MACHINE LEARNING

اَللّٰهُمَّ اُرْزُقْنِى عِلْمًا نَافِعًا وَاسِعًا عَمِيْقًا

اَللّٰهُمَّ اُرْزُقْنِی رِزْقًا وَاسِعًا حَلَالًا طَيِّبًا
مُبَارَكًا مِنْ عِنْدِكَ

WEEK 09

LOGISTIC REGRESSION

REIVEW: LINEAR REGRESSION

Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

Parameters: θ_0, θ_1

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

LINEAR REGRESSION AS CLASSIFICATION MODEL

TYPE OF MACHINE LEARNING

- Supervised Learning.
- Unsupervised Learning.
- Reinforcement Learning.

SUPERVISED LEARNING: REGRESSION

Living area (feet^2)	Price (1000\$s)
2104	400
1600	330
2400	369
1416	232
3000	540
$\vdots$	$\vdots$

SUPERVISED LEARNING: REGRESSION

- When we try to predict a number from historical data this type of supervised learning problem is called Regression Problem

SUPERVISED LEARNING: CLASSIFICATION

- What of Machine Learning ?
  - Supervised Learning
    - Classification
      - Binary Classification

SUPERVISED LEARNING: CLASSIFICATION

x1	x2	Type
-7	1	Positive
-4	4	Positive
-1	-3	Negative
+2	-2	Negative
-6	2	Positive
+4	-1	Negative
-5	3	Positive
+3	0	Negative
+1	5	Positive
+2	+1	Negative

CLASSIFICATION: MORE FORMALLY

Given: Training data: $(x_1, y_1), \ldots, (x_n, y_n) / x_i \in \mathbb{R}^d$ and y_i is discrete (categorical/qualitative), $y_i \in \mathbb{Y}$.

Example $\mathbb{Y} = \{-1, +1\}, \mathbb{Y} = \{0, 1\}$.

Task: Learn a classification function:

$$f : \mathbb{R}^d \longrightarrow \mathbb{Y}$$

Linear Classification: A classification model is said to be linear if it is represented by a linear function f (linear hyperplane)

Credits: EDX ColumbiaX CSMM.101x Artificial Intelligence (AI)

CLASSIFICATION: EXAMPLE

1. Email Spam/Ham $\rightarrow$ Which email is junk?
2. Tumor benign/malignant $\rightarrow$ Which patient has cancer?
3. Credit default/not default $\rightarrow$ Which customers will default on their credit card debt?

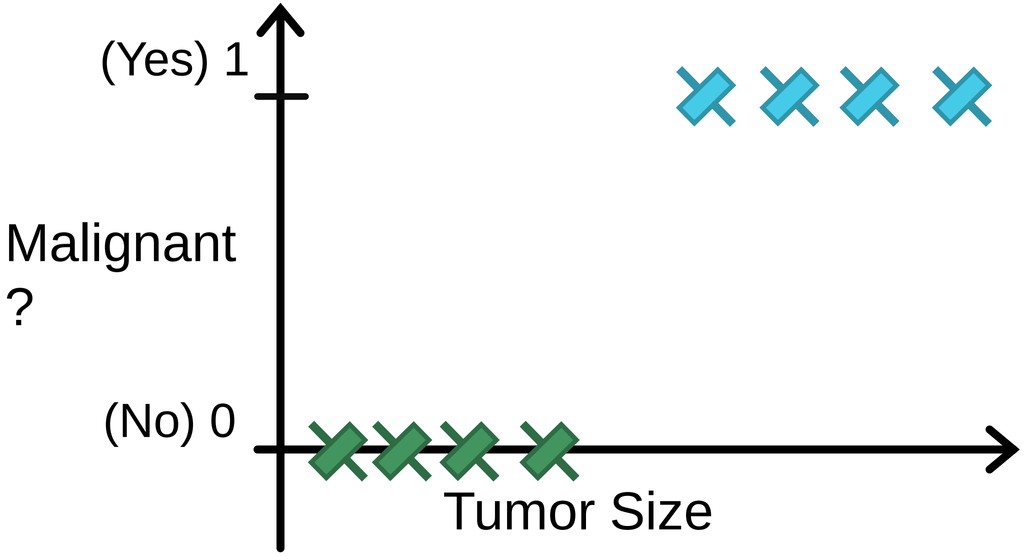
Balance	Income	Default
300	$20,000.00	no
2000	$60,000.00	no
5000	$45,000.00	yes
.	.	.
.	.	.
.	.	.

$$y \in \{0, 1\}$$

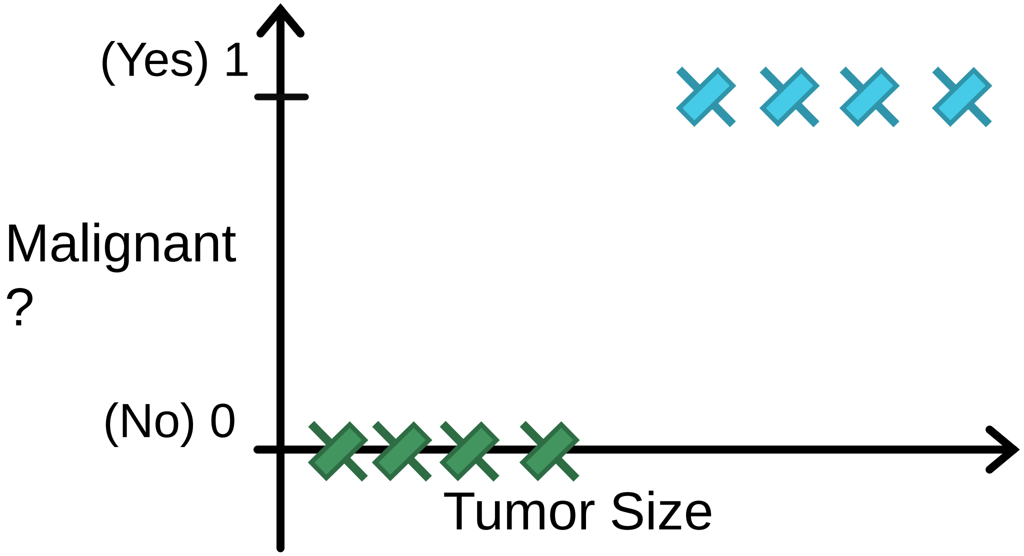
0: "Negative Class" (e.g., benign tumor)
1: "Positive Class" (e.g., malignant tumor)

