Lecture 11. classification.

1. Classification metrics

① accuracy $= \dfrac{\text{\# of correctly classified samples} \leftarrow (N_c)}{\text{\# of total samples} \leftarrow N}$

② error rate $= 1 - \dfrac{N_c}{N} = \dfrac{N - N_c}{N}$ ← misclassified.

③ confusion matrix (cfm)

suppose we have 3 classes: $\underbrace{0,\ 1,\ 2}_{\text{class labels}}$

True ← target labels	predicted class 0	predicted class 1	predicted class 2
0	5	1	0
1	1	4	1
2	2	0	4

cfm: count the # of data samples of a predicted class at a specific target class

		predicted		
		0	1	2
target	0	5	1	0
	1	1	4	1
	2	2	0	4

→ # of samples with target label '0'

→ # ...... label '1'

→ # ...... label '2'

Total data samples: $N = (5+1) + (1+4+1) + (2+0+4)$

$= 18$

$$\text{accuracy} = \frac{5+4+4}{18} = \frac{13}{18}$$

$$\text{error rate} = 1 - \frac{13}{18} = \frac{5}{18}$$

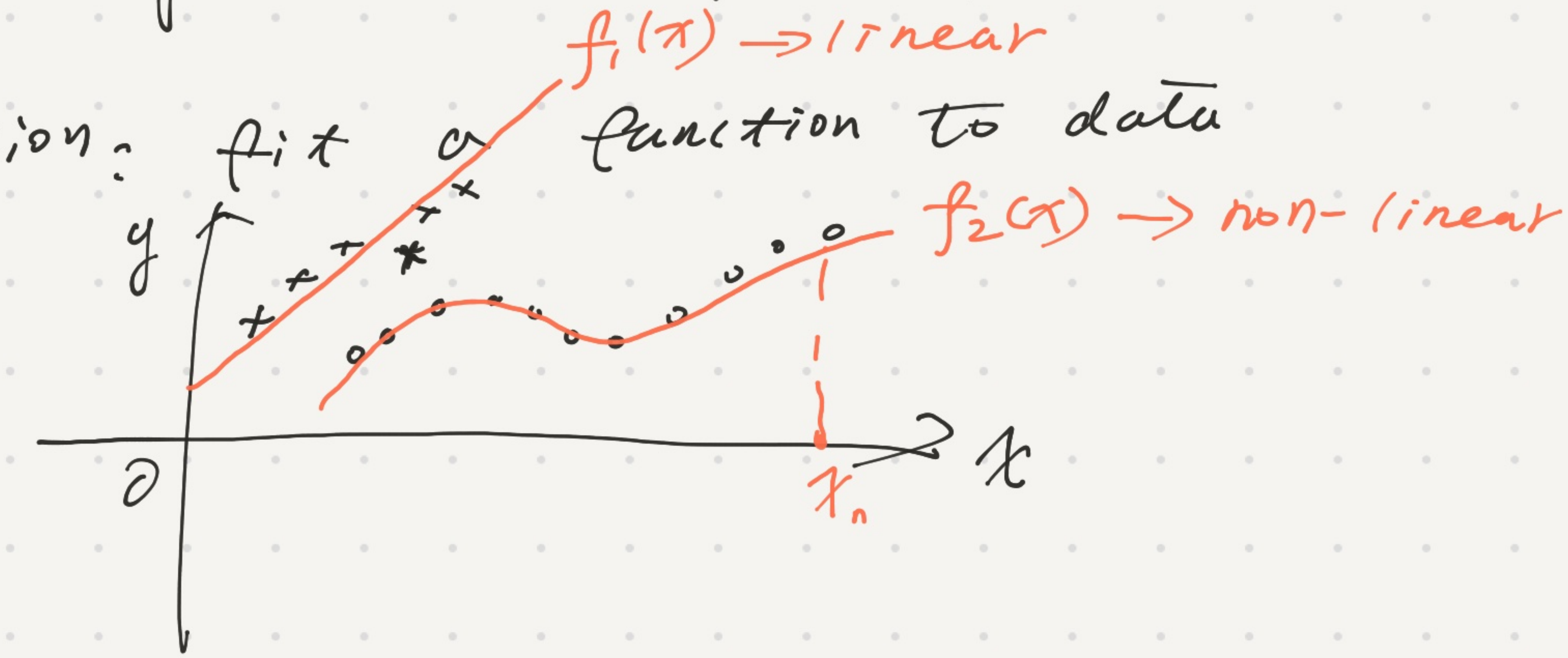
recall rate for class '0' : $\dfrac{\text{\# of correctly classified samples of '0'}}{\text{\# of samples of '0'}}$

