linear regression

$$y = XW$$

we want to solve W.

In other words we want to find the best W that minimize loss function = cost function

MSE :

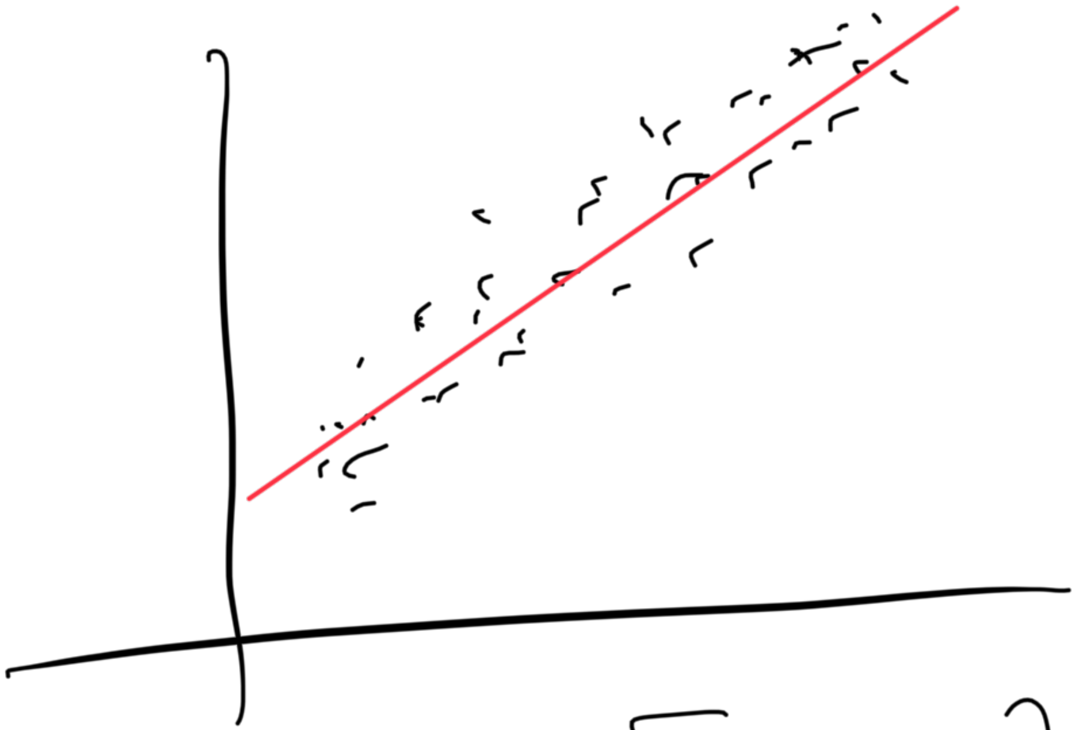
$$J(\underline{w}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \underbrace{Xw}_{\hat{y}_i})^2$$

$$\mathcal{L}(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$$

$$= (y_i - Xw)^2$$

2 cases

1st



$$X = \begin{bmatrix} 1 & x_1^1 \\ \vdots & \vdots \\ 1 & x_N^1 \end{bmatrix}$$

$$W = \begin{bmatrix} b \\ w_1 \end{bmatrix}$$

2nd

$$X = \begin{bmatrix} 1 & x_1 & \cdots & x_1^{m} \\ \vdots & \vdots & & \vdots \\ 1 & x_N & & x_N^m \end{bmatrix}$$

$$W = \begin{bmatrix} b \\ w_1 \end{bmatrix}$$

$$\begin{bmatrix} w_2 \\ \vdots \\ w_m \end{bmatrix}$$

How to find w?