Credits: EDX ColumbiaX CSMM.101x Artificial Intelligence (AI)

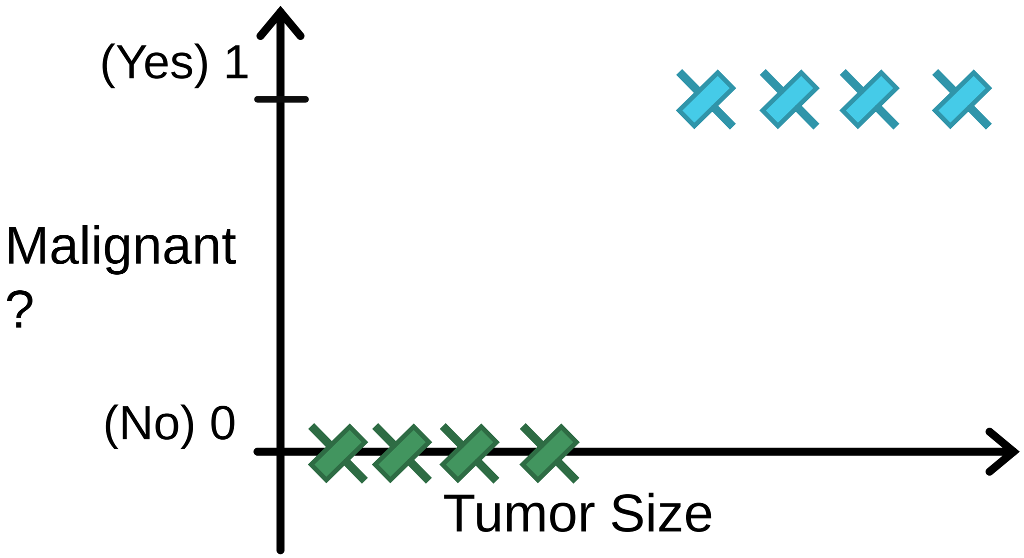
CLASSIFICATION: DATA VISUALIZATION



CLASSIFICATION: HOW TO CLASSIFY

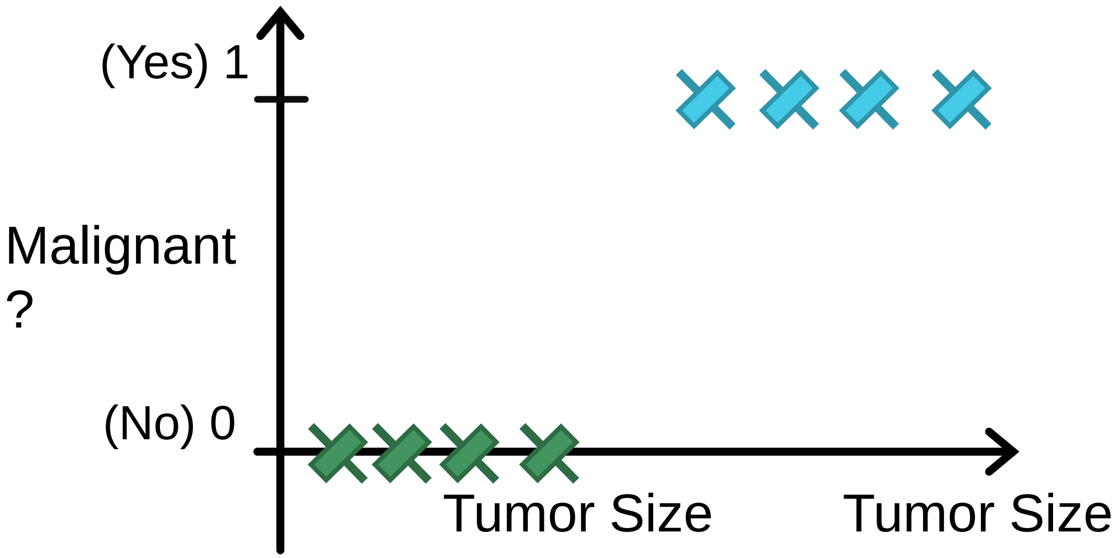


CLASSIFICATION



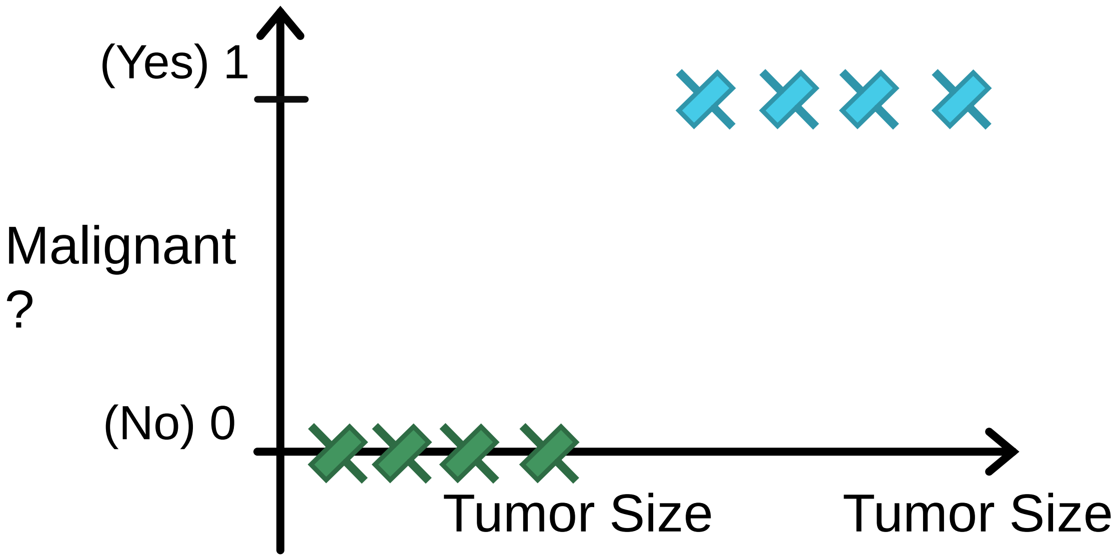
Can we Use Linear Regression as Binary Classifier ?

