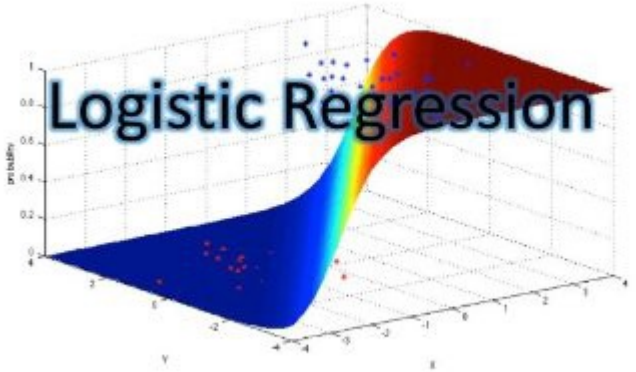


Logistic Regression

~Abhishek Kumar

Logistic Regression

- Supervised
- It is not a regression but a Classfication algorithm
- Example:
  - Spam/Not spam,
  - Tumor: Malignant/Benign

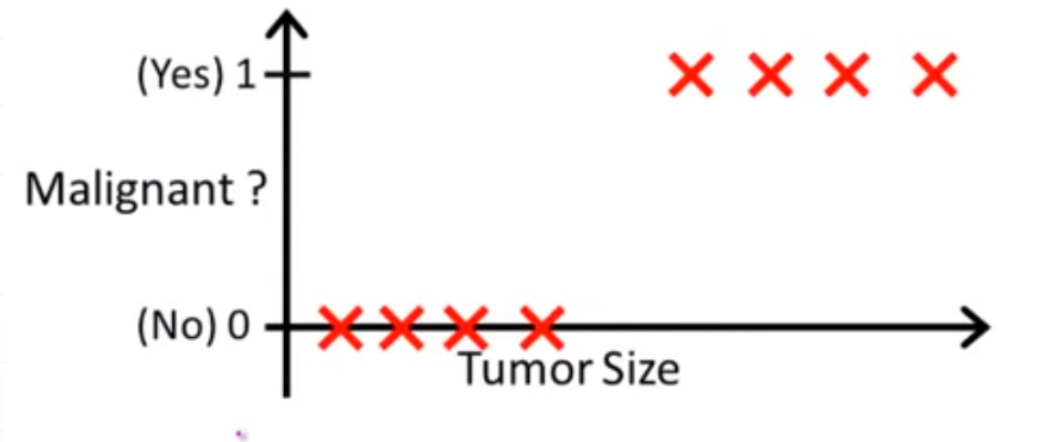
In case of Linear Regression

- Threshold classifier output

h(x) at 0.5

if h(x) >= 0.5, predicts “y=1”

if h(x) < 0.5, predicts “y=0”



- Classification y = 0 or 1

h(x) can be >1 or <0

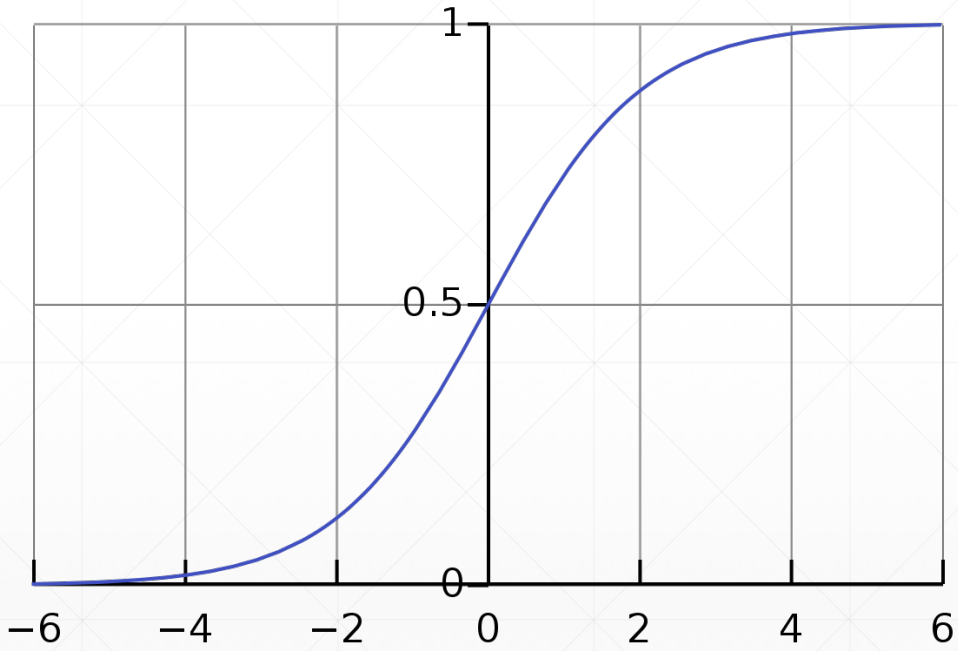
- Logistic Regression 0 <= h(x) <= 1

Logistic Regression

- Want 0 <= h(x) <= 1
- h(x) = g(mx+c)

