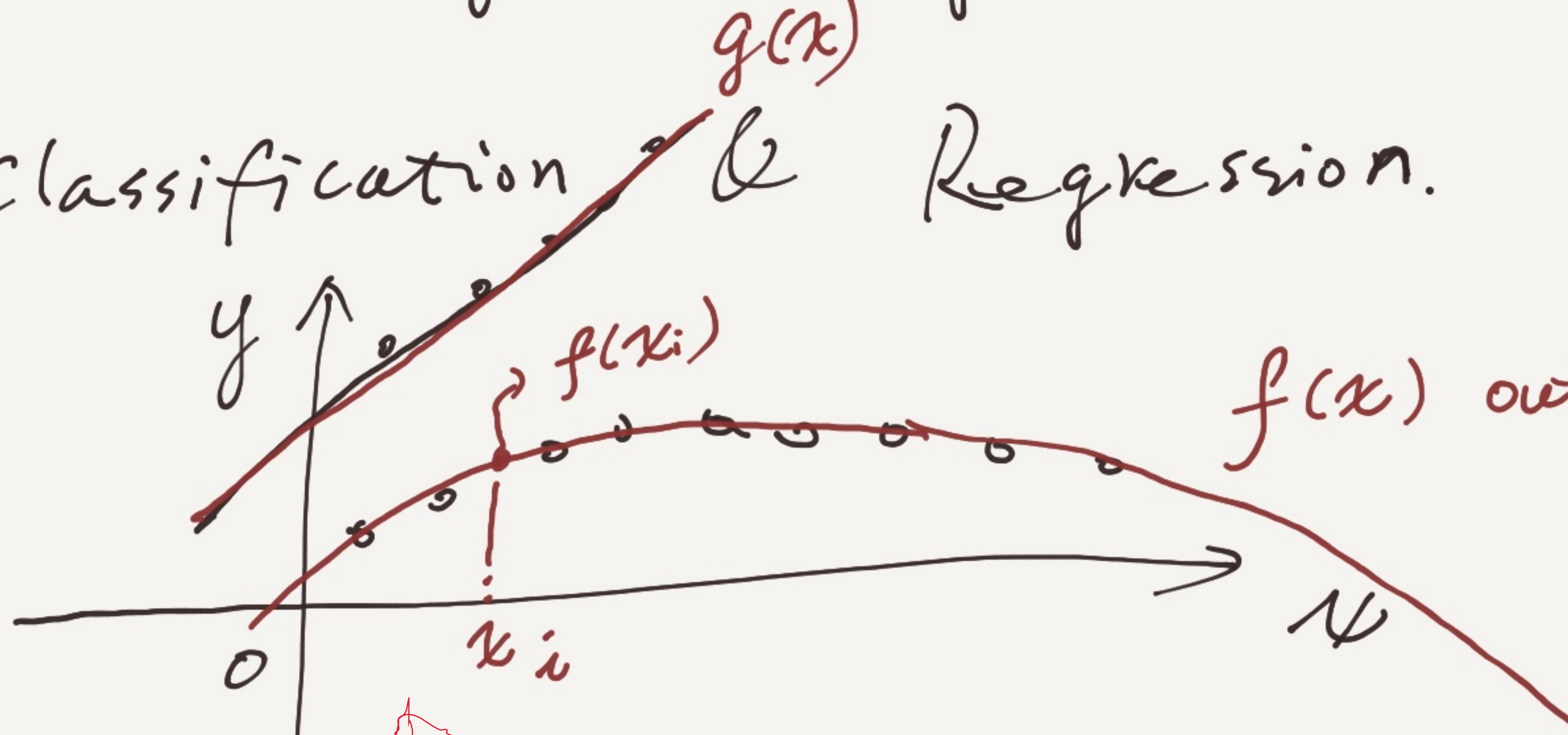
Lecture 11. Logistic Regression

1. Classification & Regression



Regression examples.

$f(x)$ outputs continuous values.

e.g., $[0, 1]$

classification.

$h(x)$ outputs discrete values.