LINEAR REGRESSION AS BINARY CLASSIFIER



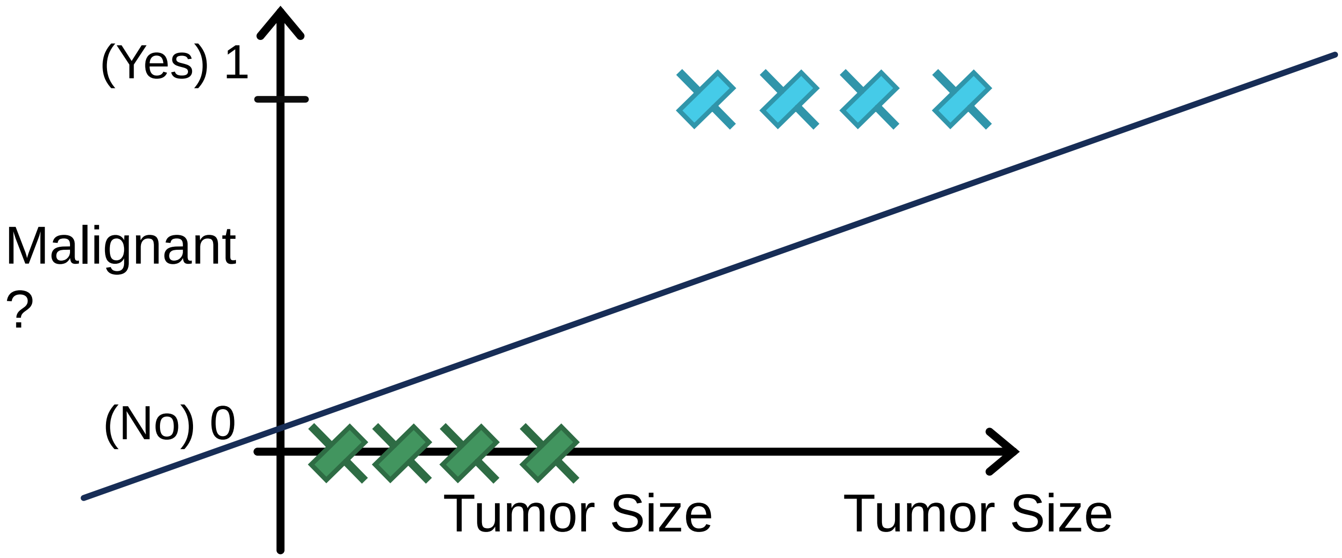
ANY IDEA HOW ?

LINEAR REGRESSION AS BINARY CLASSIFIER



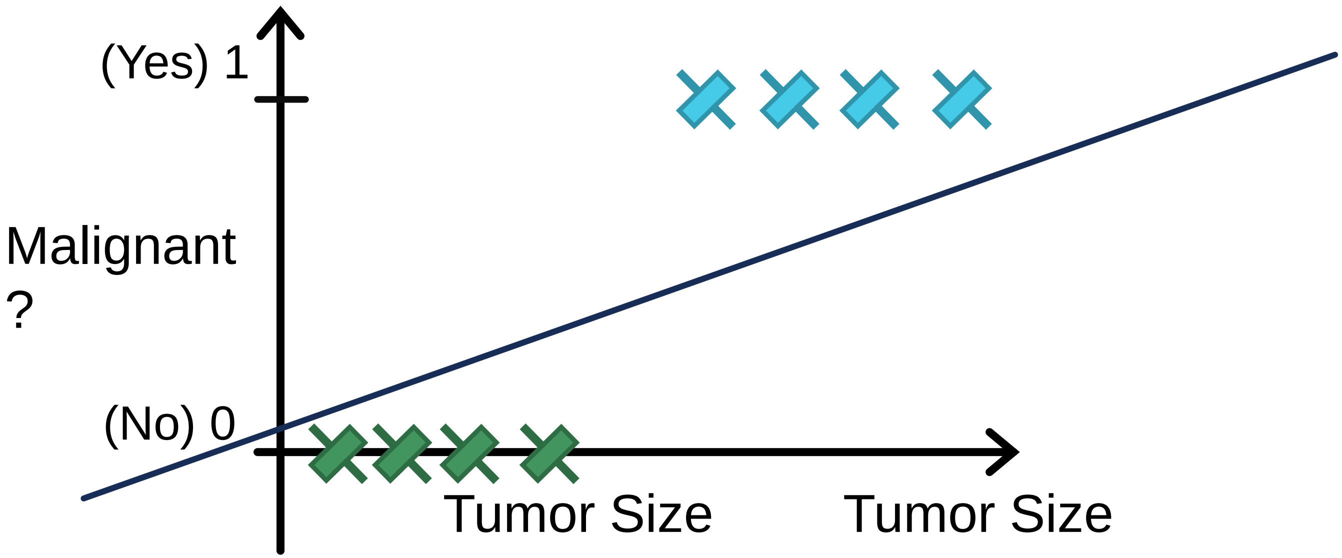
We shall find a line with minimum error and this line predict the value.

