Goals:
① understand the math behind logistic regression
② develop intuitions about learning NN

ICCS482 Deep Learning

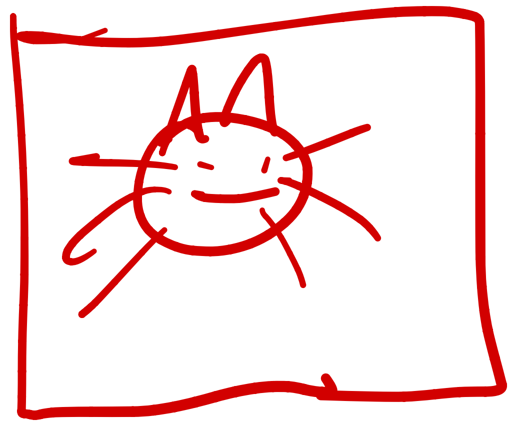
Lecture 3: Logistic Regression

Sunsern Cheamanunkul, Sep 10, 2020.

Binary Classification

1 (cat) or 0 (non cat)

Logistic Regression

 $\rightarrow$ $[x_1 \; x_2 \; x_3 \; \dots \; x_n] \longrightarrow y \in \{0, 1\}$

feature vector

$$X = \left\{ \begin{array}{l} (\vec{x}^{(1)}, y^{(1)}) \\ (\vec{x}^{(2)}, y^{(2)}) \\ \vdots \\ (\vec{x}^{(m)}, y^{(m)}) \end{array} \right.$$

$\leftarrow$ an example

$\vec{x} \in \mathbb{R}^{n_x}$

n_x is th # of features

Logistic Regression

$\tilde{x}^{(new)} \rightarrow \boxed{\text{model}} \rightarrow 0/1$

control the shift of the sigmoid

$\tilde{x}^{(new)} \rightarrow \boxed{z = \tilde{w}\cdot\tilde{x} + b} \rightarrow \boxed{\text{some number}}$