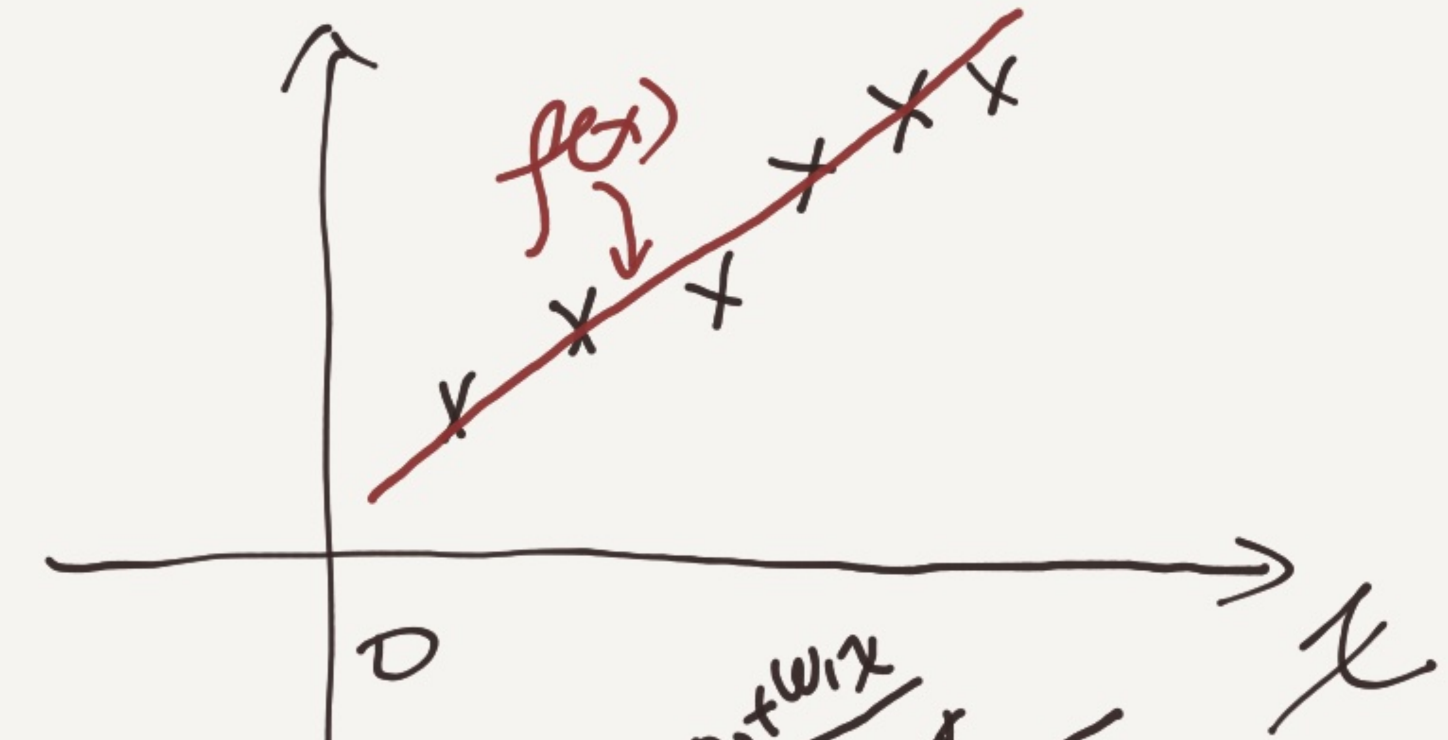
e.g., $\{0, 1\}$

, $\{-1, 1\}$

2. Linear Regression

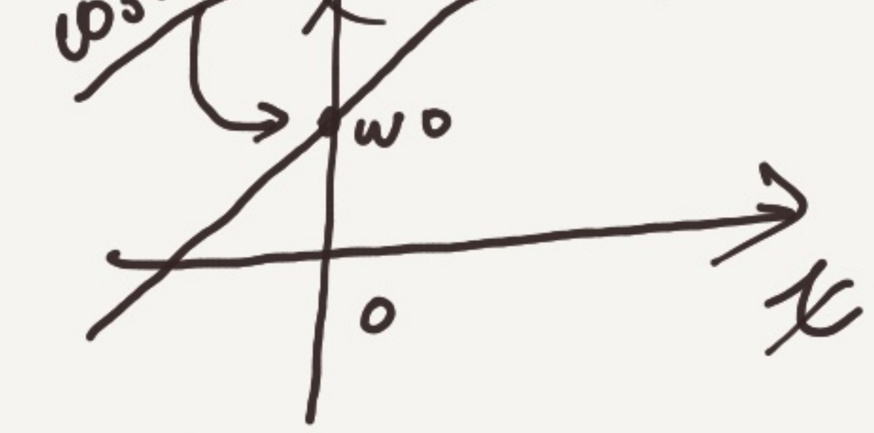
model

input x → linear function $f(x)$ → $\tilde{y}$ (prediction)