LINEAR REGRESSION AS BINARY CLASSIFIER



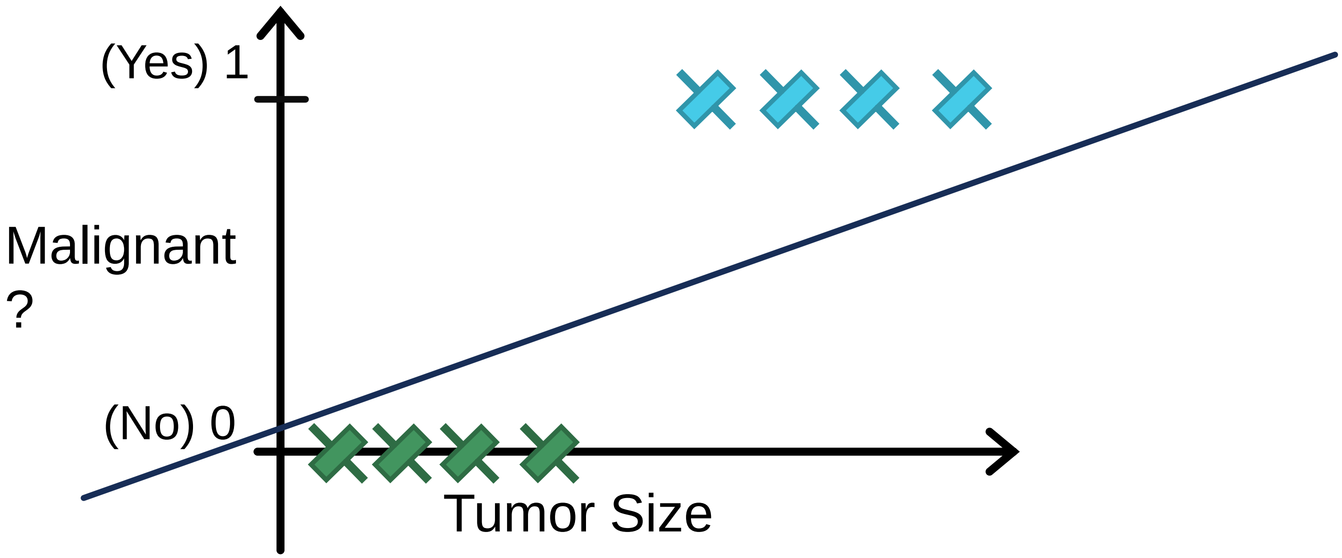
However line return the real values but we need 1 or 0 so what can we do ?

LINEAR REGRESSION AS BINARY CLASSIFIER



We can add **threshold** on y value if the line predict value greater than 0.5 we shall shall predict **Malignant**

LINEAR REGRESSION AS BINARY CLASSIFIER

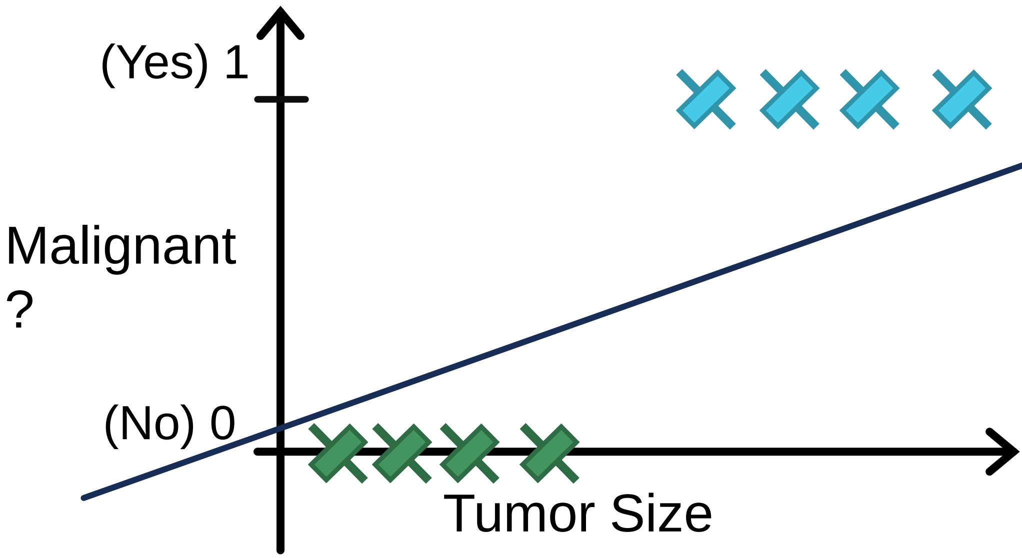


Threshold classifier output $h_\theta(x)$ at 0.5:

If $h_\theta(x) \geq 0.5$, predict "y = 1"

If $h_\theta(x) < 0.5$, predict "y = 0"