$\mathbb{R}$

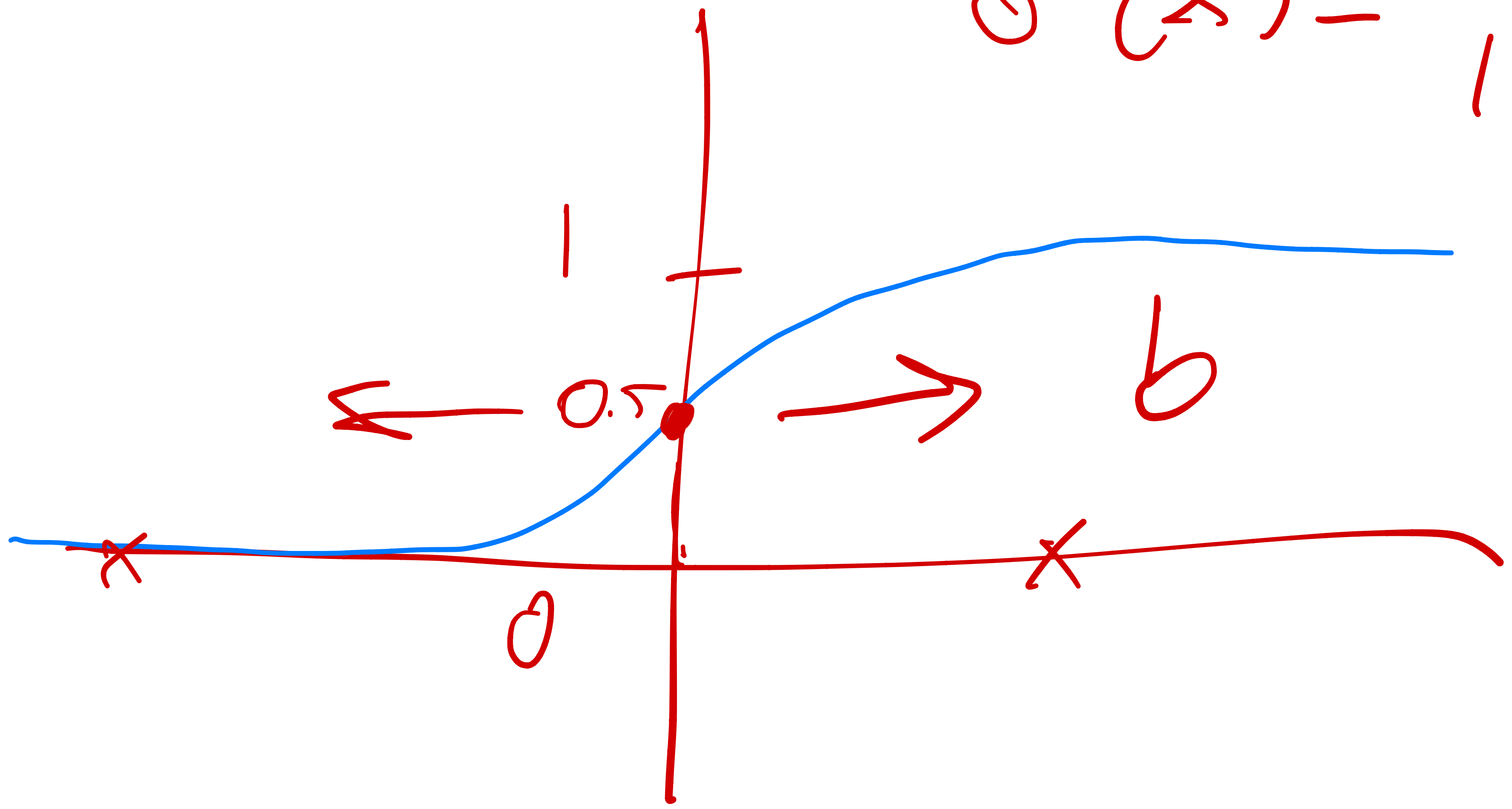
$\sigma(x) = \frac{1}{1+e^{-x}}$

$\tilde{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$\tilde{x} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

$\tilde{w}\cdot\tilde{x} = 1\cdot 3 + 2\cdot 4$

$= 11$



Logistic Regression

$\hat{y} = \sigma(\vec{w}\cdot\vec{x}+b)$

How to train the model?