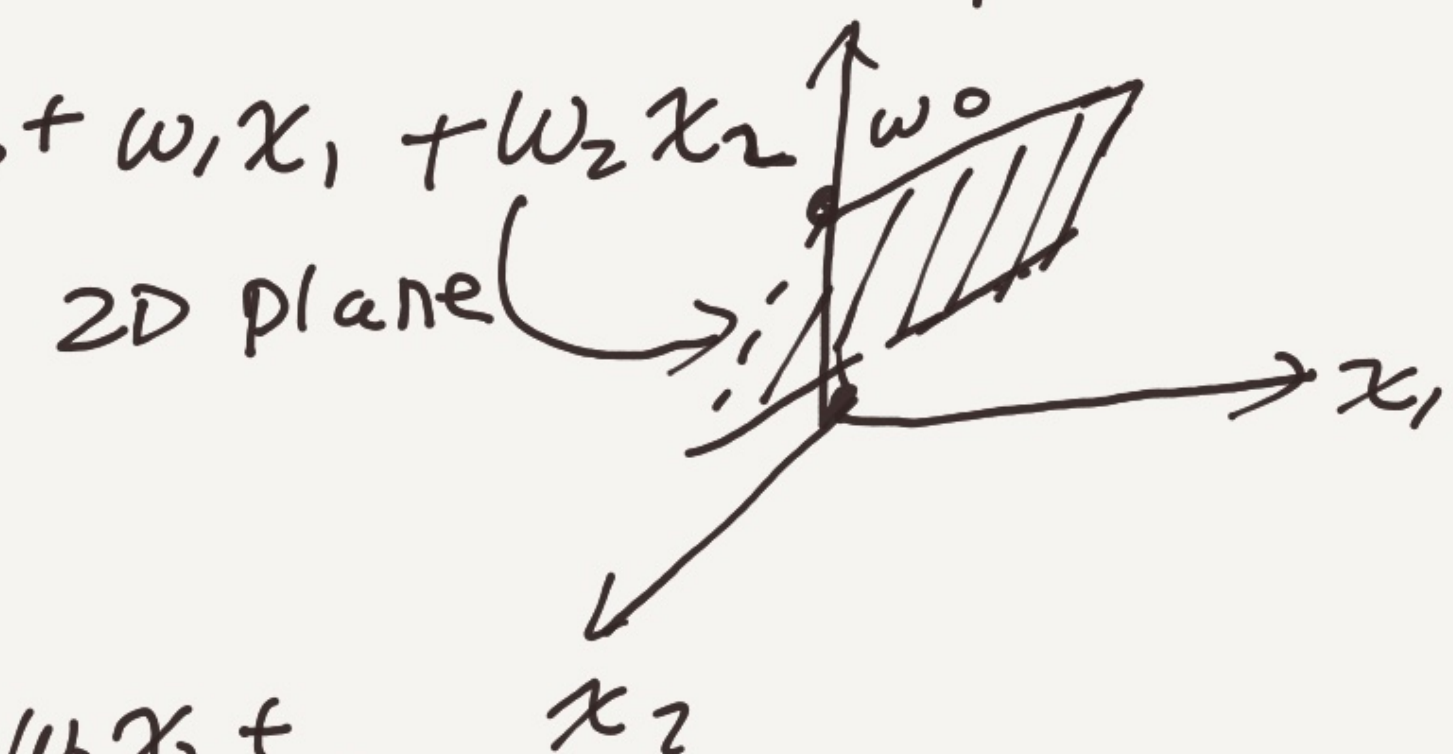


① Input:

x : 1d (one feature) → ② Model: $f(x) = w_0 + w_1 x$



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



⋮

nd : $x = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix}$ → $\underbrace{f(x) = w_0 + w_1 x_1 + w_2 x_2 + \dots + w_n x_n}_{\text{hyperplane.}}$

All above models are linear models.

