Logistic Regression

10 October 2022 21:44

Classification

Question	Answer "y"	
Is this email spam?	no	yes
Is the transaction fraudulent?	no	yes
Is the tumor malignant?	no	yes

y can only be one of two values

"binary classification"

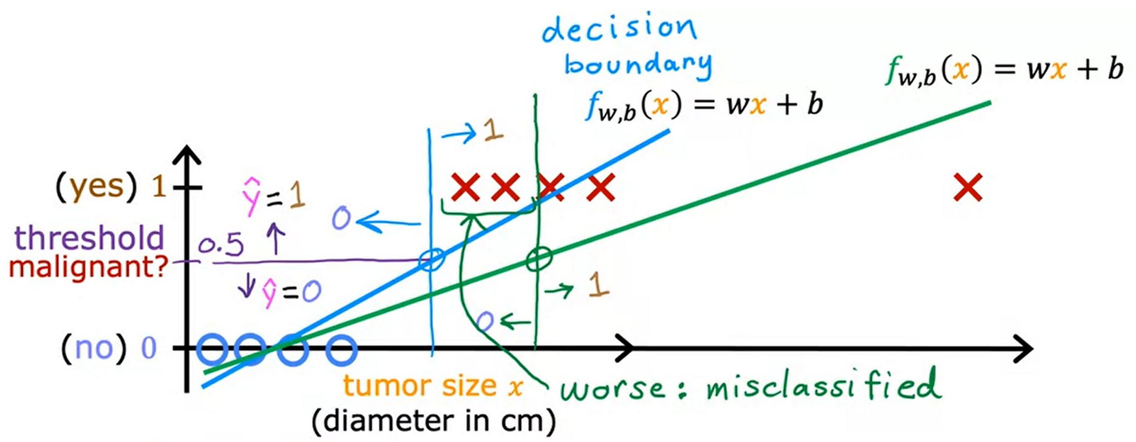
class = category

false true

0 1

useful for classification

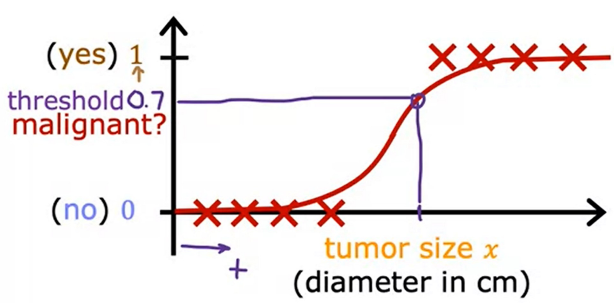
"negative class" "positive class"

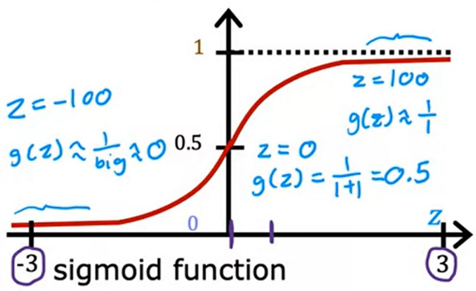


if $f_{w,b}(x) < 0.5 \rightarrow \hat{y} = 0$

if $f_{w,b}(x) \geq 0.5 \rightarrow \hat{y} = 1$

Want outputs between 0 and 1





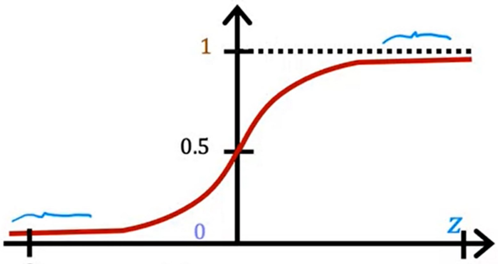
sigmoid function

logistic function

outputs between 0 and 1

$g(z) = \frac{1}{1+e^{-z}} \quad 0 < g(z) < 1$

Want outputs between 0 and 1



sigmoid function

logistic function

$$\boxed{\begin{array}{l}\text{outputs between } 0 \text{ and } 1\\ g(z) = \frac{1}{1+e^{-z}} \qquad 0 < g(z) < 1\end{array}}$$

