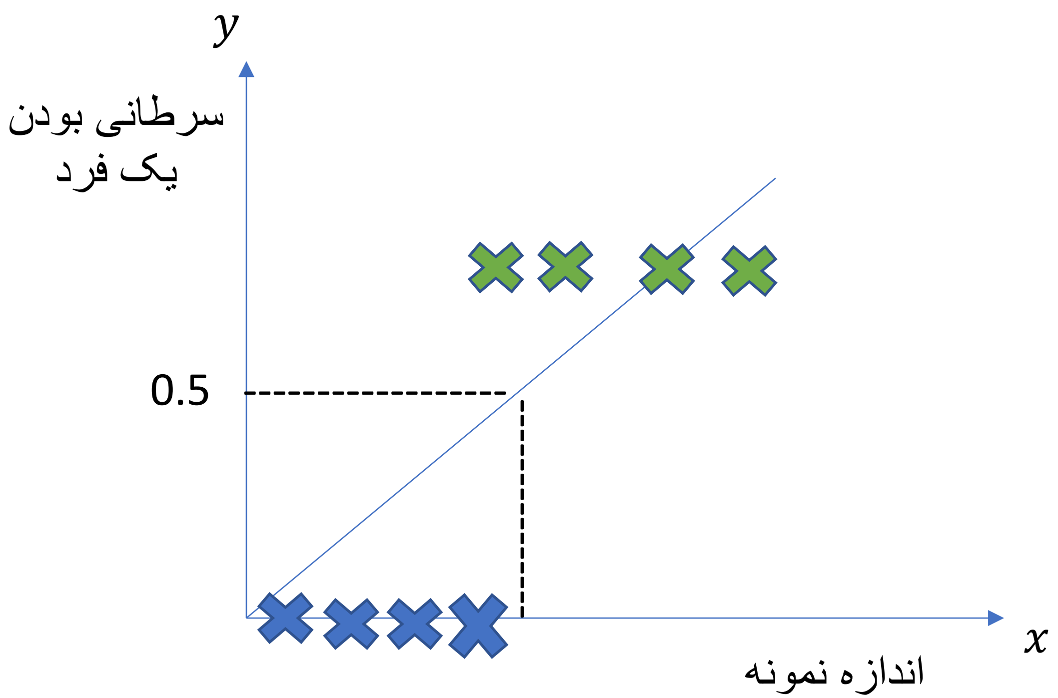
$y \in \{0, 1\}$

0: “Negative Class”

1: “Positive Class”



Classification



$If: h_\theta(x) \geq 0.5 \rightarrow predict, y = 1$
$If: h_\theta(x) \leq 0.5 \rightarrow predict, y = 0$

$0 \leq h_\theta(x) \leq 1$
$0 \leq h_\theta(x) = P(y = 1)|x) \leq 1$

Logistic Regression

مبانی رگرسیون لجستیک:

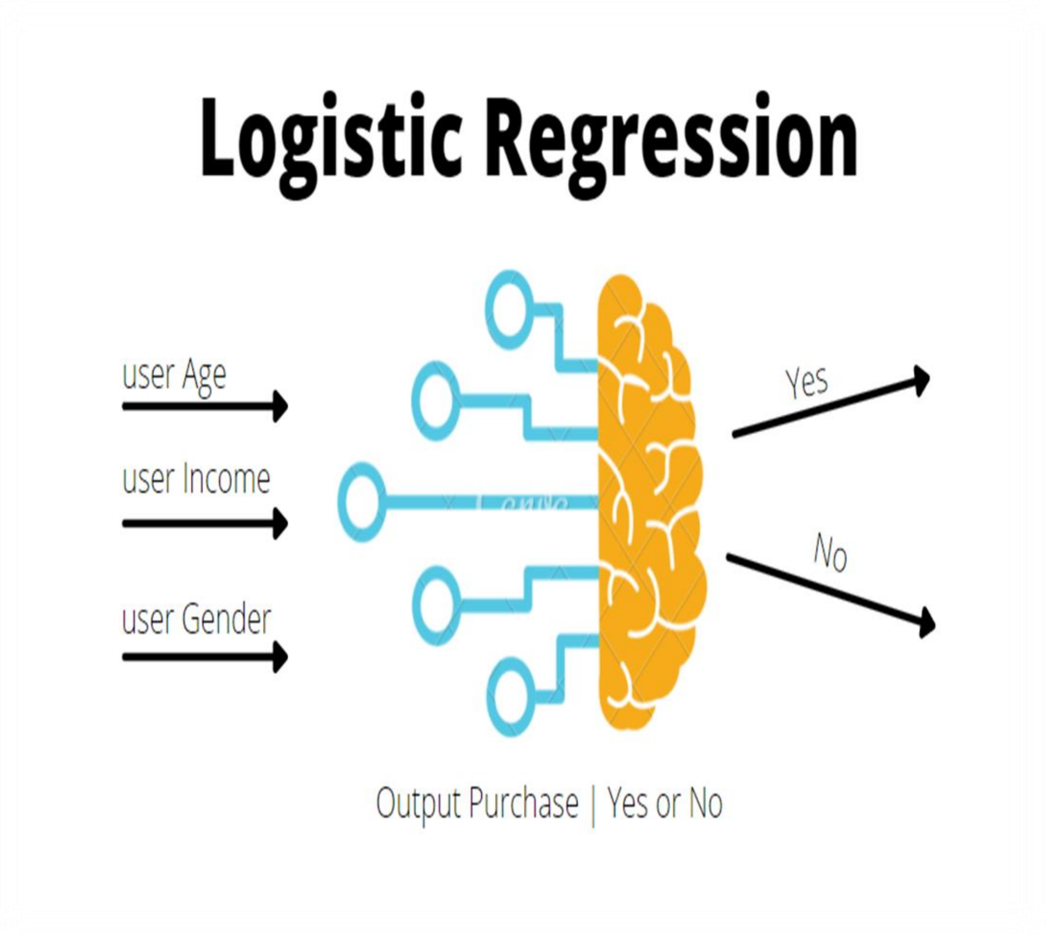
- عمدتاً برای مسائل طبقه‌بندی دودویی استفاده می‌شود.
- احتمال تعلق یک ورودی به یکی از دو کلاس را مدل‌سازی می‌کند.

مرزهای تصمیم‌گیری خطی:

- رگرسیون لجستیک به‌طور معمول مرزهای تصمیم‌گیری خطی را در فرم استاندارد خود ایجاد می‌کند.

کاربردها:

- در مسائلی مانند طبقه‌بندی تصویر، تشخیص پزشکی و همچنین جایی که سطوح تصمیم پیچیده برای بهبود دقت ضروری هستند، استفاده می شود.