Simplify the model definition of linear model:

$$f(x) = \underbrace{w}_{\text{model parameter(s)}}{}^T \cdot x \qquad w = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{pmatrix} \qquad x = \begin{pmatrix} x_0 = 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}$$

data

3 **Model training**: Key is to find out the vector w that can best fit the data:

w^*: best parameter vector.

$=$

$$x \longrightarrow \boxed{(w^*)^T x} \longrightarrow \hat{y} \longleftrightarrow y$$

prediction

4 **Loss/objective / cost function**: defines the difference between predictions ($\hat{y}$) and true targets (y)

$$L_1 = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad \text{(mean square error)}$$

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

$\hat{y}_i$ is the predictoin for the ith data sample,

y_i is true target.

N: # of data samples

Formulate the linear regression problem as:

$$w^* = \underset{w}{\arg\min}\ \frac{1}{N}\sum_{i=1}^{N} (\hat{y}_i - y_i)^2 = \underset{w}{\arg\min}\ \frac{1}{N}\sum_{i=1}^{N} (w^T x - y_i)^2$$

The algorithms to solve the above equation are called

optimization algorithms / optimizer(s)

3. Logistic Regression.

Linear Regression: $x \longrightarrow \boxed{\overset{\text{linear}}{f(x)}} \longrightarrow \hat{y}$ continuous value

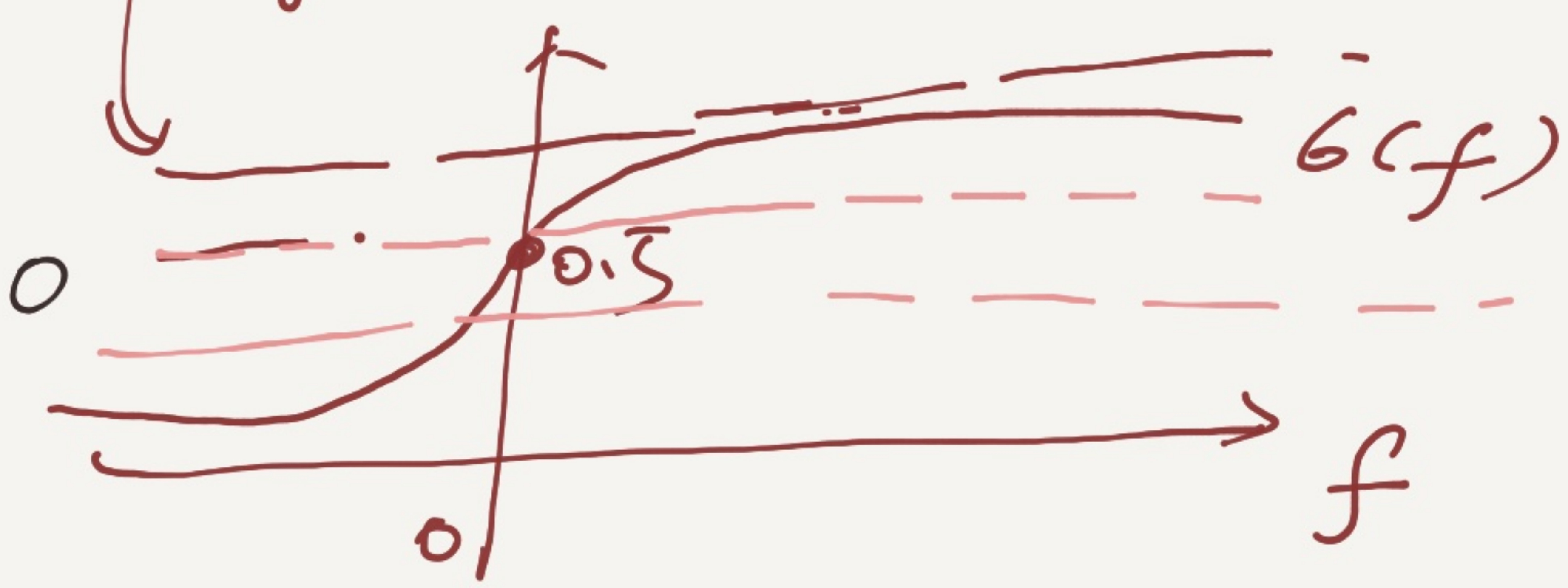
Classification: Logistic Regression: $x \longrightarrow \boxed{f(x)} \longrightarrow \hat{y} \longrightarrow \boxed{\sigma} \xrightarrow{p \in [0,1]} \boxed{T} \longrightarrow 0 \text{ or } 1$