5/6

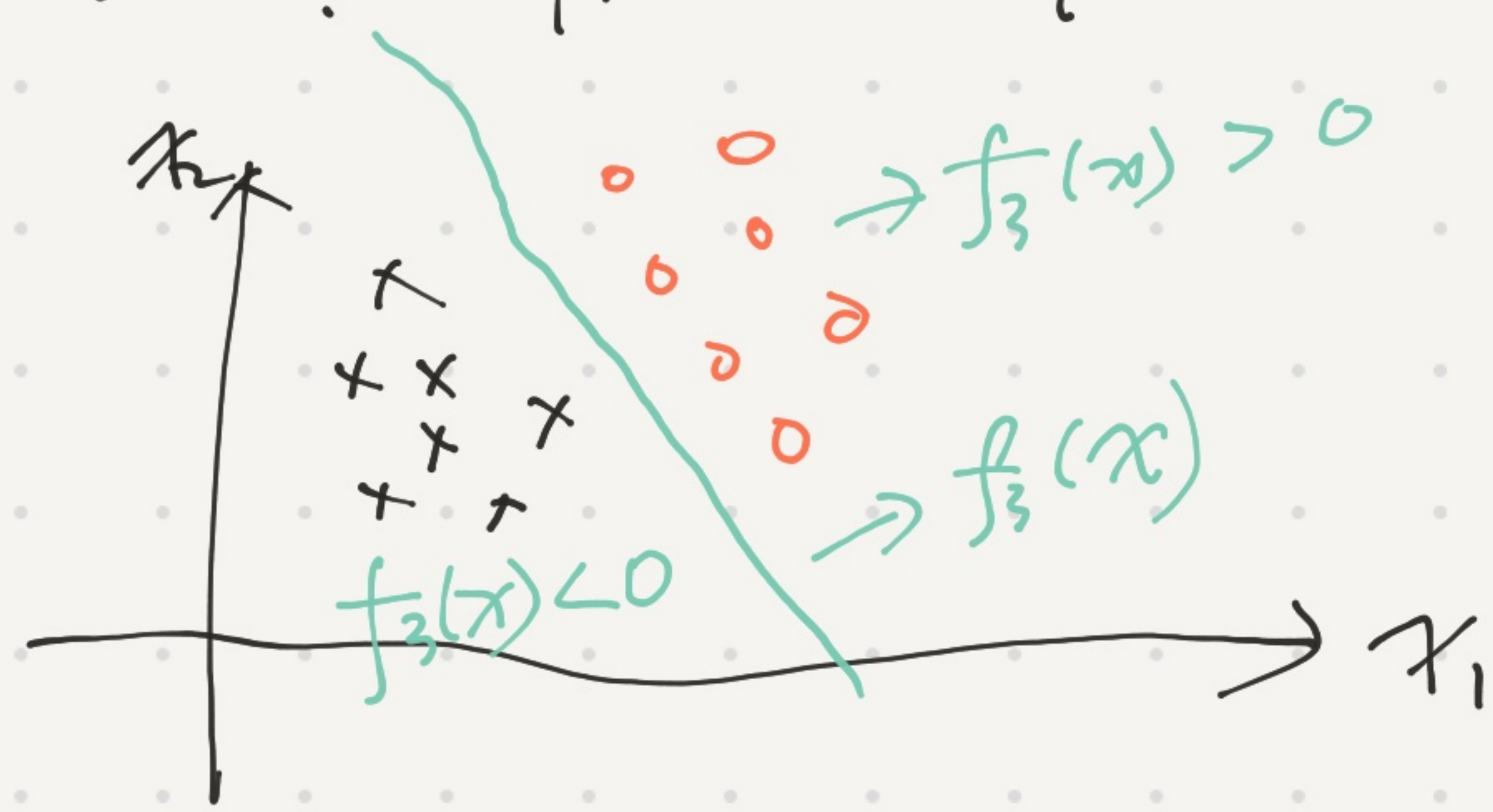
recall rate for class '2' : $4/6 = 2/3$

2. Logistic regression (classification)

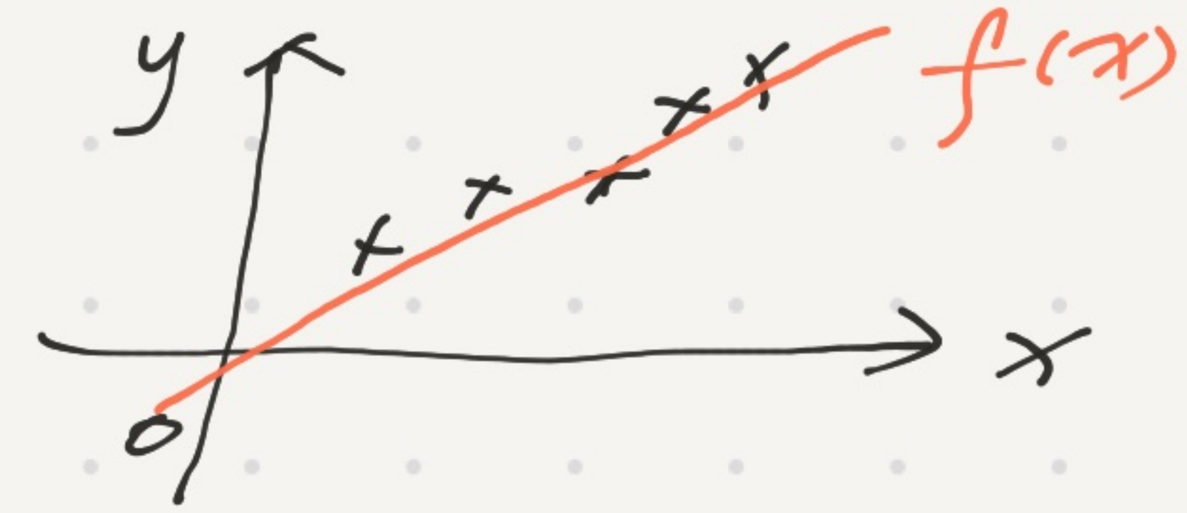
Regression: fit a function to data



classification: fit a function to seperate data

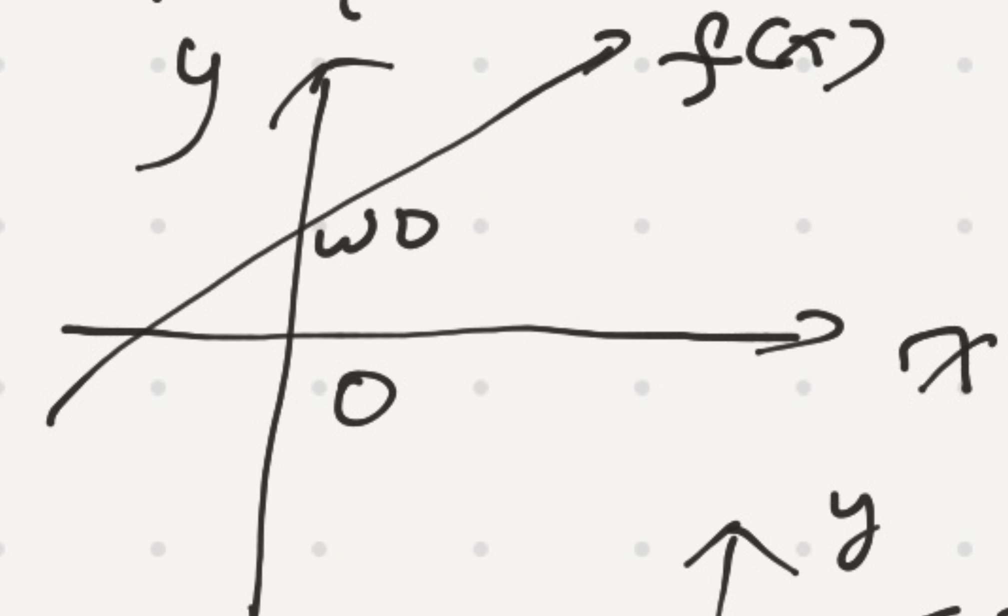


2.1 Linear regression



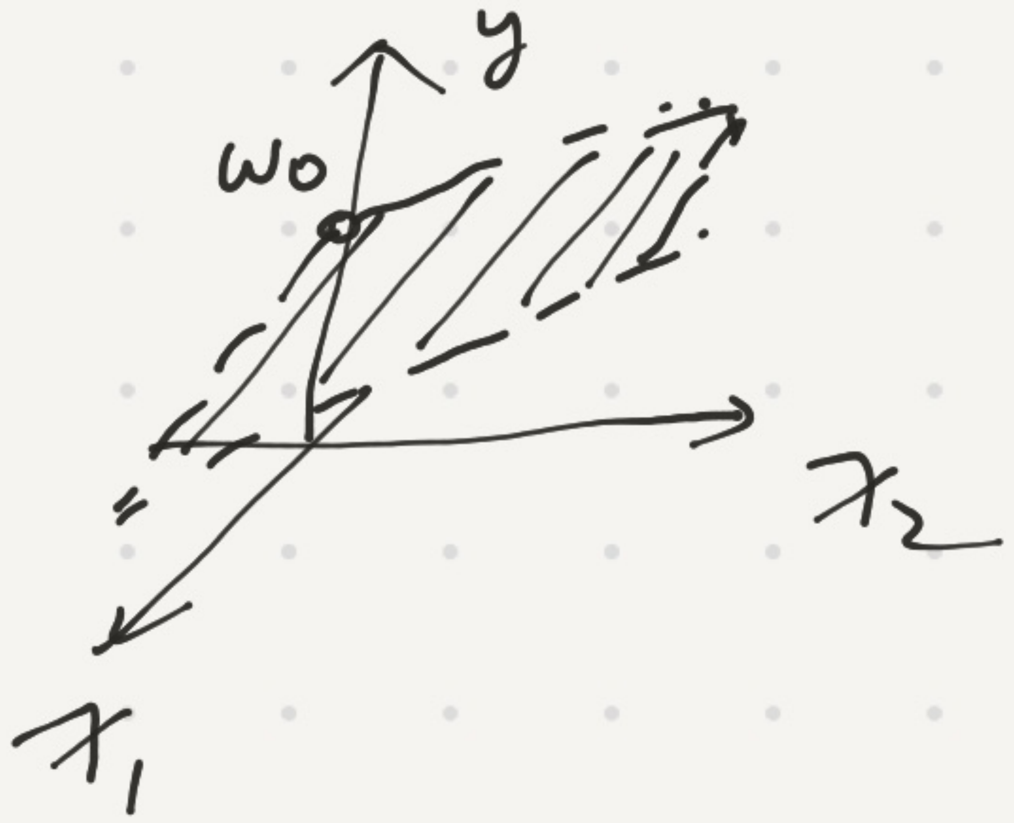
$$\underset{\text{input}}{x} \longrightarrow \boxed{\overset{\text{model}}{f(x)}} \longrightarrow \hat{y} \rightarrow \text{estimation of the target}$$

Model: suppose we are looking for linear function



input x: 1d: $f(x) = \underbrace{w_0 + w_1 x}_{\text{line}}$