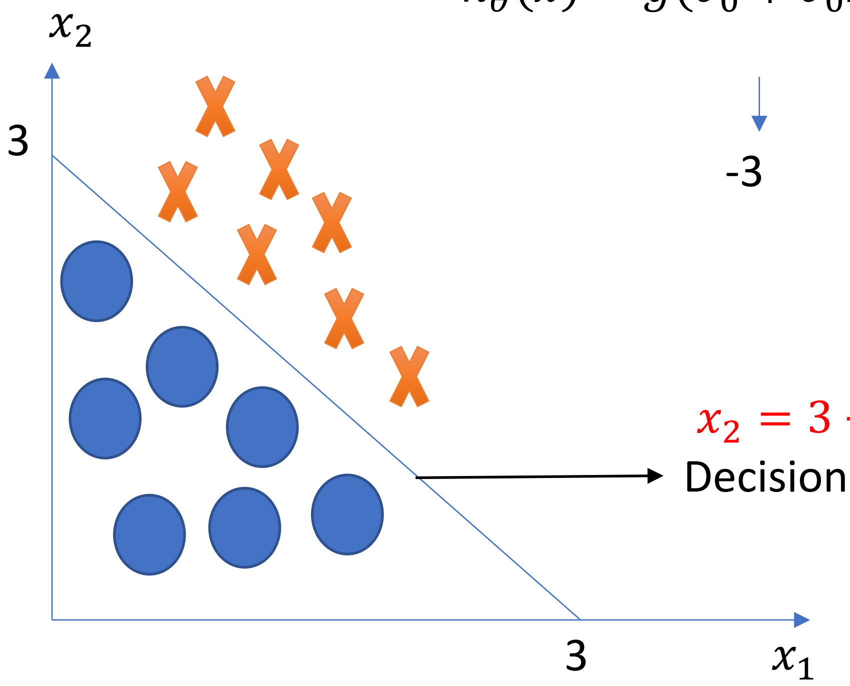


Decision Boundary

$$\theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

$$h_\theta(x) = g(\theta_0 + \theta_0 x_1 + \theta_2 x_2)$$

-3 1 1



$x_2 = 3 - x_1$

Decision Boundary

Predict $y = 1, if \underbrace{-3 + x_1 + x_2 \geq 0}_{\theta^T X}$

$$x_1 + x_2 \geq 3$$

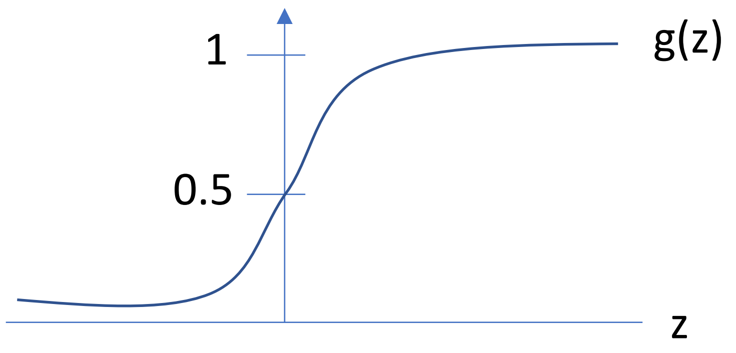
Sigmoid Function: Logistic Function

Want $0 \leq h_\theta(x) \leq 1$

$$h_\theta(x) = g(\theta^T x)$$

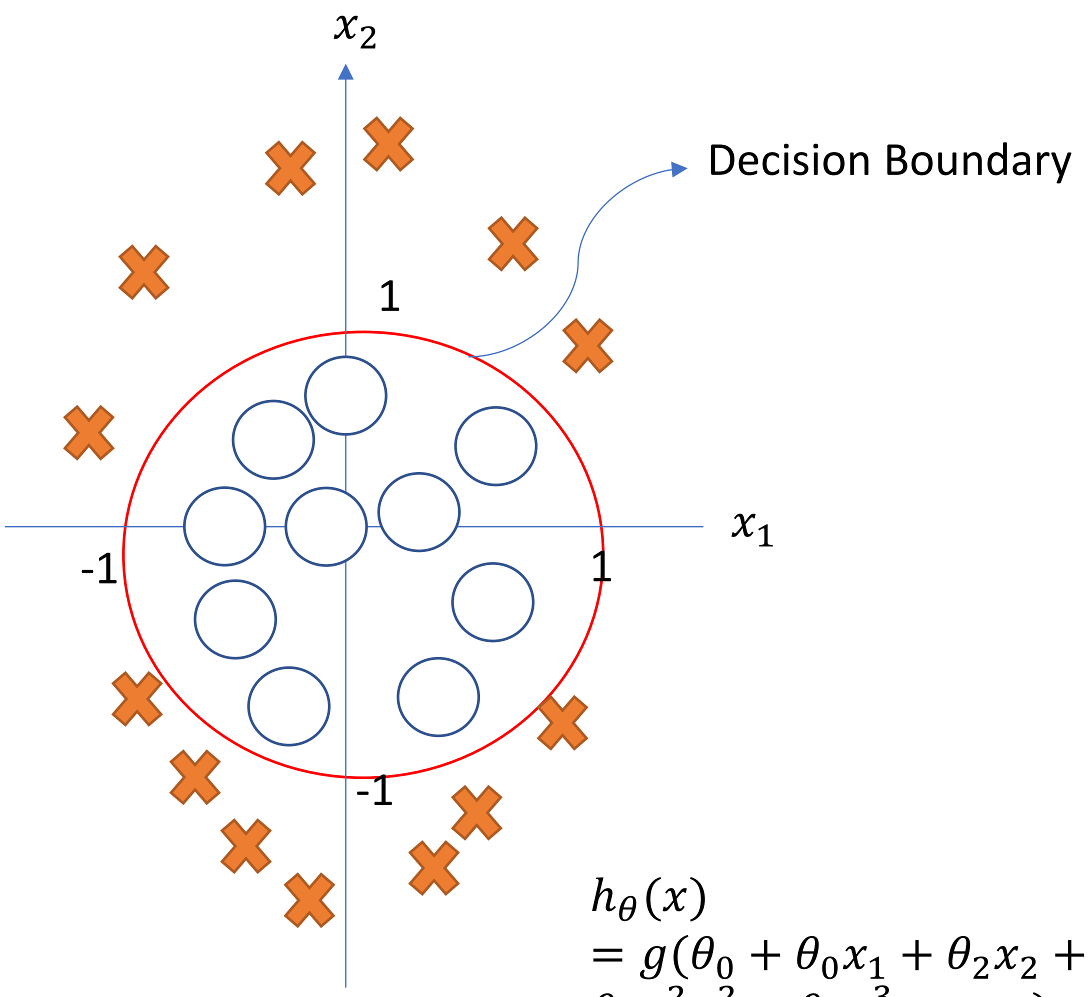
$$g(z) = \frac{1}{1 + e^{-z}}$$

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