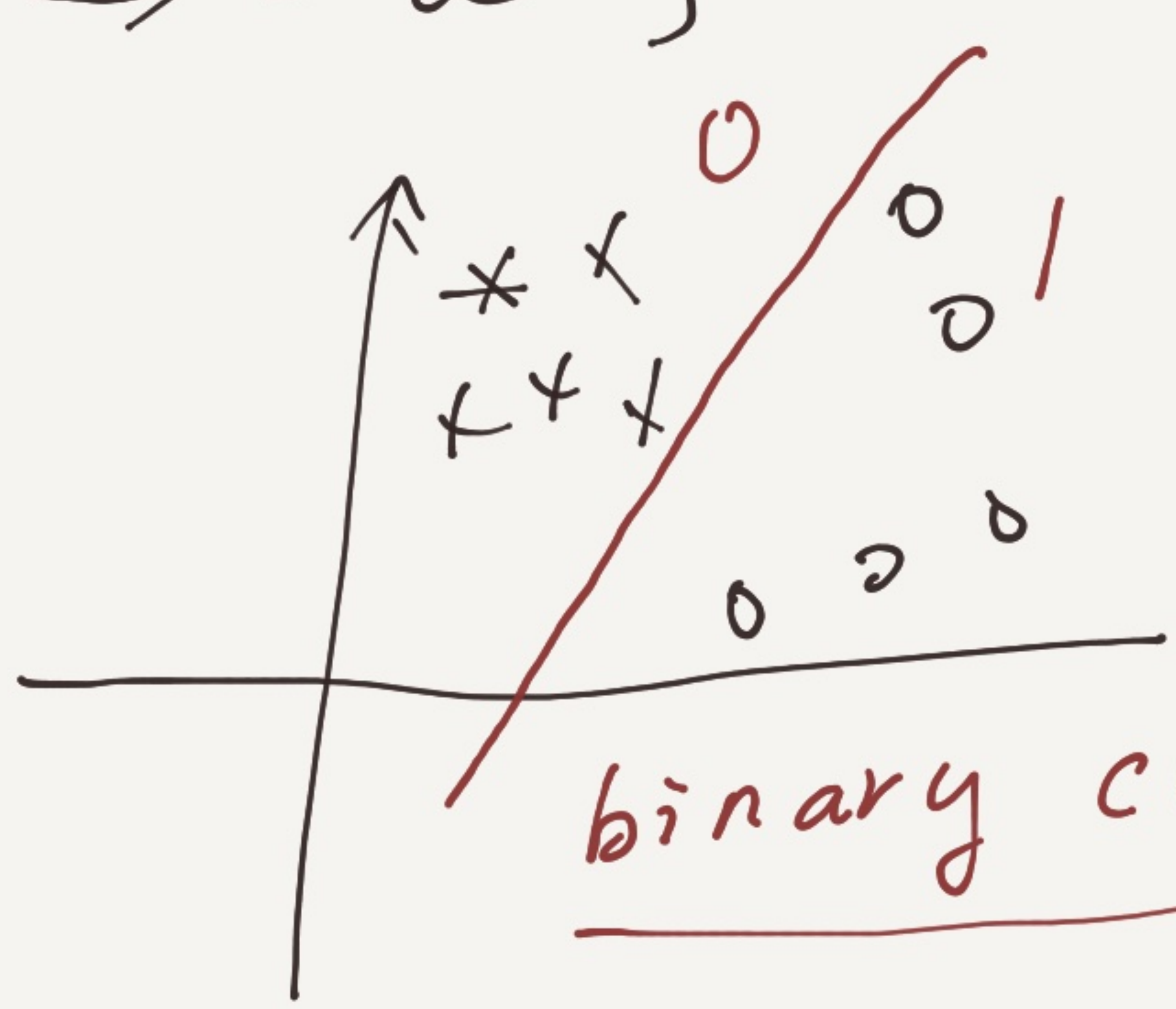
$$\sigma(f) = \frac{1}{1+e^{-f}} \in [0,1]$$