→ Figure out the optimal settings for $\vec{w}$ and b

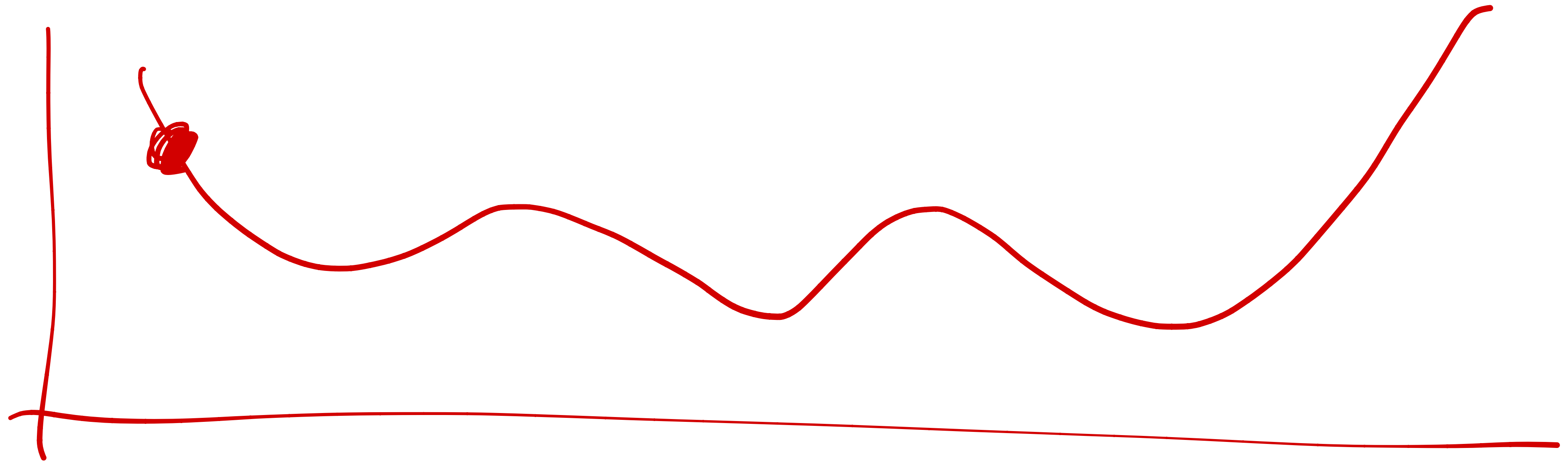
objective function

Loss function: $\ell(y,\hat{y}) = \cancel{\frac{1}{2}(\hat{y}-y)^2}$

this is non-convex