Model output example : $P(y = 1|x) = 0.8$

Non-Linear Decision Boundaries



-1 0 0

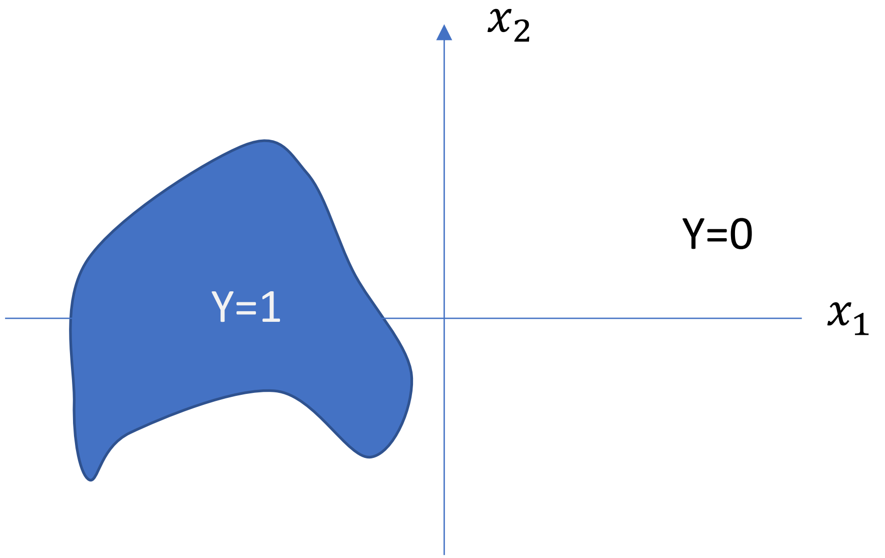
$$h_\theta(x) = g(\theta_0 + \theta_0 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

$$\theta = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Predict $y = 1, if \underbrace{-3 + x_1^2 + x_2^2 \geq 0}_{\theta^T X}$

$$x_1{}^2 + x_2{}^2 \geq 1$$

$$h_\theta(x) = g(\theta_0 + \theta_0 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_1^2 x_2 + \theta_5 x_1^2 x_2^2 + \theta_6 x_1^3 x_2 + \ldots)$$