Logistic function

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

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





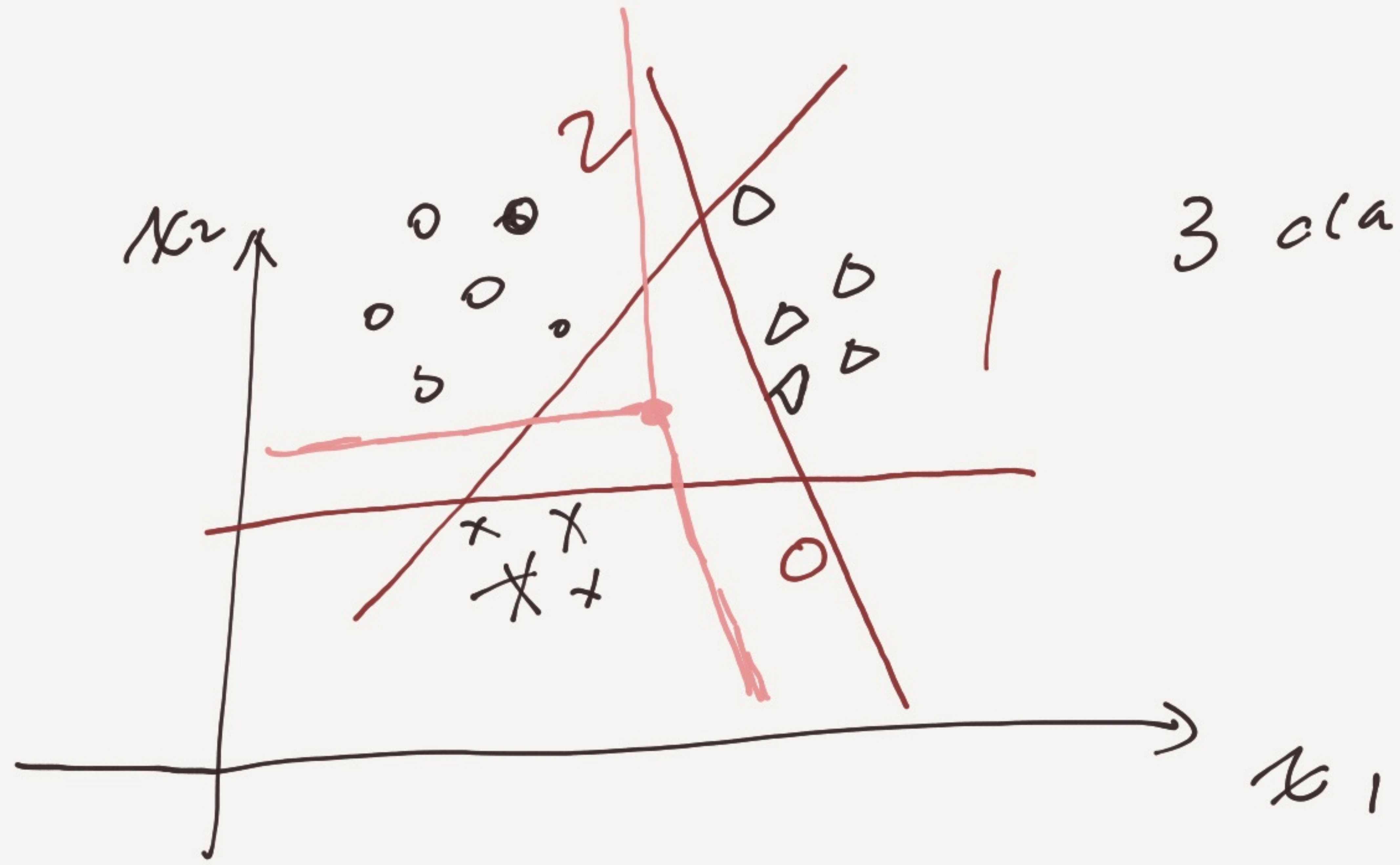
binary classifier

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

Thresholding

$p \in [0,1]$

Multiple-class classification using binary classifiers



3 classes.

$$T(p) = \begin{cases} 0 & p < 0.3 \\ 1 & 0.3 \le p \le 0.67 \\ 2 & 0.67 \le p \le 1 \end{cases}$$

(1) train multiple binary classifiers, and each classifies one from the rest

(2)multiple binary classifiers, and each classifies every two classes.