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

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

$$\downarrow z \downarrow$$

$$\boxed{g(z) = \frac{1}{1+e^{-z}}}$$

$$f_{\vec{w},b}(\vec{x}) = g(\underbrace{\vec{w} \cdot \vec{x} + b}_{z}) = \frac{1}{1 + e^{-(\vec{w}\cdot\vec{x}+b)}}$$

"logistic regression"

$$f_{\vec{w},b}(\vec{x}) = g(\underbrace{\vec{w} \cdot \vec{x} + b}_{z}) = \frac{1}{1 + e^{-(\vec{w}\cdot\vec{x}+b)}}$$

$= P(y = 1|x; \vec{w}, b)$ 0.7 0.3

0 or 1? threshold

Is $f_{\vec{w},b}(\vec{x}) \geq \overbrace{0.5}$?

Yes: $\hat{y} = 1$ No: $\hat{y} = 0$

When is $f_{\vec{w},b}(\vec{x}) \geq 0.5$?

$$g(z) \geq 0.5$$

$$z \geq 0$$

$\vec{w} \cdot \vec{x} + b \geq 0$ $\qquad$ $\vec{w} \cdot \vec{x} + b < 0$

$\hat{y} = 1$ $\qquad$ $\hat{y} = 0$

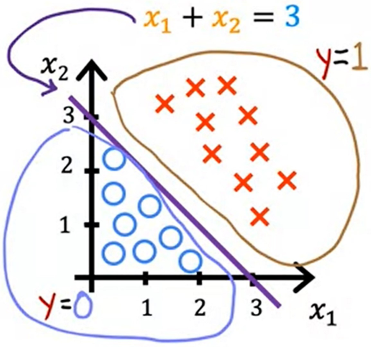
Decision boundary

$$f_{\vec{w},b}(\vec{x}) = g(z) = g(\underbrace{\underset{1}{w_1} x_1 + \underset{1}{w_2} x_2 + \underset{-3}{b}})$$

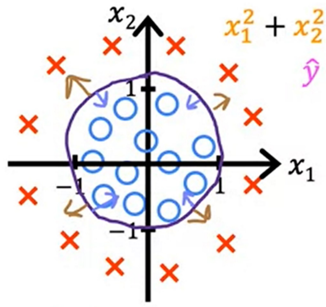
Decision boundary $z = \vec{w} \cdot \vec{x} + b = 0$

$z = x_1 + x_2 - 3 = 0$

$x_1 + x_2 = 3$



Non-linear decision boundaries



$x_1^2 + x_2^2 \geq 1$

$\hat{y} = 1$

$$f_{\vec{w},b}(\vec{x}) = g(z) = g(\overbrace{\underset{1}{w_1} x_1^2 + \underset{1}{w_2} x_2^2 + \underset{-1}{b}}^{z})$$

decision boundary $\quad z = x_1^2 + x_2^2 - 1 = 0$

$$x_1^2 + x_2^2 = 1$$

$x_1^2 + x_2^2 < 1$

$\hat{y} = 0$

Cost

cost

$$J(\vec{w}, b) = \frac{1}{m}\sum_{i=1}^{m} L(\underbrace{f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)}}_{\text{loss}})$$

$$= \begin{cases} -\log\left(f_{\vec{w},b}(\vec{x}^{(i)})\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}(\vec{x}^{(i)})\right) & \text{if } y^{(i)} = 0 \end{cases}$$

convex
↳ can reach a global minimum

find w, b that minimize cost J

Simplified cost function

loss

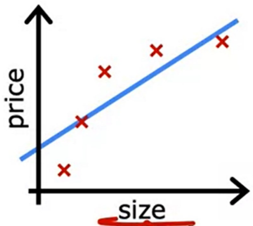
$$L(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)}) = -\underbrace{y^{(i)}\log\left(f_{\vec{w},b}(\vec{x}^{(i)})\right) - (1 - y^{(i)})\log\left(1 - f_{\vec{w},b}(\vec{x}^{(i)})\right)}$$

cost

$$J(\vec{w}, b) = \frac{1}{m}\sum_{i=1}^{m}\left[\underbrace{L(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)})}\right]$$