LINEAR REGRESSION AS BINARY CLASSIFIER



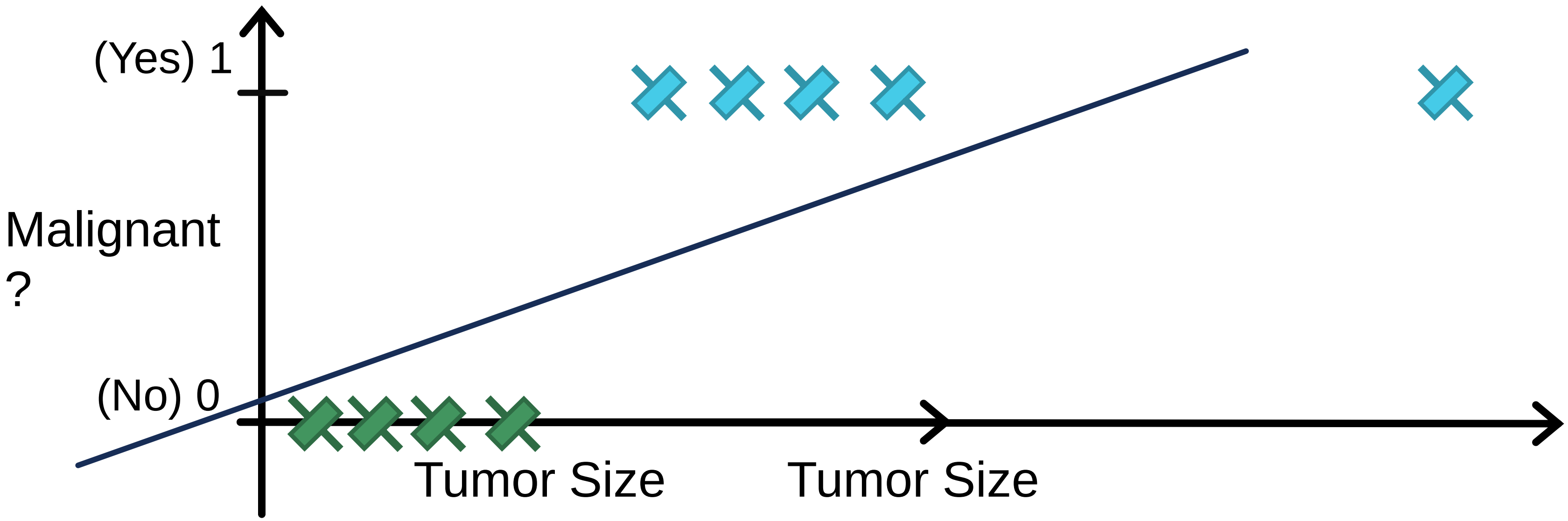
Can you highlight any problem with it ?

Threshold classifier $h_\theta(x)$ output at 0.5:

If $h_\theta(x) \geq 0.5$, predict "y = 1"

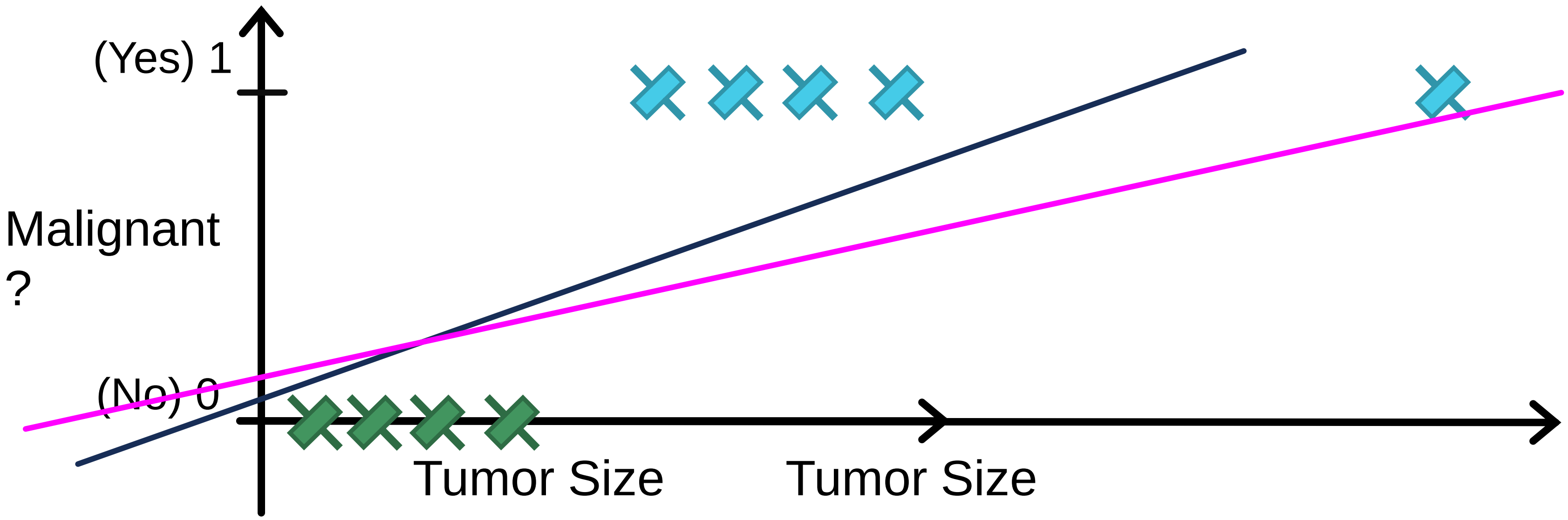
If $h_\theta(x) < 0.5$, predict "y = 0"