$$\frac{\partial J}{\partial w} = 0$$

$$\Rightarrow w = (X^T X)^{-1} X^T y$$

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

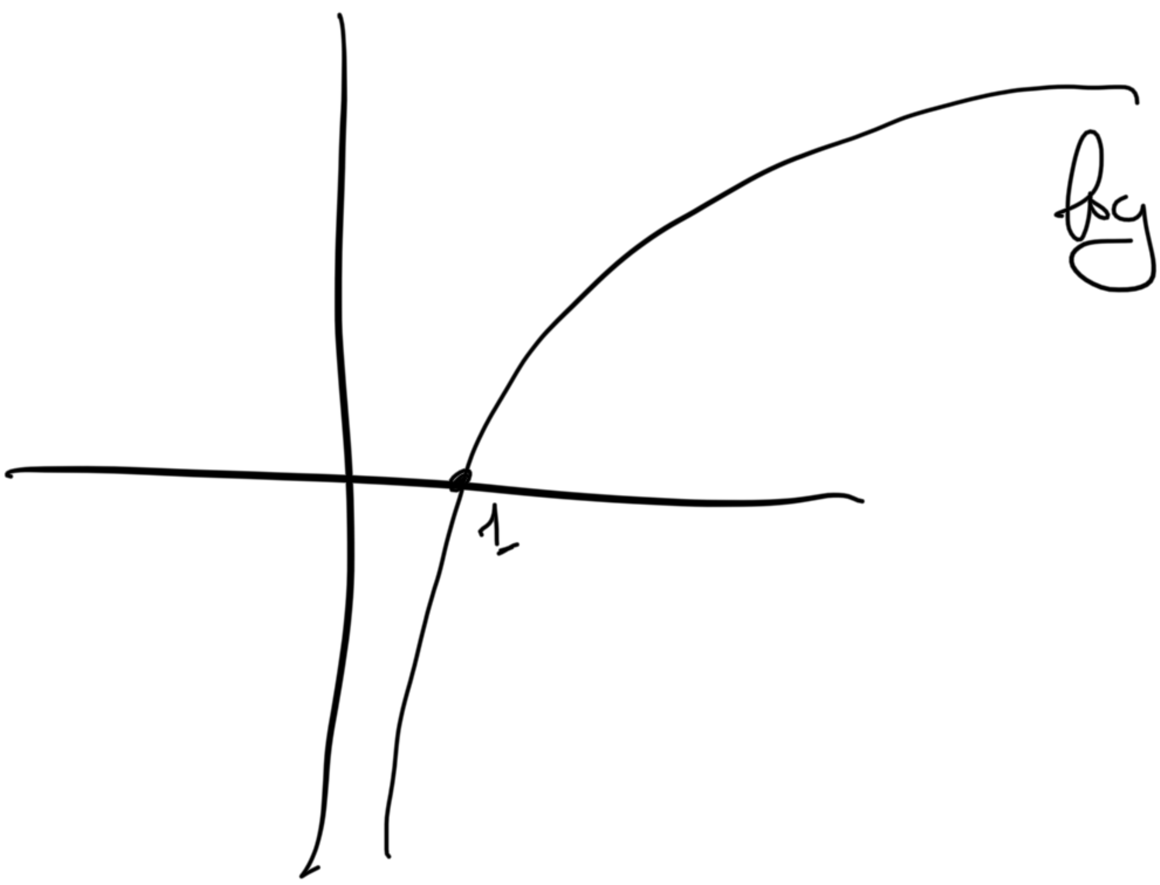
$$+ (1-y_i) \log(1-\hat{y}_i)$$

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

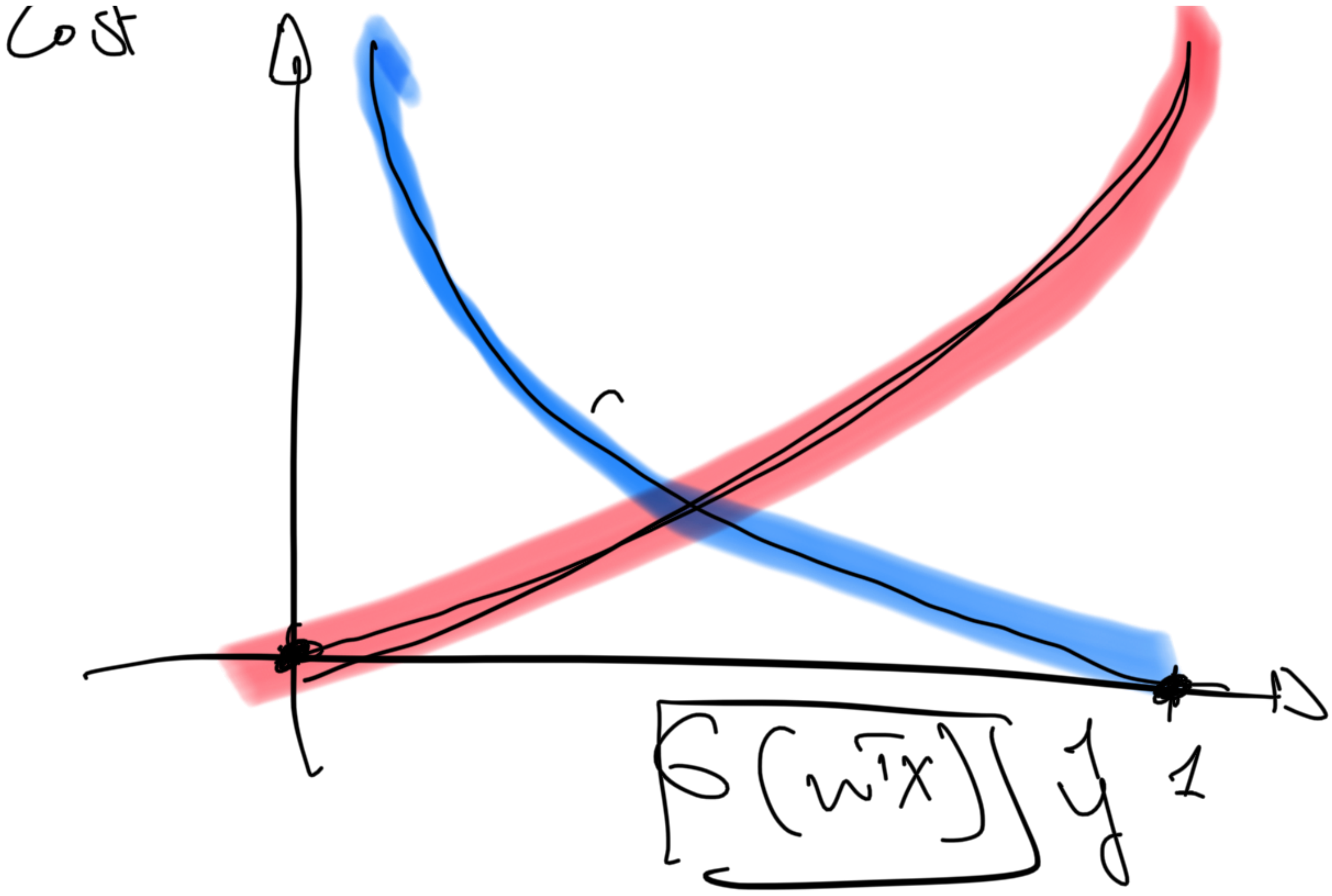
$$\text{Cost}(y, \hat{y}) = -y \log(\hat{y}) - (1-y)\log(1-\hat{y})$$

if $y = 1 \Rightarrow \text{Cost}(y, \hat{y}) = -\log(\hat{y})$

if $y = 0 \Rightarrow \text{Cost}(y, \hat{y}) = -\log(1-\hat{y})$



$$\begin{cases} \boxed{-\log(\sigma(w^T x)) \quad \text{if } y = 1} \\ \boxed{-\log(1 - \sigma(w^T x)) \quad \text{if } y = 0} \end{cases}$$



