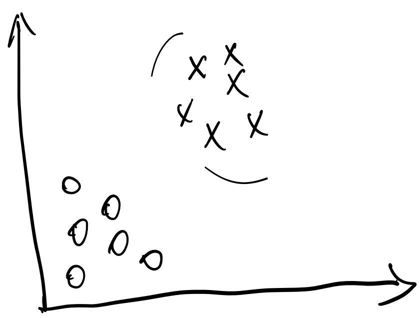
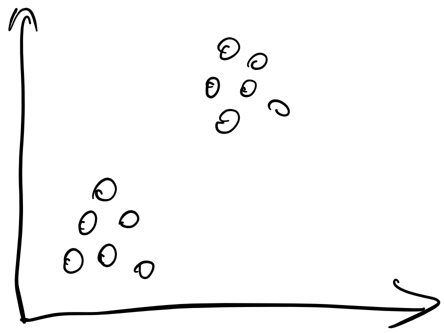
1. Supervised vs Unsupervised learning

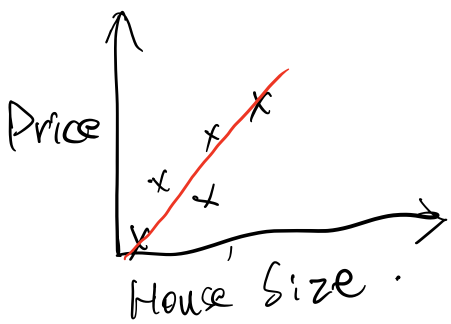


with label in train set



without label

2. Linear Regression



$$f(x) = wx + b$$

$\hat{y}$: prediction value.

y : actual value.

How to evaulate the model ?

Use Cost function:

$$J(f(x)) = \frac{1}{2m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

$$= \frac{1}{2m} \sum_{i=1}^{m} (y_i - (wx_i + b))^2$$

Minimize Cost Function:

Gradient Descent:

$\alpha > 0$

$\frac{\partial}{\partial w} J > 0$ on the right side.

$$w' = w - \alpha \frac{\partial}{\partial w} [J \ f(x)]$$

$$b' = b - \alpha \frac{\partial}{\partial b} [\ (f(x))]$$

$$w' = w - \frac{1}{2m} \sum_{i=1}^{m} (y_i - w x_i - b) 2x_i$$

$$= w - \frac{1}{m} \sum_{i=1}^{m} (_{w,b}(x^i) - y^i) x^i$$

$$b' = b - \frac{1}{m} \sum_{i=1}^{m} f_{w,b}(x^i) - y^i)$$

$$w' = w - \alpha \frac{\partial}{\partial w} J(f_{w,b}(x))$$

$$b' = w - \alpha \quad {}_{b} J(f_{w,b}(x))$$

The problem of α

$\alpha = 0.001$

α too small, too many iterations needed.

$\alpha = 0.001$

$\alpha = 1000$

$\alpha = 0.1$

$\alpha = 1000$

α too big, can't converge.

Univariate regression :

$y = wx + b$ x : size of house

Multiple linear regression :

$y = w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + b$

x_1 : size x_2 : # of bedrooms.

x_3 : Age x_4 : # of floors.

Generalization :