$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ 2d: $\underbrace{f(x) = w_0 + w_1 x_1 + w_2 x_2}_{\text{plane}}$



$\vdots$

$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ nd: $f(x) = \overset{x_0=1}{w_0} + \underbrace{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}_{\text{hyperplane.}}$

$$\underline{f(x) = \mathbf{w}^T \cdot \mathbf{x}} \qquad \mathbf{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{pmatrix}_{(n+1)\times 1} \qquad \mathbf{x} = \begin{pmatrix} x_0 = 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Model training (fit): is to find $\overbrace{\text{out a vector } w}^{w^*}$ that can best fit the data

$$f(x) = w^T \cdot x = (w^*)^T \cdot x \rightarrow \hat{y}$$

$$x \longrightarrow \boxed{f(x, w)} \rightarrow \hat{y} \longrightarrow y$$

Loss function

cost function

objective function

$$L = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \quad \text{(mean square error)}$$

$$= \frac{1}{N} \sum_{i=1}^{N} (w^T \cdot x_i - y_i)^2$$

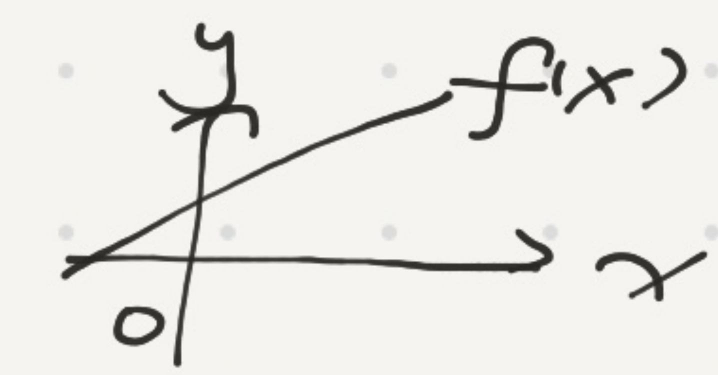
Goal is to:

$$w^* = \underset{w}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} (w^T \cdot x_i - y_i)^2$$

Optimization Algorithms, optimizers.

Gradient decent → GD.