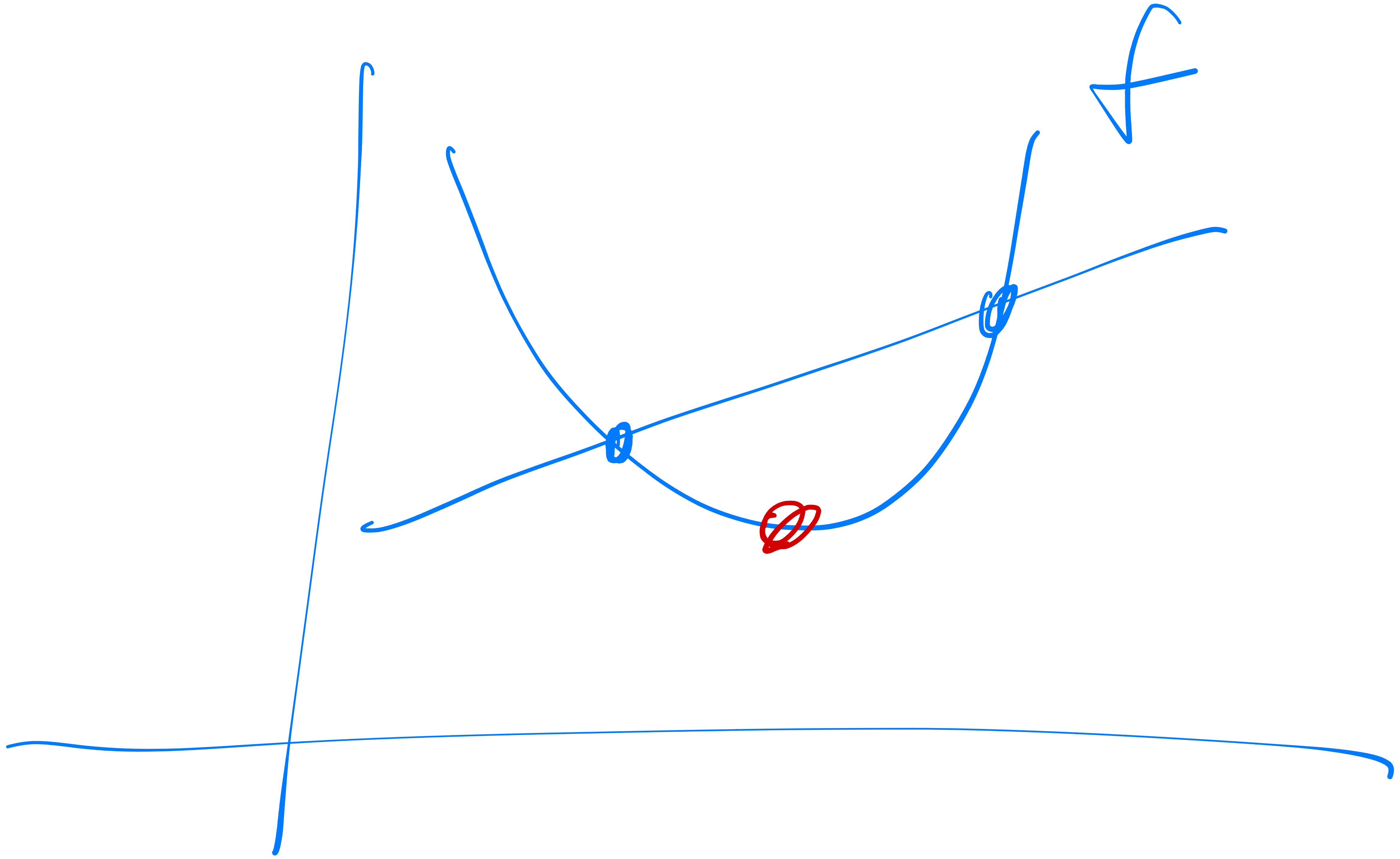
square loss

cost

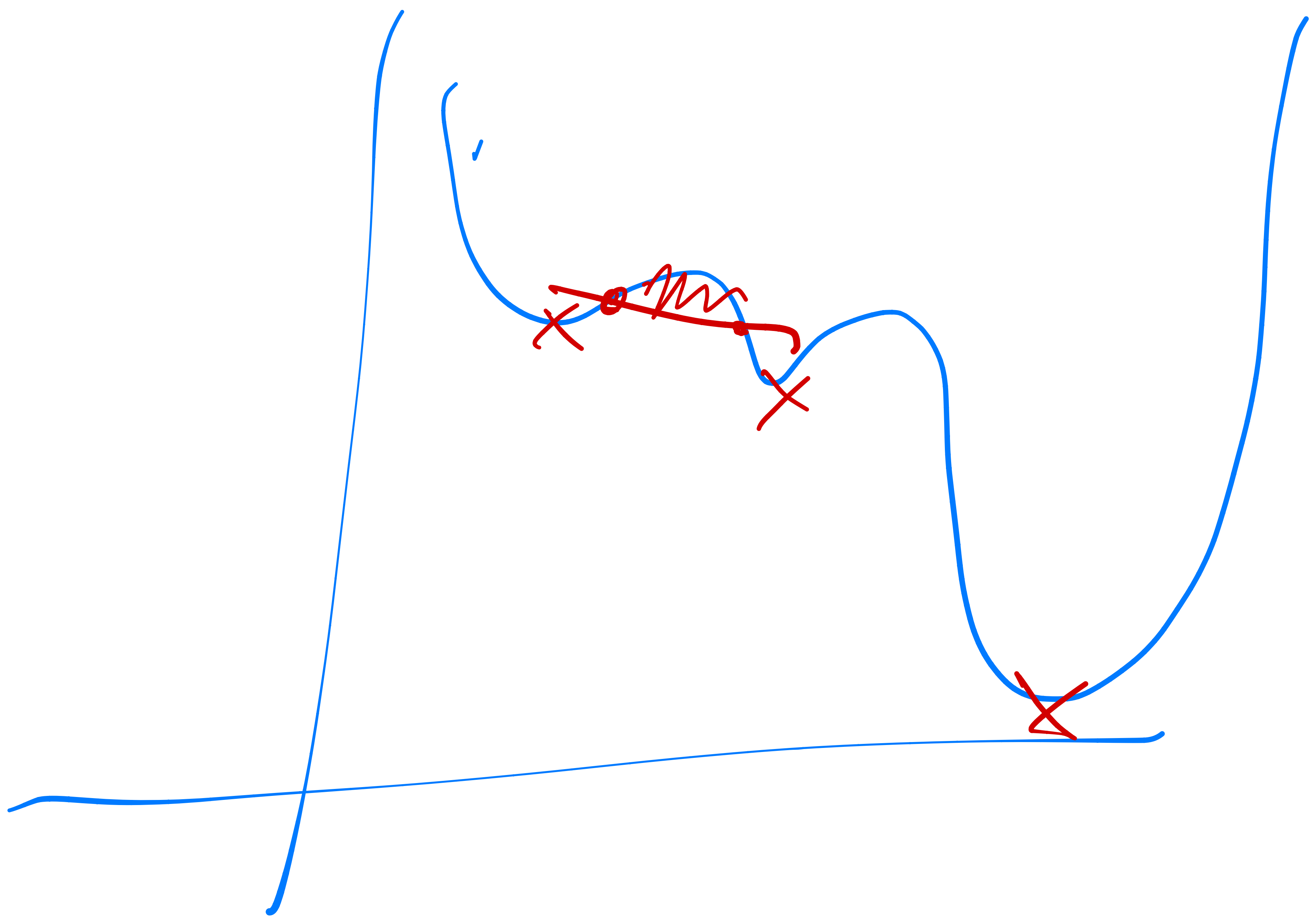


Logistic Regression

convex



non-convex



Logistic Regression

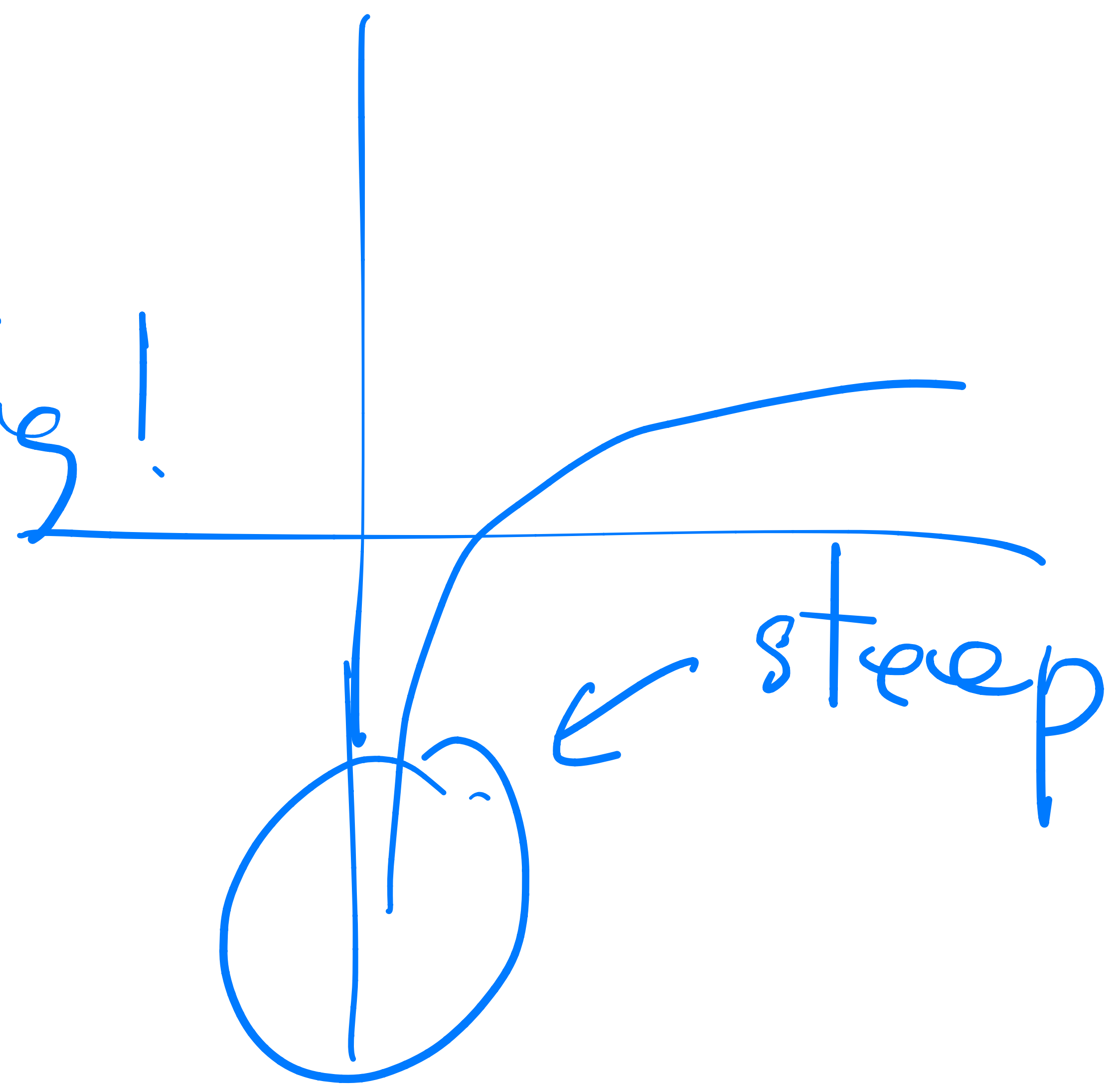
works well in classification

Log loss (entropy loss)

$$\ell(\hat{y}, y) = -(y \cdot \log \hat{y} + (1-y)\log(1-\hat{y}))$$

if $y=1$, $\ell(\hat{y}, y) = -\log \hat{y}$

very big!



if $y=0$, . . .