LINEAR REGRESSION AS BINARY CLASSIFIER



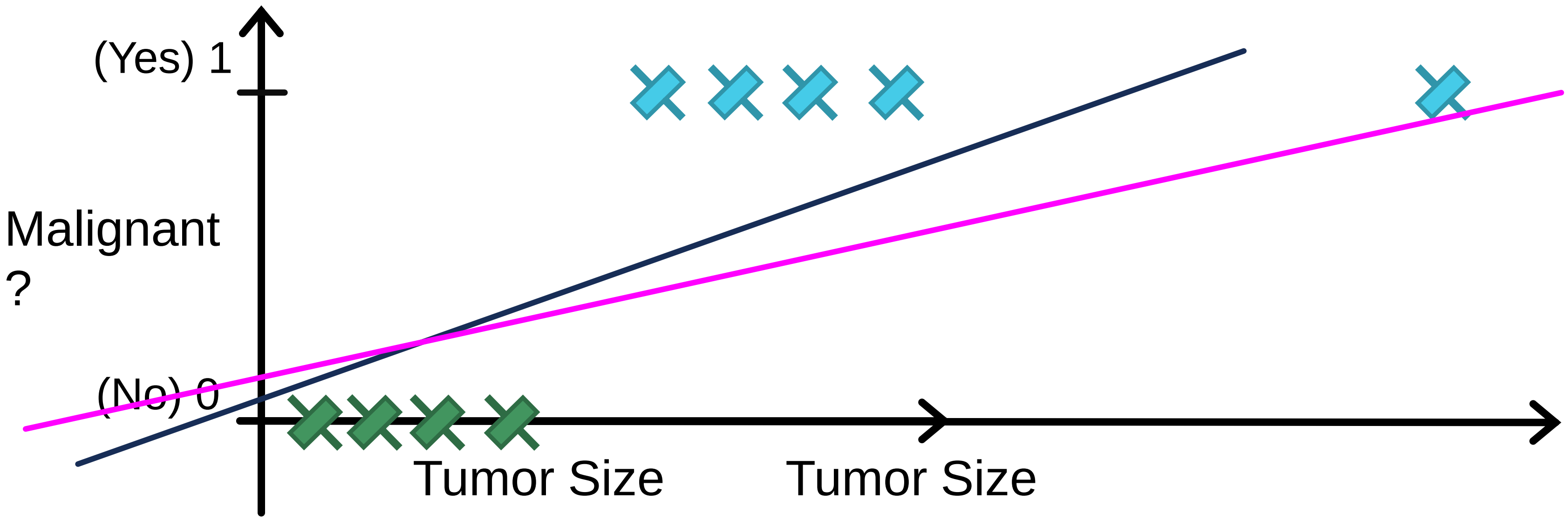
1. single instance can change the predication line drastically and make the predication with a lot of errors.

LINEAR REGRESSION AS BINARY CLASSIFIER



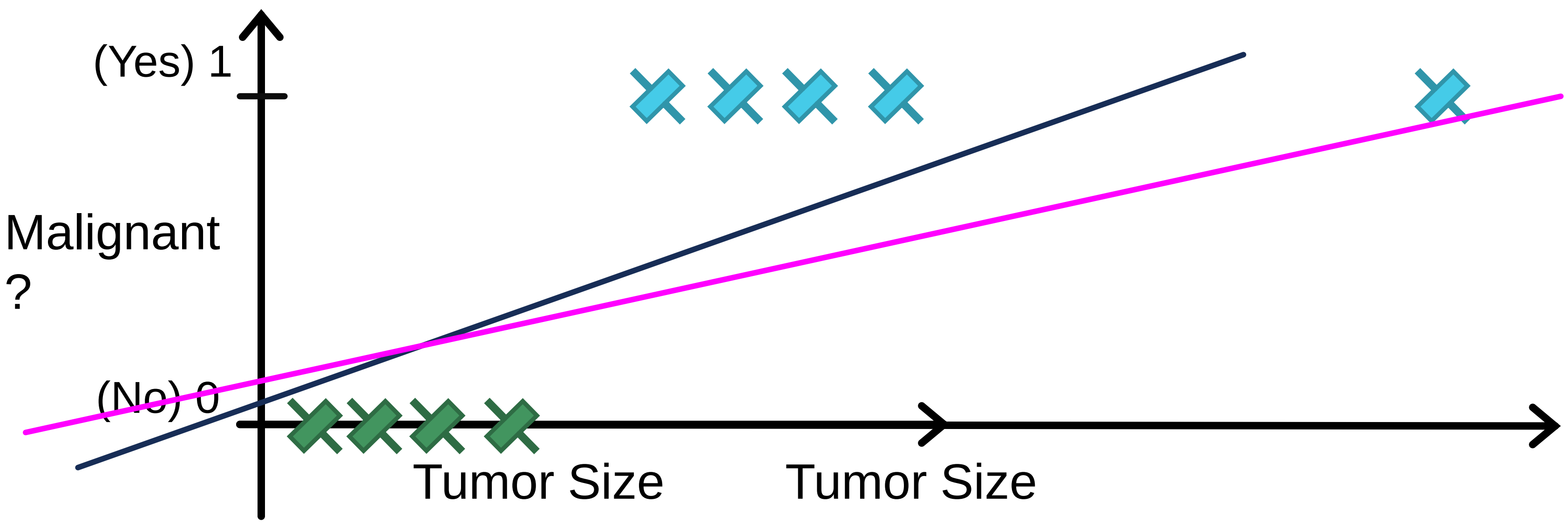
1. single instance can change the predication line drastically and make the predication with a lot of errors.

LINEAR REGRESSION AS BINARY CLASSIFIER



2. For comparative large or comparative small tumor size the predication could be **greater than 1** and **less than 0**

LINEAR REGRESSION AS BINARY CLASSIFIER



3. We can't predict Malignant Cell with any certainty. we want to predict how likely is a Tumor size is Malignant. That is output a probability between 0 and 1 that a cell is malignant