Cost Function

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

$$x \in \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix},\ x_0 = 1,\ y \in \{0,1\}$$

Linear Regression: $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2$

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

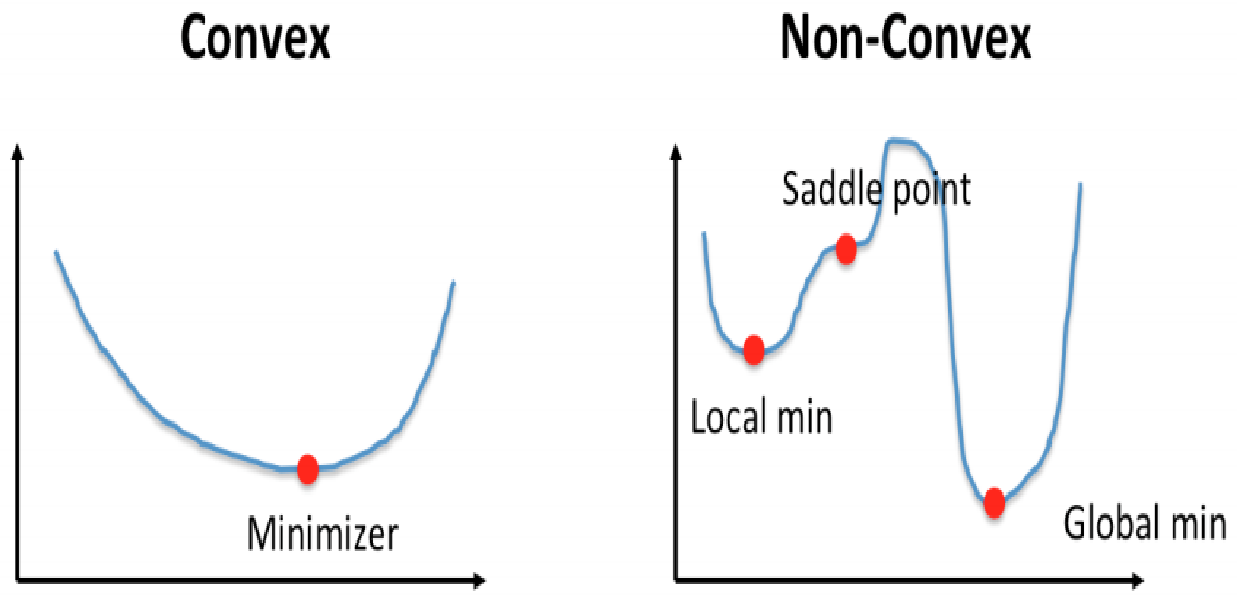
Logistic Regression: $h_\theta(x^i) = g(\theta^T x^i)$

MSE: $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (g(\theta^T x^i) - y^i)^2$

Convex VS Non-Convex Cost Function

Convex Cost Function:
•Bowl-shaped with a single global minimum.
•Easier optimization, guarantees finding the global minimum.

Non-Convex Cost Function:
•Contains multiple local minima.
•**Challenges**: Optimization algorithms can get stuck in local minima.
•**Impact**: Risk of poor model performance, as finding the global minimum is difficult.