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

- Sigmoid/ Logistic function
- Fit the parsmetrs to our data



Interpretation of our model

h(x) = estimated probability that y = 1, input x

Tumor example

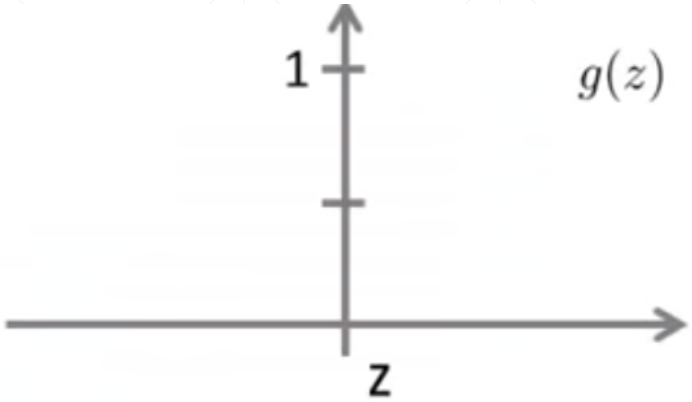
h(x) = 0.7 ---- Pateint has 70% chance of tumor being malignant

P(y = 0) + P(y = 1) = 1

Logistic regression

$$h_\theta(x) = g(\theta^T x)$$

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



We can suppose:

“y = 1” if h(x) >= 0.5

g(z) >= 0.5 when z >= 0

h(x) = g(mx + c) >= 0.5 so mx+c >= 0

“y=0” if h(x) < 0.5

g(z) < 0.5 when z < 0

h(x) = g(mx + c) < 0.5 so mx+c < 0

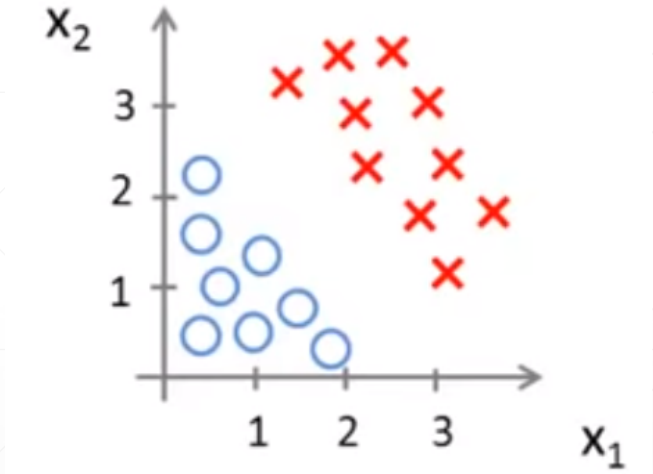
Decision Boundary

$h(x) = g\ (c + m_1x_1 + m_2x_2)$

$c = -3,\ m_1 = 1,\ m_2 = 1$

Predict “y =1” if $-3 + x_1 + x_2 >= 0$

$x_1 + x_2 >= 3$



Seprates the region where y = 1 and y = 0

Decision boundary is the property of the parameters of the hypothesis not the property of dataset

Another example

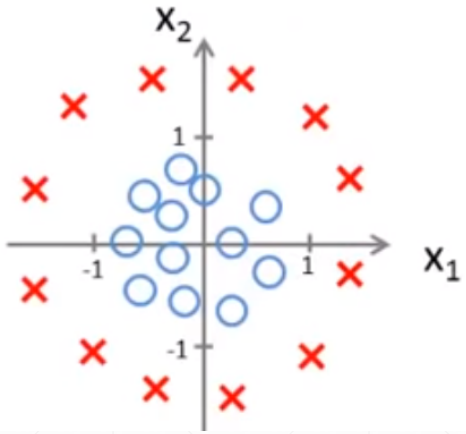
$h(x) = g\,(c + m_1x_1 + m_2x_2 + m_3x_1^2 + m_4x_2^2)$

$c=0\ ,\ m_1 = 0,\ m_2 = 0,\ m_3 = 1,\ m_4 = 1$

Predict “y = 1” if $-1 + x_1^2 + x_2^2 >= 0$

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

Even we can do it for higher polynomial



Cost function

Linear regression: $J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$

Non – convex function

Come up with a different function that is convex

Logistic regression cost function

$$\text{Cost}(h(x), y) = \begin{cases} -\log(h(x)) & \text{if } y = 1 \\ -\log(1 - h(x)) & \text{if } y = 0 \end{cases}$$

Discussion

Thank you!