2.2 logistic regression



linear function

$$x \to \boxed{f(x)} \to \hat{y}$$

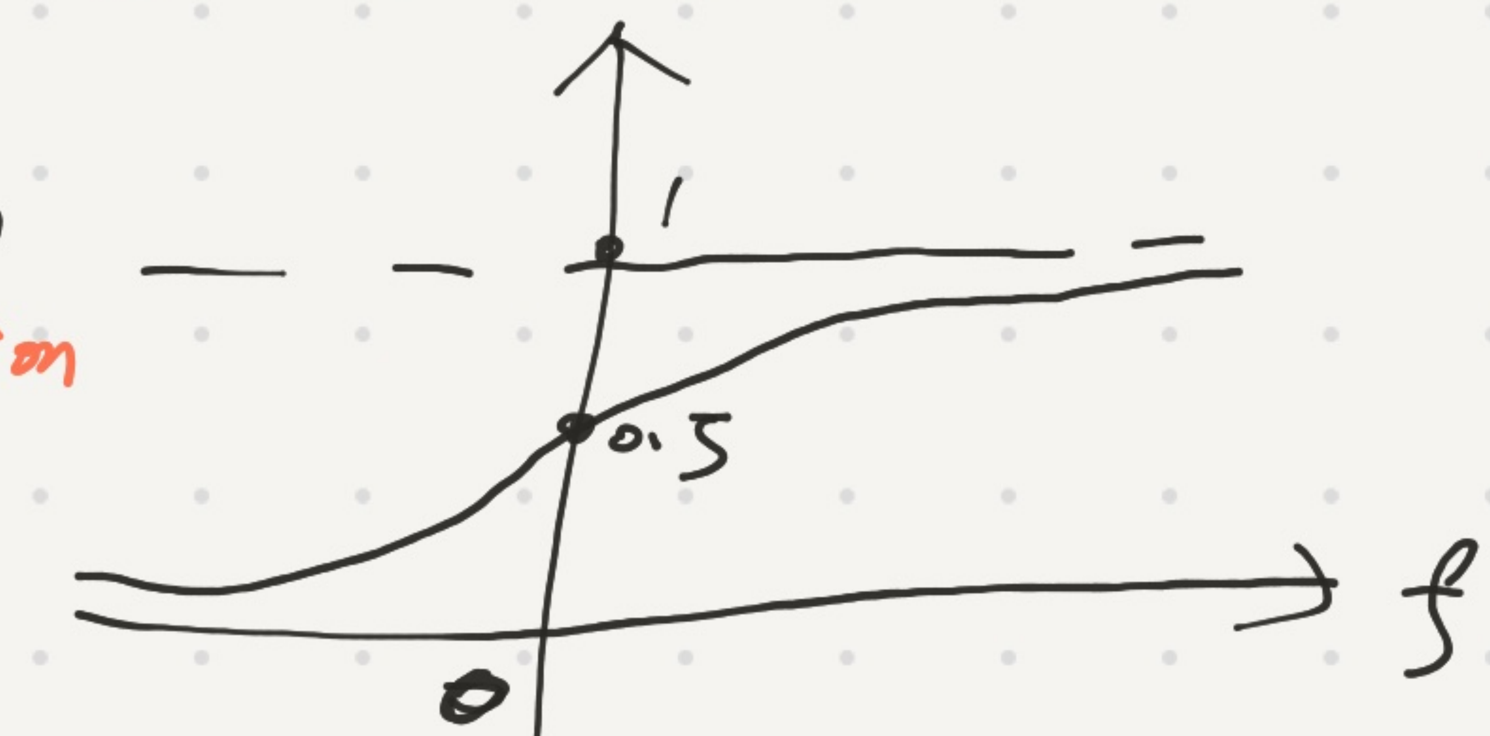
$\hat{y} \to$ continuous value

thresholding

$$\boxed{x \to \boxed{f(x)} \xrightarrow{\hat{y}}} \boxed{\sigma} \xrightarrow{[0,1]} \boxed{T} \to 0, \text{or } 1$$

Logistic function

sigmoid function



$$\sigma(f) = \frac{1}{1+e^{-f}} \in [0,1]$$

$$f \to +\infty \to e^{-f} \to 0 \longrightarrow \sigma(f) = 1$$

$$f \to -\infty \to e^{-f} \to +\infty \to \sigma(f) = 0$$

$$T(p) = \begin{cases} 1, & \text{if } p \geq \underline{0.5} \\ 0, & \text{if } p < \underline{0.5} \end{cases}$$

$p \in [0,1]$