$$\text{Cost} = 0 \Rightarrow \text{if } \hat{y} = y \Rightarrow \hat{y} = 1$$

$$\sigma(w^T x) = 1$$

$$\sigma(w^T x) \rightarrow 0$$

$$\lim_{\rightarrow 0} \log \rightarrow -\infty$$

$$\text{Cost} \rightarrow +\infty$$

$$\text{Cost} = 0 \quad \text{if } \hat{y} = y \Rightarrow \hat{y} = 0$$

$$\sigma(w^T x) = 0$$

$$\sigma(w^T x) \rightarrow 1$$

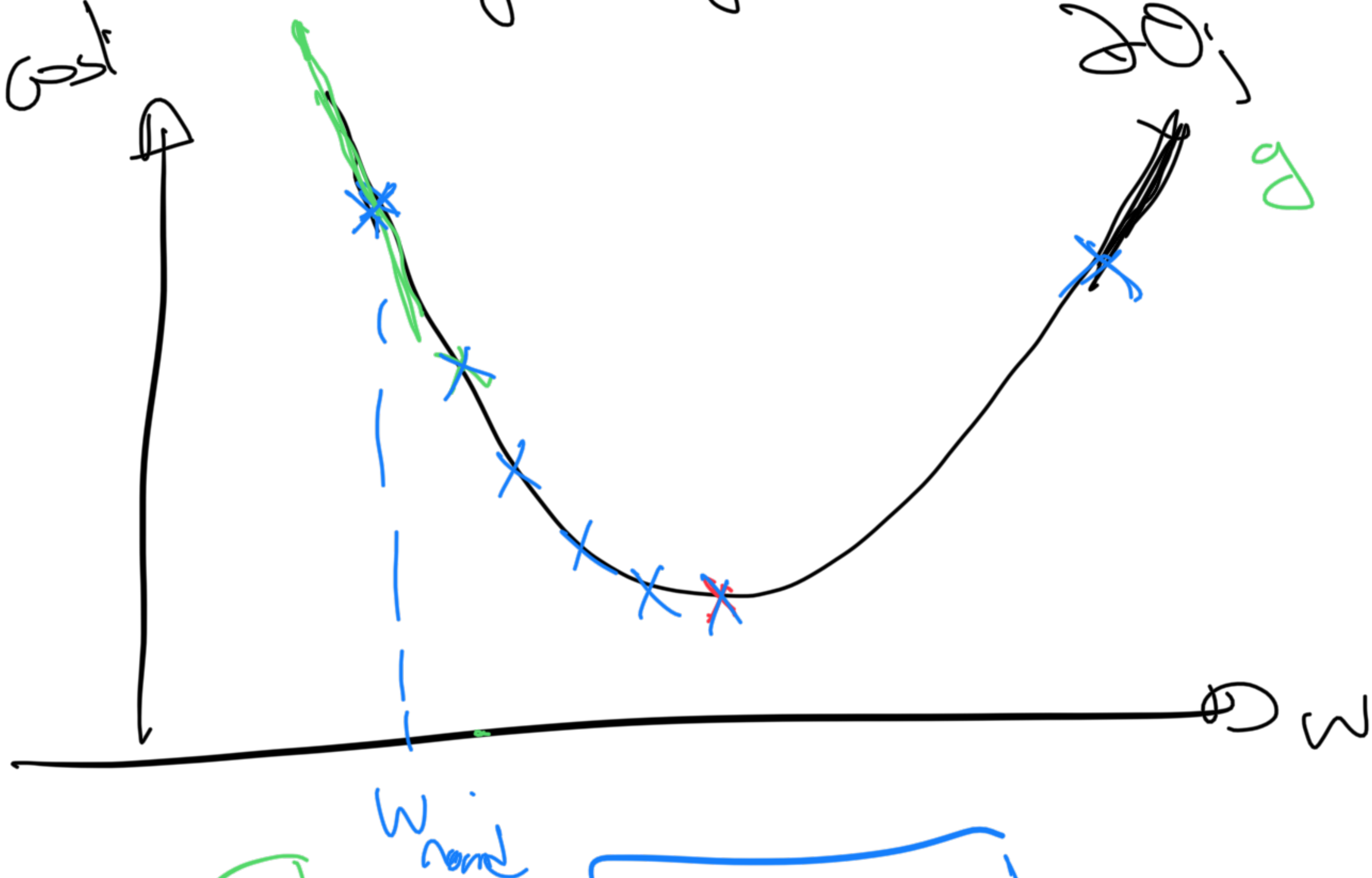
$$1 - \sigma(w^T x) \to 0$$

Gradient Descent:

Initialize θ

Repeat until convergence

$$\theta_j := \theta_j - \alpha \frac{\partial J}{\partial \theta_j}$$



$g' > 0$

$\boxed{g' < 0}$

w_{rand} to get bigger

$$w_{rand} \leftarrow w_{rand} - \alpha g'$$

≥ 0

W round will get bigger