Logistic Regression

MSE Cost Function:

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}((h_\theta(x^{(i)}) - y^{(i)})^2)$$

Binary Cross Entropy cost function:

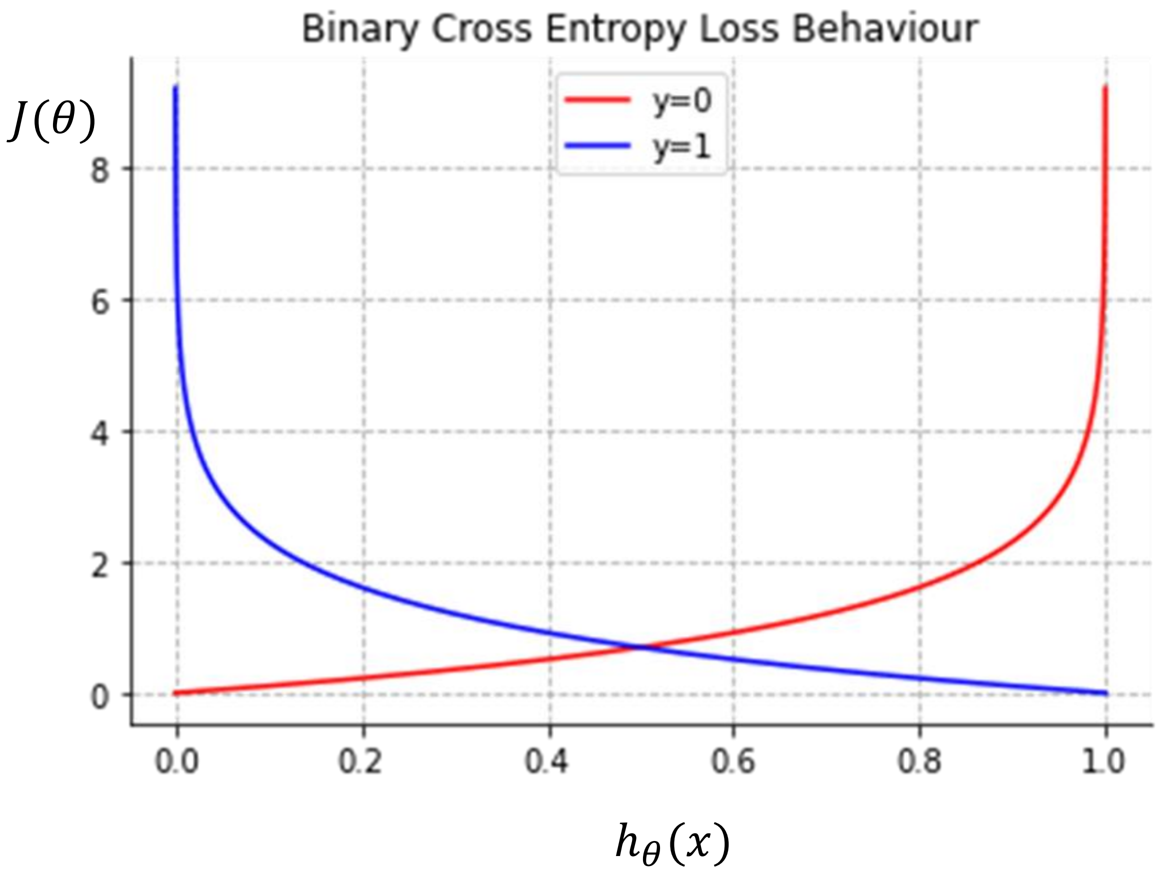
$$cost(h_\theta(x), y) = \begin{cases} -log(h_\theta(x)), & if \quad y = 1 \\ -log(1 - h_\theta(x)), & if \quad y = 0 \end{cases}$$

if y=1 and $h_\theta(x) = 1 \Rightarrow cost = 0;$

if y=1 and $h_\theta(x) = 0 \Rightarrow cost = \infty;$

if y=0 and $h_\theta(x) = 0 \Rightarrow cost = 0;$

if y=0 and $h_\theta(x) = 1 \Rightarrow cost = \infty;$



Logistic Regression

$$cost(h_\theta(x), y) = \begin{cases} -log(h_\theta(x)), & if \quad y = 1 \\ -log(1 - h_\theta(x)), & if \quad y = 0 \end{cases}$$