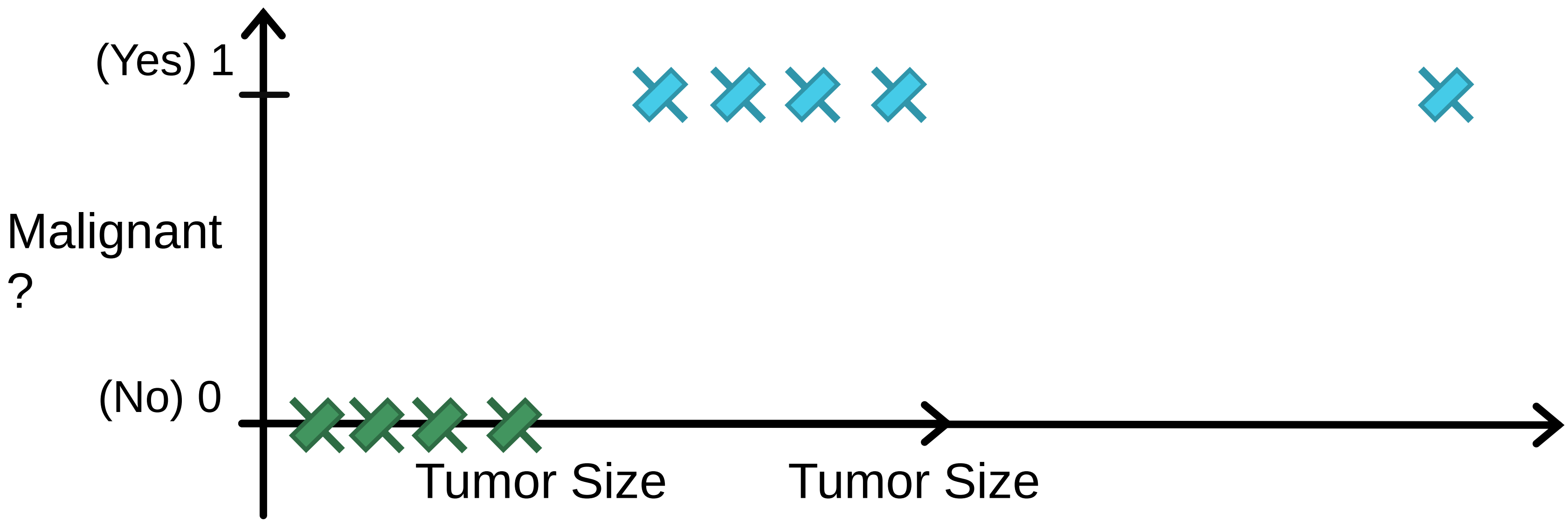
CONCLUSION

- Can we use linear regression for classification ?
- Yes. However...
  - Works only for Binary classification (2 classes).
  - Won't work for Multiclass classification e.g.,Y = Malignant, Benin, Unknown, Critical
  - If we use linear regression, some of the predictions will be outside of [0,1].
  - Model can be poor.

LOGISTIC REGRESSION

LOGISTIC REGRESSION

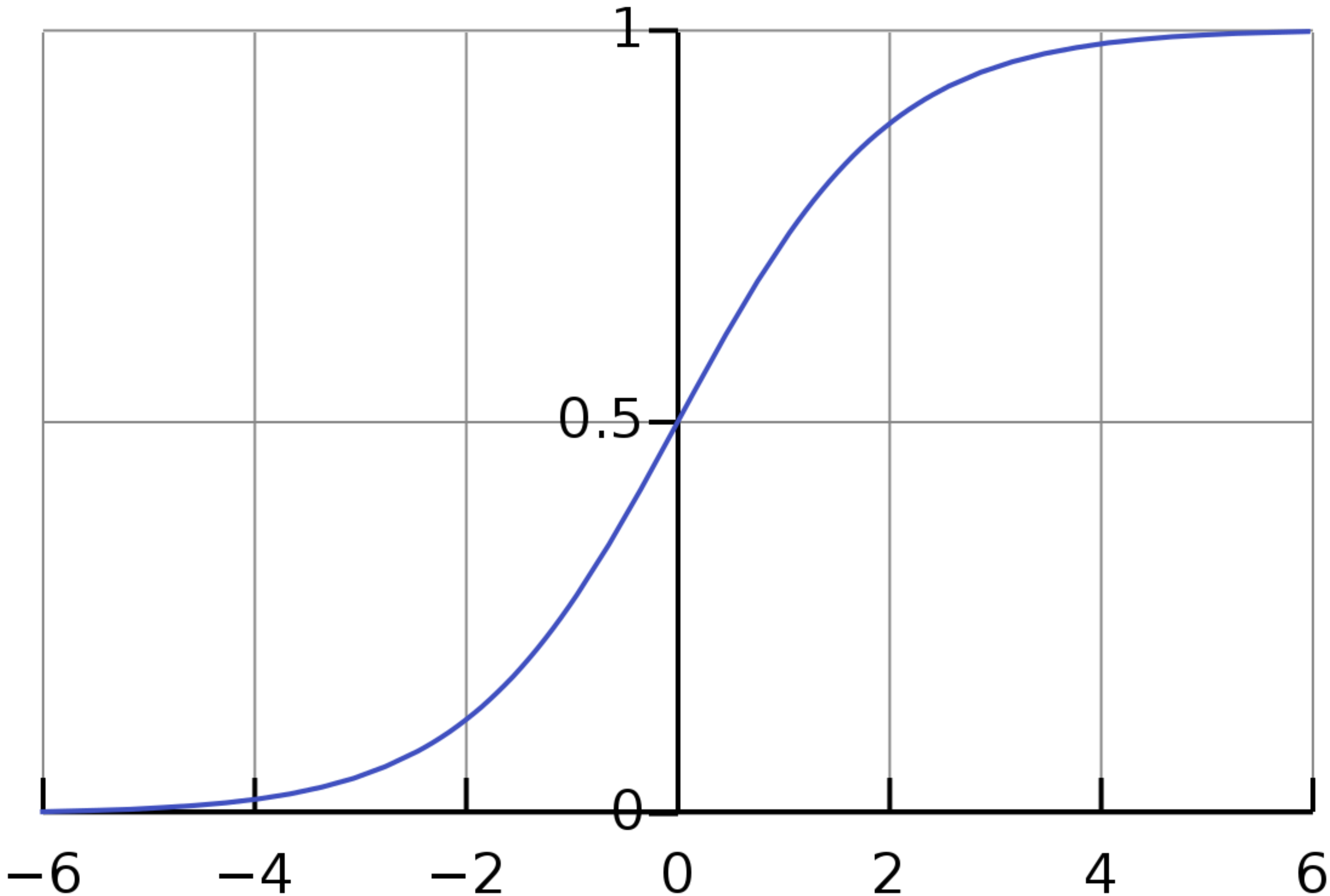
- We need a function that bound values between 0 and 1 so we can consider it as the probability for one class



