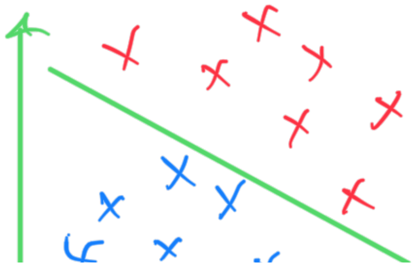
Logistic Regression:

Machine learning

- Supervised learning
  - Regression
    - Predict continuous values
    - linear regression
  - Classification
    - predict discrete labels
    - logistic regression
- Unsupervised learning

Binary logistic regression

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

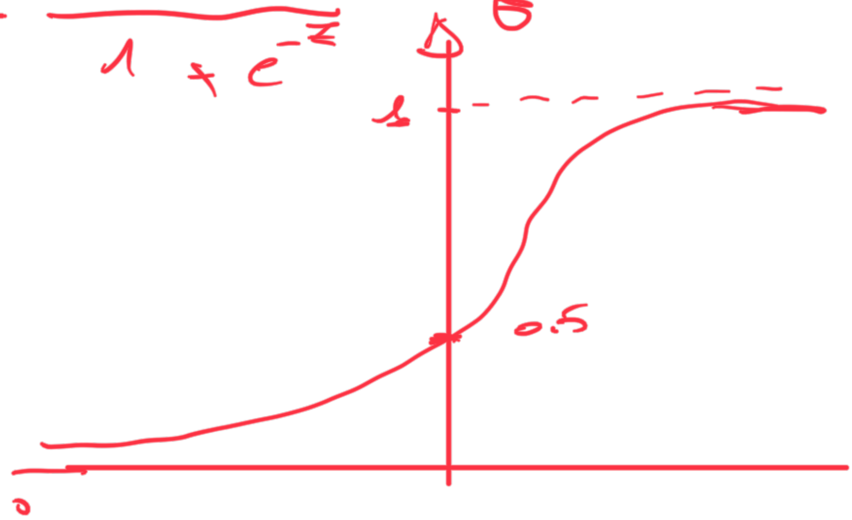


problem $\longrightarrow$ $\boxed{\hat{y} = x\theta}$ $\quad y \in \{0,1\}$

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

$$\boxed{\hat{y} = \sigma(\underbrace{x\theta}_{z})}$$

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



$\lim_{z \to -\infty} \sigma(z) = 0$

$\lim_{z \to +\infty} \sigma(z) = 1$

Assume a Threshold

Predict $\hat{y} = 1$ if $\sigma(z) \geq 0.5$

Predict $\hat{y} = 0$ if $\sigma(z) < 0.5$

finding good θ

$\Downarrow$

loss function

$\Longrightarrow$ define a new loss function

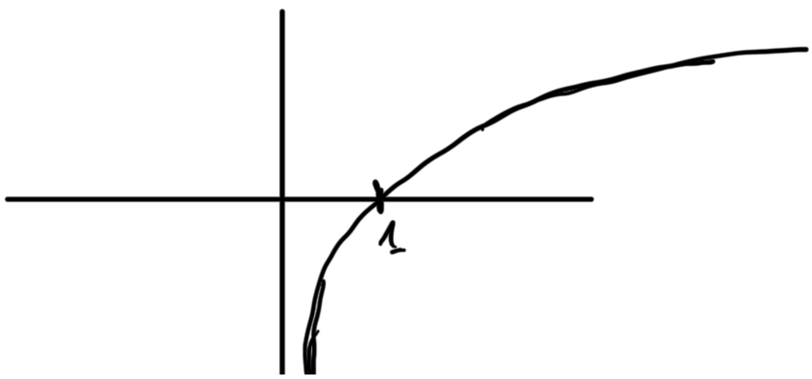
Binary cross entropy loss

$$J(\theta) = -\frac{1}{N} \sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)$$

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

$$\text{cost}(y,\hat{y}) = -y \log(\hat{y}) - (1-y)\log(1-\hat{y})$$

$$\text{Cost}(y,\hat{y}) = \begin{cases} -\log(\hat{y}) & \text{if } y=1 \\ -\log(1-\hat{y}) & \text{if } y=0 \end{cases}$$



$\lim_{0} \log(x) = -\infty$

$\lim_{+\infty} \log(x) = +\infty$

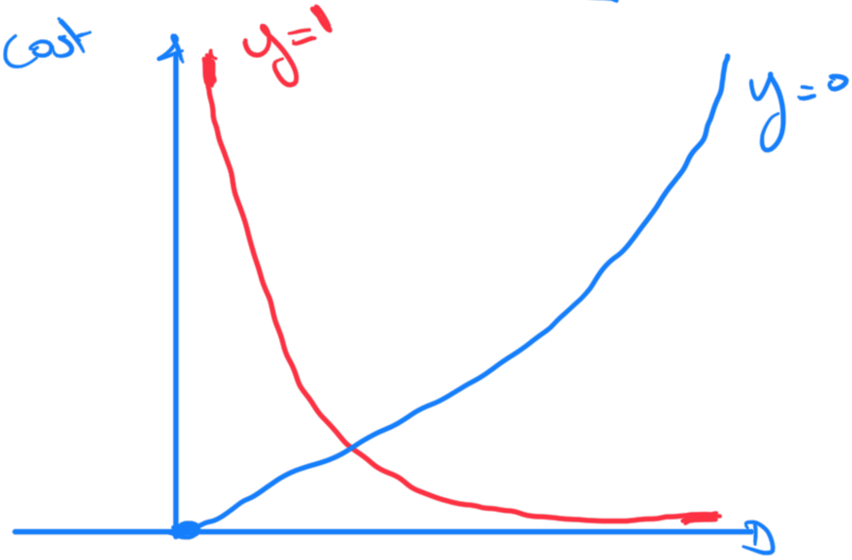
$\log(1) = 0$

14

$$\text{cost}\begin{cases} -\log(\hat{y}) & \text{if } y=1 \\ -\log(1-\hat{y}) & \text{if } y=0 \end{cases}$$

$$\begin{cases} \boxed{-\log(\sigma(X\theta)) \quad \text{if } y=1} \\ \boxed{-\log(1-\sigma(X\theta)) \quad \text{if } y=0} \end{cases}$$