Logistic Regression

Cost Function (objective function)

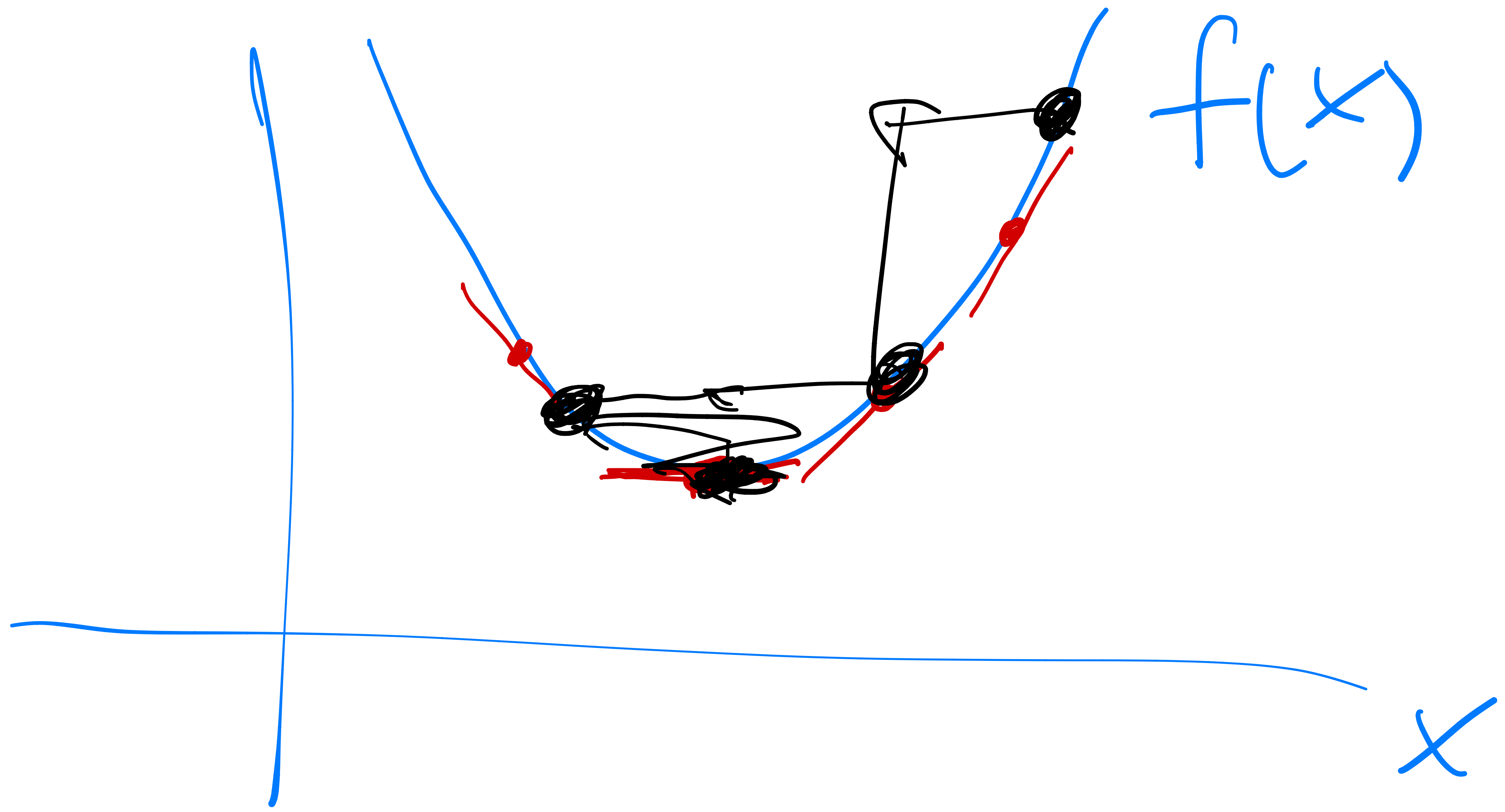
$$J(\vec{w}, b) = \underbrace{\frac{1}{m} \sum_{i=1}^{m} \ell(\hat{y}^{(i)}, y^{(i)})}_{\text{average loss over the training set.}}$$

size of the train set

Logistic Regression

How to find the best $\vec{w}$ and b?