$$cost(h_\theta(x), y) = -ylog\big(h_\theta(x)\big) - (1 - y)log(1 - h_\theta(x))$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \underbrace{-y^i log\left(h_\theta\left(x^i\right)\right) - (1 - y^i)log(1 - h_\theta(x^i))}_{T_i}$$

$$\min_\theta J(\theta)$$

Gradient Descent: Repeat Unit Convergence :

repeat{

$$\theta_j = \theta_j - \alpha \frac{dJ(\theta)}{d_\theta} \quad J = 0, \ldots, n$$

}until convergence

$$\frac{dJ(\theta)}{d_\theta} = ? \qquad J(\theta) = \frac{1}{m}\sum_{i=1}^{m} T_i \qquad \frac{dJ(\theta)}{d_\theta} = \frac{1}{m}\sum_{i=1}^{m} \frac{dT_i}{di}$$

$$T_i = -[y^i \log h_\theta(x^i) + (1 - y^i)\log(1 - h_\theta(x^i))] \qquad h_\theta(x^i) = \sigma(\theta^T x^i) = \sigma(z^i) = \frac{1}{1+e^{-z^i}}$$

$$T_i = -[y^i \log\sigma(z^i) + (1 - y^i)\log(1 - \sigma(z^i)]$$

$$(1)\ \frac{dT_i}{d\sigma(z^i)} = -\left[\frac{y^i}{\sigma(z^i)} + (1 - y^i).\frac{-1}{1 - \sigma(z^i)}\right] = -\left[\frac{y^i}{\sigma(z^i)} - \frac{1 - y^i}{1 - \sigma(z^i)}\right]$$

$$(2)\ \frac{d\sigma(z^i)}{dz^i} = \frac{e^{-z^i}}{(1 + e^{-z^i})^2} = \frac{1}{1 + e^{-z^i}}.\frac{e^{-z^i}}{1 + e^{-z^i}} = \sigma(z^i).(1 - \sigma(z^i)) \qquad ,$$

$$(3)\ \frac{dz^i}{d\theta_j} = x_j^i$$

$$z^i = \theta^T x^i = \theta_0 x_0 + \theta_1 x_1 + \ldots \theta_n x_n$$

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

From (1), (2) and (3):

$$\frac{dT_i}{d\theta_j} = -\left[\frac{y^i}{\sigma(z^i)} - \frac{1-y^i}{1-\sigma(z^i)}\right]\sigma(z^i).\,(1-\sigma(z^i))x_j^i$$

$$= -\left[y^i.\left(1-\sigma(z^i)\right) - \left(1-y^i\right).\sigma(z^i)\right]x_j^i$$

$$= -\left[y^i - \sigma(z^i)\right]x_j^i = \left[\sigma(z^i) - y^i\right]x_j^i$$

$$\frac{dy(\theta)}{d\theta_j} = \frac{1}{m}\sum_{i=1}^{m}\frac{dT_i}{d\theta_j} = \frac{1}{m}\sum_{i=1}^{m}\frac{dT_i}{dz^i}.\frac{dz^i}{d\theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left(\underbrace{\sigma(z^i)}_{h_\theta(x^i)} - y^i\right).x_j^i$$

$$GD{:}\ \mathrm{R}peatUntiConvergance\{$$

$$\theta_j = \theta_j - \alpha\sum_{i=1}^{m}\left(h_\theta\left(x^i\right) - y^i\right).x_j^i$$

$$\}$$

Logistic regression on m examples

$\theta_1 \leftarrow \text{random} \quad \theta_2 \leftarrow \text{random} \quad b \leftarrow \text{random}$

