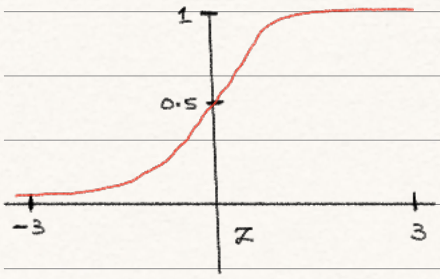
Subject: Classification / /

Logistic Regression:

Sigmoid/logistic function → output b/w 0 & 1

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



$z >>$, $g(z) \approx 1$

$z <<$, $g(z) \approx 0$

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

$z = w \cdot x + b$

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

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

Decesion boundary:

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

Pick threshold

if $f_{w,b}(x) \geq 0.5$?

Yes: $\hat{y} = 1$

Subject: / /

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

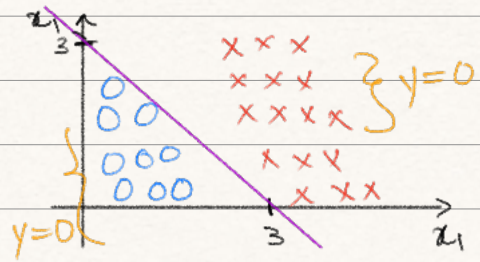
$$g(z) \geq 0.5$$

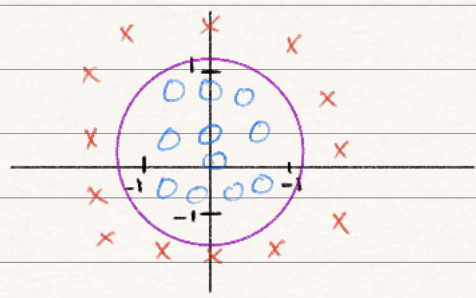
$$z \geq 0$$

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

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

Example,





$$z = x_1 + x_2 - 3 = 0$$

$$x_1 + x_2 = 3$$

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

$$z = x_1^2 + x_2^2 - 1 = 0$$

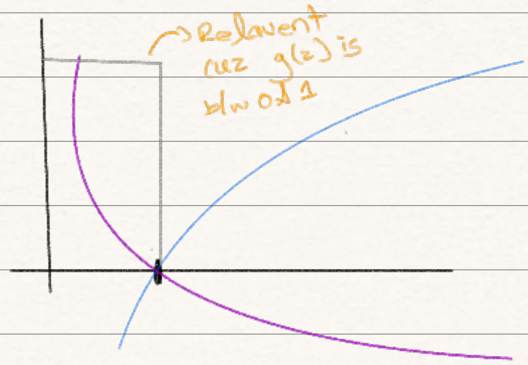
$$x_1^2 + x_2 = 1$$

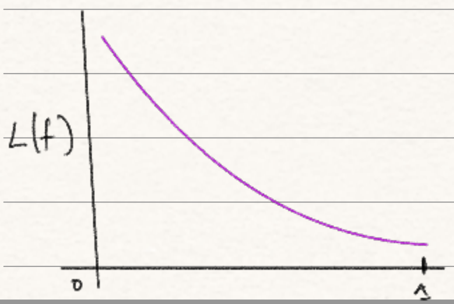
Cost function:

we don't use CF same as Linear Regression because it doesn't create a convex 'U' func

Loss Function

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



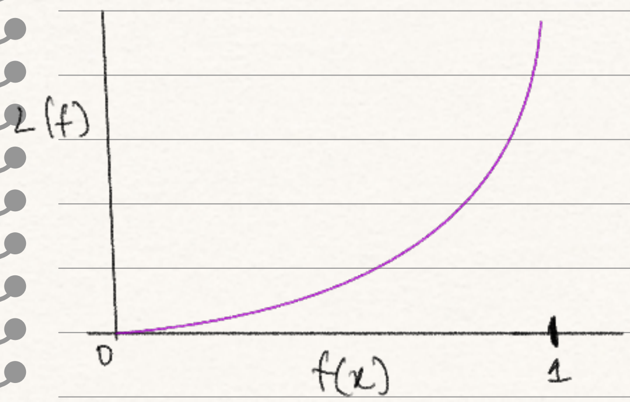


if $y^{(i)} = 1$

As $f(x^{i}) \rightarrow 1$ then loss $\rightarrow 0$

As $f(x^{i}) \rightarrow 0$ then loss $\rightarrow \infty$

Subject: / /



if $y^i = 0$

As $f_{\vec{w},b}(\vec{x}^{(i)}) \to 1$ then loss $\to \infty$

As $f_{\vec{w},b}(\vec{x}^{(i)}) \to 0$ thes loss $\to 0$

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

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

$\downarrow$

simlified version

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

Gradient descent:

repeat {

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

no. of features

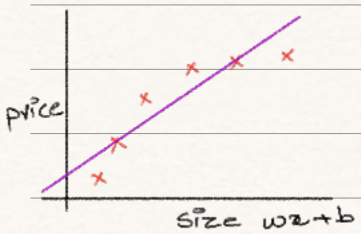
$$b = b - \alpha \underbrace{\frac{\partial}{\partial w_j} J(\vec{w},b)}$$ $$\frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b}(\vec{x}^{(i)}) - y^{(i)}\right)$$

}

Subject: / /

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

The problem of overfitting:



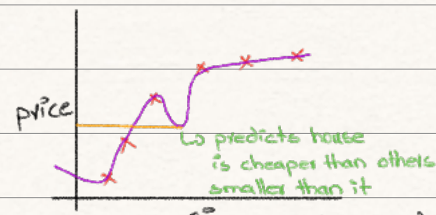
$wx+b$

less features

Underfit

- Does not fit the training set well

High bais



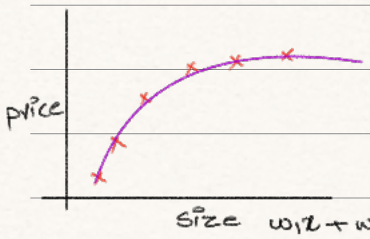
$w_1x+w_2x^2+w_3x^3+w_4x^4+b$

more features

Overfit

- Fits extremely well

high variance



$w_1x+w_2x^2+b$

Just right

just right

- Fits well

Generalization

↳ Good prediction on unseen examples

Addressing Overfitting:

- Collect more training model
- Use fewer features
- Reduce size of parametres w_j

all features

\+

less data

↓

overfit

select features } feature selection

↓

just right

Cons:

- Useful features lost

Subject: / /

Regularization:

- Reduces size of parametres w_j to remove over fitting
- Penalize all parameters

Regularization Parametre

$$J(\vec{w},b) = \frac{1}{2m}\sum_{i=0}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$$

Penalizing b makes little difference

$$\ldots + \frac{\lambda}{2m} b^2$$

$\lambda = 0 \longrightarrow$ over fit

$\lambda = 10^{10} \longrightarrow$ under fit

Regularized Gradient descent For Linear Regression:

repeat {

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

no. of features

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

}

Regularized Gradient descent For Logistic Regression:

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

$$+ \frac{\lambda}{2m}\sum_{j=1}^{m} w_j^2$$

Subject: / /

repeat {

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

no. of features

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

}