$f_{w,b}(x) = w_1 x_1 + \cdots + w_n x_n + b$

$\vec{w} = [w_1 \ w_2 \ w_3 \cdots w_n]$

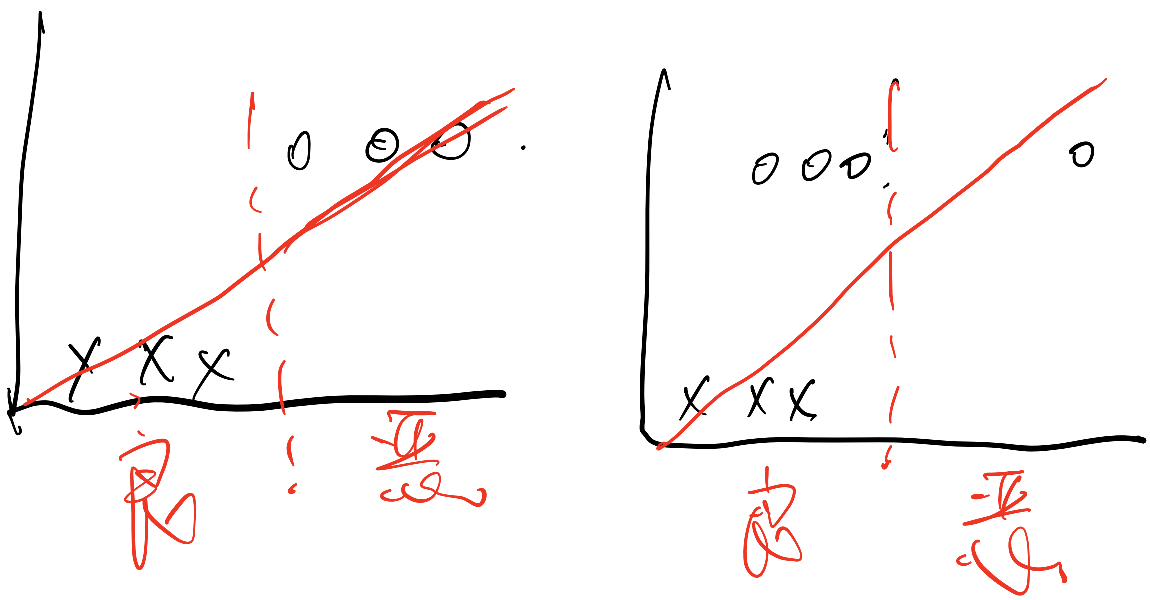
$\vec{x} = [x_1 \ x_2 \ x_3 \cdots x_n]$

dot product

$$f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b =$$

$$w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_n x_n + b$$

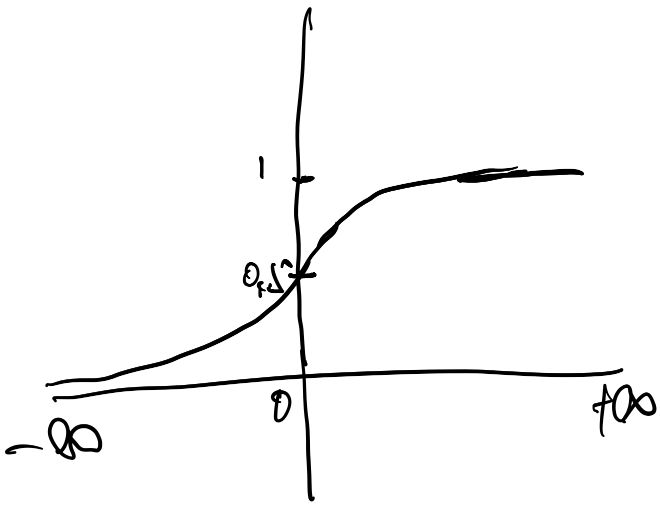
Logistic function:

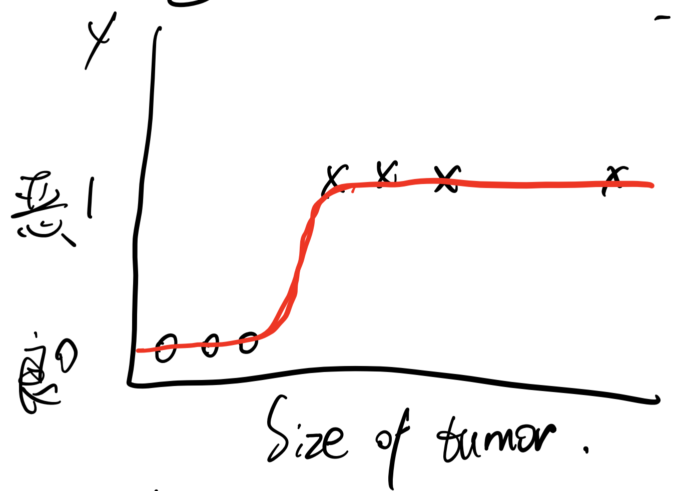


Drawback of Linear Func

$$y = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(wx+b)}}$$

$Z = wx + b$



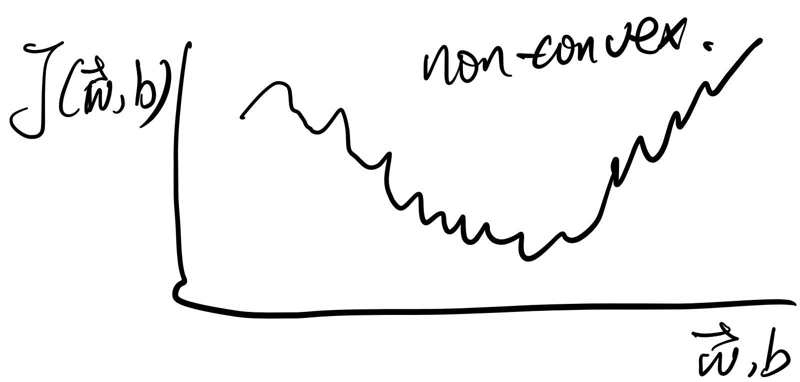


$y = 0.8$, 80% probability of being 恶

Squared Error Cost Function:

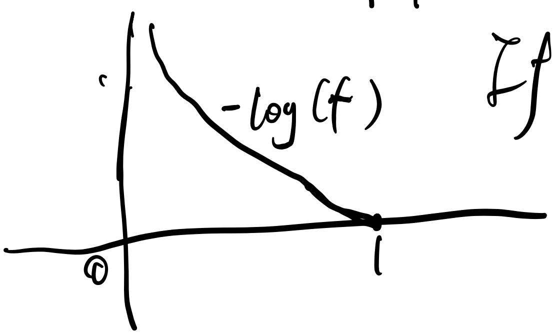
$$J(f_{w,b}(x)) = \frac{1}{2m}\sum_{i=0}^{m-1}(f_{w,b}(x_i) - y_i)^2.$$

Apply SECF on Logistic Regression



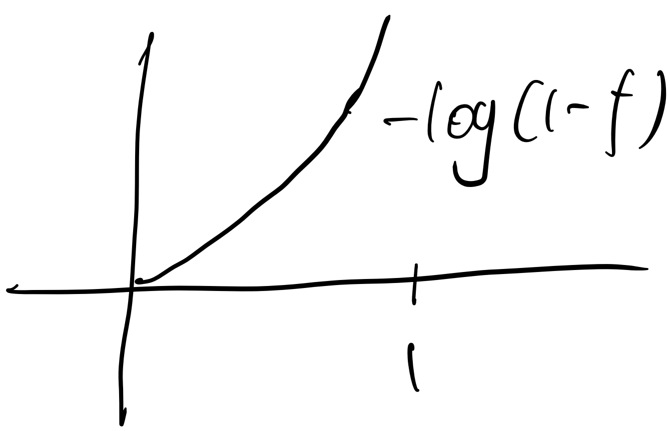
Logistic Loss Function

$$L(f_{w,b}(x^i), y^i) = \begin{cases} -\log(\underbrace{f_{w,b}(x^i)}_{\text{predicted}}), & \underbrace{y=1}_{\text{actual}}. \\ -\log(1-f_{w,b}(x^i)), & y=0 \end{cases}$$



If $f = 1$, $-\log(f) = 0$

$f = 0$, $-\log(f) = \infty$



If $f = 0$, $-\log(1-f) = 0$.

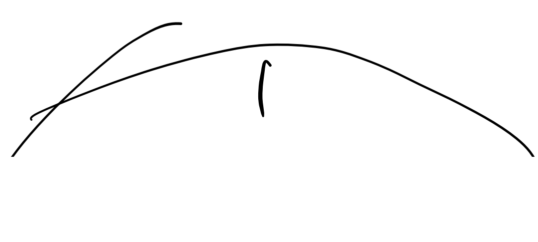
$$f = 1 \quad , \quad -\log(1-f) = \infty$$

$$L(f_{w,b}(x^i), y^i) = \begin{cases} -\log(\underbrace{f_{w,b}(x^i)}_{\text{predicted}}) , & \underbrace{y=1}_{\text{actual}} \\ -\log(1 - f_{w,b}(x^i)) , & y = 0 \end{cases}$$

$$L = -y^i \log(f(x^i)) - (1-y^i) \log(1-f(x^i))$$

Cost Function

$$J(f_{\vec{w},b}(\vec{x})) = -\frac{1}{m} \sum_{i=0}^{m-1} [y^i \log(f(x^i)) + (1-y^i) \log(1-f(x^i))$$



$$y = \frac{\quad}{1 + e^{-z}}$$

$$\vec{z} = \vec{w} \cdot \vec{x} + b.$$

features.

Data	Size (x_1)	Age (x_2)	# bed (x_3)	# floor (x_4)
x^1				
x^2				
$\vdots$ m				
x^n				

Data points

$$\vec{x}^m = [x_1^m \quad x_2^m \quad x_3^m \quad x_4^m]$$

$$\odot$$

$$[w^1 \quad w^2 \quad w^3 \quad w^4]$$