$$\theta^t = \begin{bmatrix} \theta_1^t \\ \theta_2^t \\ b^t \end{bmatrix} \quad \theta^{t+1} = \begin{bmatrix} \theta_1^{t+1} \\ \theta_2^{t+1} \\ b^{t+1} \end{bmatrix}$$

$$\|\theta^{t+1} - \theta^t\|_2 \leq \varepsilon$$

Repeat{

$J = 0; \quad d\theta_1 = 0; \quad d\theta_2 = 0; \quad db = 0;$

For i=1 to m

$$z^{(i)} = \theta^T x^{(i)} + b$$

$$a^{(i)} = \sigma\left(z^{(i)}\right)$$

$$J \mathrel{+}= \left[y^{(i)} Log a^{(i)} + \left(1 - y^{(i)}\right) Log\left(1 - a^{(i)}\right)\right]$$

$$dz^{(i)} = a^{(i)} - y^{(i)}$$

$$d\theta_1 \mathrel{+}= x_1^{(i)}\, dz^{(i)} \quad d\theta_2 \mathrel{+}= x_2^{(i)}\, dz^{(i)} \quad db \mathrel{+}= dz^{(i)}$$

$$d\theta = \begin{bmatrix} d\theta_1 \\ d\theta_2 \\ db \end{bmatrix}$$

$$\|d\theta\| \leq \varepsilon = 10^{-4}$$

$J /= m; \quad d\theta_1 /= m; \quad d\theta_2 /= m; \quad db /= m;$

$\theta_1 = \theta_1 - \alpha\, d\theta_1 \quad \theta_2 = \theta_2 - \alpha\, d\theta_2 \quad b = b - \alpha\, db$

} until convergence

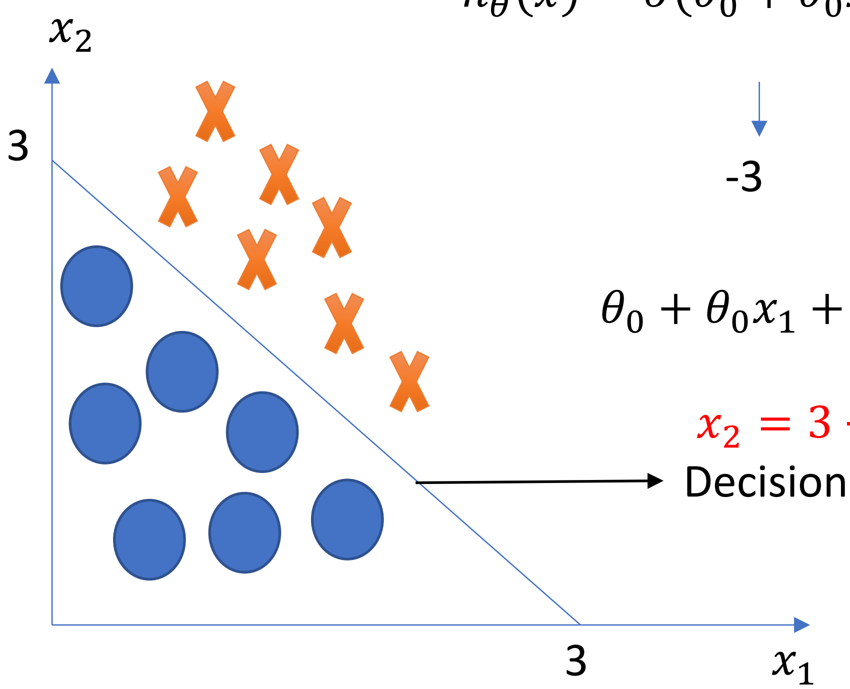
Decision Boundary of logistic regression

$$\theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

$$h_\theta(x) = \sigma(\theta_0 + \theta_0 x_1 + \theta_2 x_2)$$

-3 1 1

$\theta_0 + \theta_0 x_1 + \theta_2 x_2$=-3+ $x_1 + x_2 = 0$



$x_2 = 3 - x_1$

Decision Boundary