$\hat{y} = \sigma(X\theta)$



$\hat{y} = \sigma(X\theta)$

if $y=1$ and $\hat{y} = \sigma(X\theta) \to 0$

$-\log(\hat{y}) \to +\infty$

$\hat{y} = \sigma(X\theta) \to 1$

if $y=0$ and $\hat{y} = \sigma(X\theta) \to 0$

$\log(1-\hat{y}) = 0$

if $y=0$ and $\hat{y} = \sigma(X\theta) \to 1$

How to find the optimal θ

$J(\theta)$

$\hat{y} = f(x\theta)$

$$\frac{\partial J(\theta)}{\partial \theta} = 0$$

Gradient Descent:

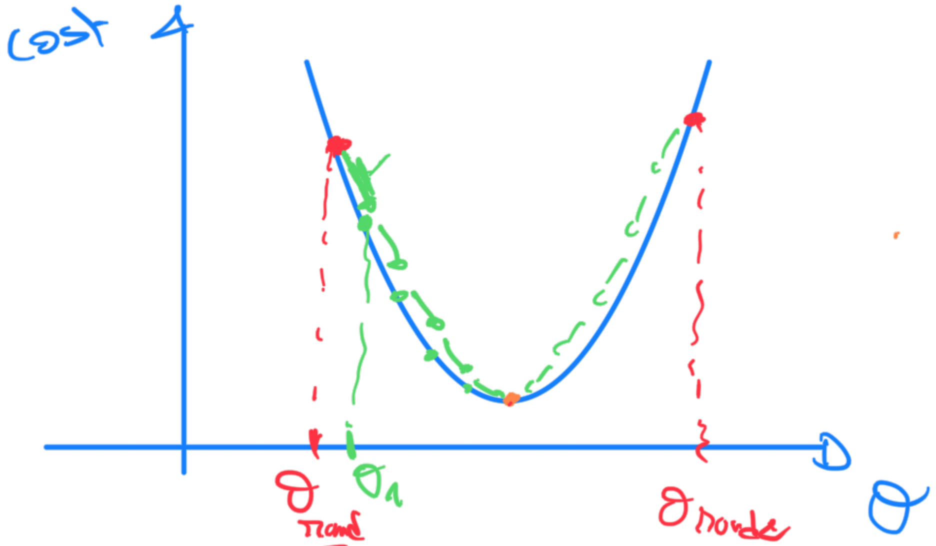
Initialize θ

Repeat until convergence

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial J}{\partial \theta_j}$$

learning rate

$j = 0 \ldots N$



$\alpha > 0$

$$\theta_{j\,new} \leftarrow \theta_{j\,start} - \alpha \frac{\partial J}{\partial \theta_{rand}}$$

< 0

g decreasing

$g' < 0$

$\theta_j - \alpha\,(-\text{num})$

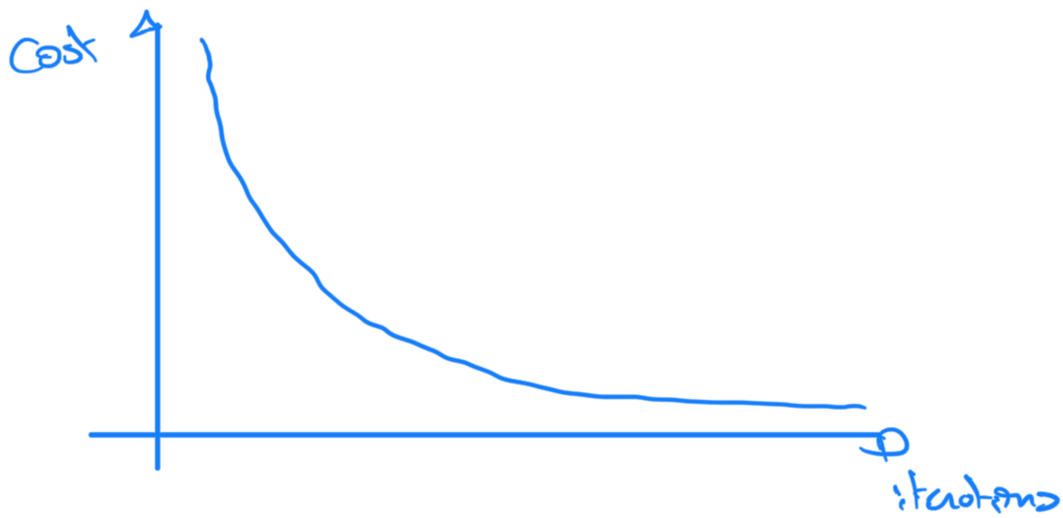
$\theta_j + \alpha\,\text{num}$

g increasing

$g' > 0$

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial J}{\partial \theta_j} > 0$$

$\theta_j - \alpha\,\text{num}$



+ mini batch (stochastic) GD

+ learning rate