"Gradient Descent"



minimum points must have $f'(x) = 0$

Logistic Regression

$$J(\vec{w}, b) = \frac{1}{m}\sum_{i=1}^{m} \ell(\hat{y}^{(i)}, y^{(i)}) = \frac{-1}{m}\sum_{i=1}^{m} y^{(i)} \cdot \log(\hat{y}^{(i)}) + (1-y^{(i)}) \cdot \log(1-\hat{y}^{(i)})$$

$$\vec{w}_0 \longrightarrow \vec{w}_1 \longrightarrow \vec{w}_2 \quad \ldots$$

$$b_0 \longrightarrow b_0 \longrightarrow b_2$$

$$\boxed{\begin{aligned} w &= w - \alpha \cdot \frac{dJ(\vec{w}, b)}{dw} \\ b &= b - \alpha \frac{dJ(\vec{w}, b)}{db} \end{aligned}}$$

Logistic Regression

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \quad b$$

$$z = \vec{w} \cdot \vec{x} + b = (x_1 \cdot w_1 + x_2 \cdot w_2) + b$$

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

$$\ell(a, y) = -\left(y \cdot \log a + (1-y) \log(1-a)\right)$$

$$J(\vec{w}, b) = \frac{1}{m} \sum_{i=1}^{n} \ell(a^{(i)}, y^{(i)})$$

Logistic Regression

$$\frac{dJ(\vec{w},b)}{dw_1} = \frac{1}{m}\sum_{i=1}^{n} \frac{d\,\ell(a^{(i)}, y^{(i)})}{dw_1}$$

$$\frac{d\,\ell(a,y)}{dw_1} = -\left(y \cdot \log a + (1-y)\log(1-a)\right)$$

$$= -y\frac{d\log a}{dw_1} + \frac{d\log(1-a)}{dw_1} - y\frac{d\log(1-a)}{dw_1}$$

$$\frac{d\log a}{dw_1} = -\frac{y}{a}\frac{da}{dw_1} + \frac{1}{1-a}\left(\frac{d(1-a)}{dw_1}\right) - \frac{y}{1-a}\frac{d(1-a)}{dw_1}$$

Logistic Regression

$$\frac{d\, l(a,y)}{da} = \frac{-y}{a} + \frac{1-y}{1-a}$$

$$\frac{d\, l(a,y)}{dz} = \frac{d\, l(a,y)}{da} \cdot \frac{da}{dz}$$

$$= \left(\frac{-y}{a} + \frac{(1-y)}{1-a}\right) \cdot \frac{da}{dz}$$

$$= \left(-\frac{y}{a} + \frac{(1-y)}{1-a}\right) \cdot \left(a(1-a)\right) = a-y$$

$$a = \sigma(z)$$

$$\frac{da}{dz} = \frac{d\, \frac{1}{1+e^{-z}}}{dz}$$

$$= a(1-a)$$

Logistic Regression

$$\frac{d\,l(a,y)}{dw_1} = \boxed{x_1 \cdot dz}$$

$$\frac{d\,l(a,y)}{dw_2} = \boxed{x_2 \cdot dz}$$

$$\frac{d\,l(a,y)}{db} = \boxed{dz}$$

$$dz = \left(\frac{-y}{a} + \frac{1-y}{1-a}\right) \cdot a(1-a)$$

$$= a - y$$

Logistic Regression

$J=0$; $dw_1=0$; $dw_2=0$; $db=0$

For $i=1$ to m

$z^{(i)} = \vec{w}\cdot x^{(i)} + b$

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

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

compute cost

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

$dw_1 \mathrel{+}= x_1^{(i)}\cdot dz^{(i)}$

$dw_2 \mathrel{+}= x_2^{(i)}\, dz^{(i)}$

$db \mathrel{+}= dz^{(i)}$

$J \mathrel{/}= m$; $dw_1 \mathrel{/}= m$; $dw_2 \mathrel{/}= m$; $db \mathrel{/}= m$

$w_1 = w_1 - \alpha\, dw_1$

$w_2 = w_2 - \alpha\, dw_2$

$b = b - \alpha\, db$

compute derivatives and update $\vec{w}$ and b

Logistic Regression