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

Regression example

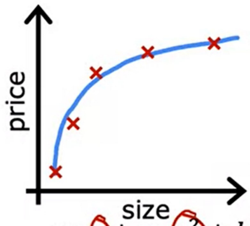


$w_1x + b$

underfit

- Does not fit the training set well

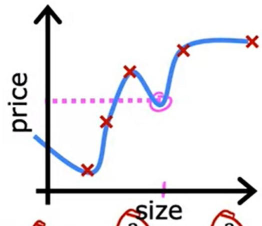
high bias



$w_1x + w_2x^2 + b$

- Fits training set pretty well

generalization



$w_1x + w_2x^2 + w_3x^3 + w_4x^4 + b$

overfit

- Fits the training set extremely well

high variance

Select features to include/exclude

size	bedrooms	floors	age	avg income	...	distance to coffee shop	price
x_1	x_2	x_3	x_4	x_5		x_{100}	y

all features

+

insufficient data

↓

overfit

selected features

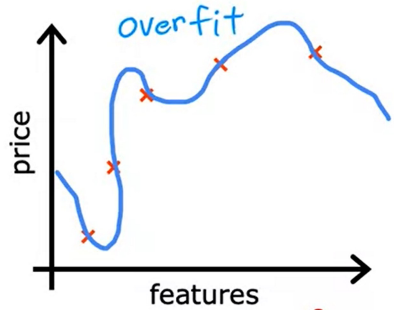
size
bedrooms
age
just right

feature selection

Regularization

Reduce the size of parameters w_j

overfit

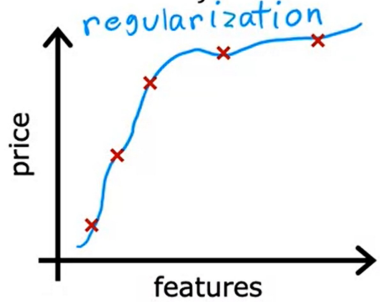


$f(x) = 28x - 385x^2 + 39x^3 - \underbrace{\cancel{174}x^4}_{0} + 100$

large values for w_j

eliminate feature

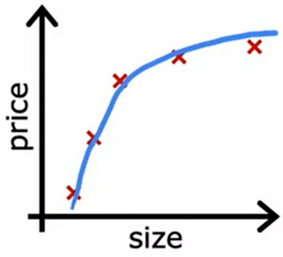
regularization



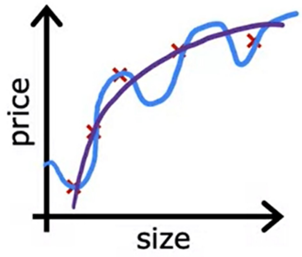
$f(x) = 13x - 0.23x^2 + \underbrace{0.000014}x^3 - \underbrace{0.0001}x^4 + 10$

small values for w_j

Intuition



$w_1x + w_2x^2 + b$



$w_1x + w_2x^2 + \underbrace{\cancel{w_3x^3}}_{\approx 0} + \underbrace{\cancel{w_4x^4}}_{\approx 0} + b$

make w_3, w_4 really small (≈ 0)

$$\min_{\vec{w},b} \frac{1}{2m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)^2 + 1000\underbrace{w_3^2}_{0.001} + 1000\underbrace{w_4^2}_{0.002}$$

Regularization

small values $w_1, w_2, \cdots, w_n, b$

simpler model $w_3 \approx 0$

less likely to overfit $w_4 \approx 0$

size	bedrooms	floors	age	avg income	...	distance to coffee shop	price
x_1	x_2	x_3	x_4	x_5		x_{100}	y

$w_1, w_1, w_2, \cdots, w_{100}, b$ n features $n = 100$

$$J(\vec{w},b) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)^2 + \overbrace{\frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2}^{\text{regularization term}} + \frac{\lambda}{2m}b^2\right]$$

can include or exclude b

"lambda" regularization parameter $\lambda > 0$

Regularized logistic regression

$$\min_{\vec{w},b} J(\vec{w},b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\vec{w},b}(\vec{x}^{(i)})\right) + (1 - y^{(i)})\log\left(1 - f_{\vec{w},b}(\vec{x}^{(i)})\right)\right] + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$$

Gradient descent

Looks same as for linear regression!

repeat {

$$w_j = w_j - \alpha\frac{\partial}{\partial w_j}J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m}\left[\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)x_j^{(i)}\right] + \frac{\lambda}{m}w_j$$

$j = 1 \ldots n$

logistic regression

$$b = b - \alpha\frac{\partial}{\partial b}J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)$$

}

don't have to regularize