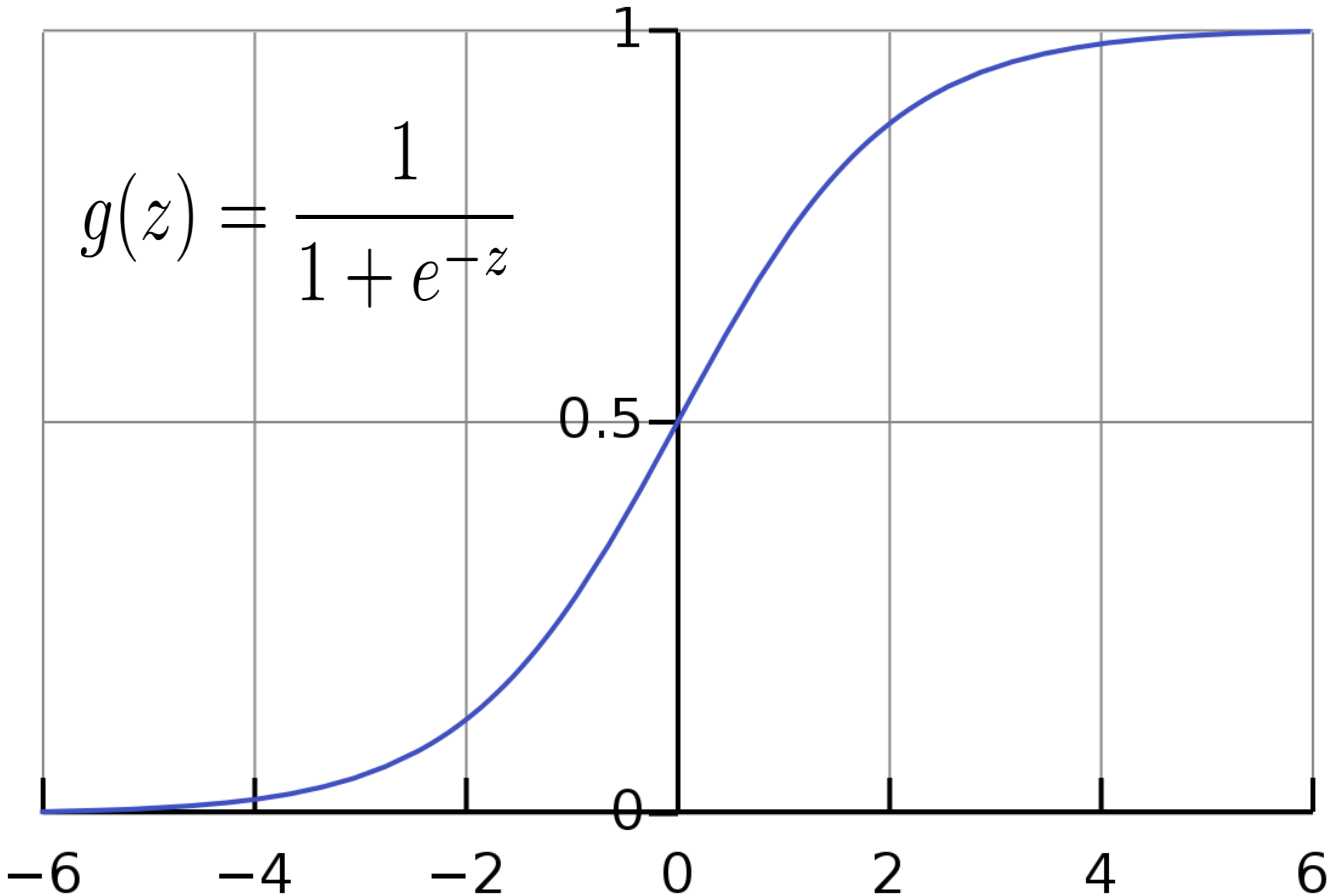
LOGISTIC REGRESSION

- Can you recall any function that return 0 for smaller values of given x an return 1 for larger values of given x.

SIGMOID FUNCTION



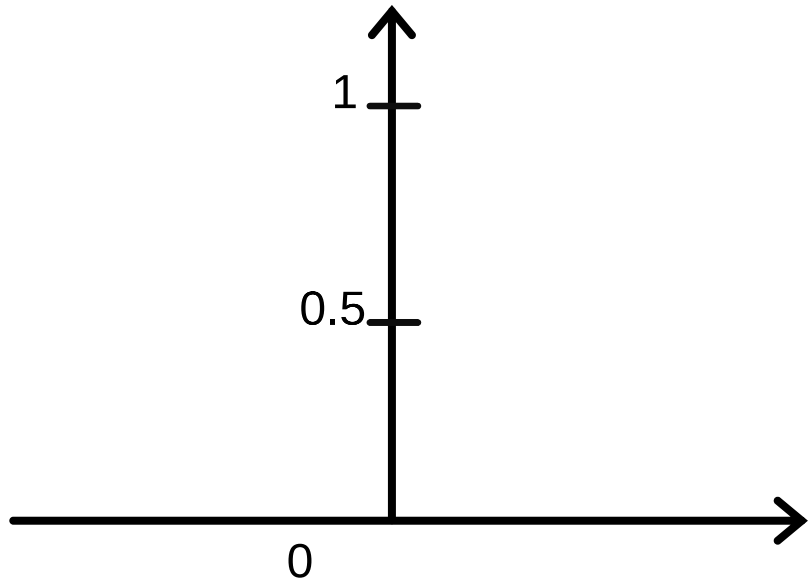
SIGMOID FUNCTION



LOGISTIC REGRESSION: HYPOTHESIS ?

Want $0 \leq h_\theta(x) \leq 1$

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



- Sigmoid function
  Logistic function

HYPOTHESIS OF LOGISTIC REGRESSION

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

$$z = \theta^T x$$

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

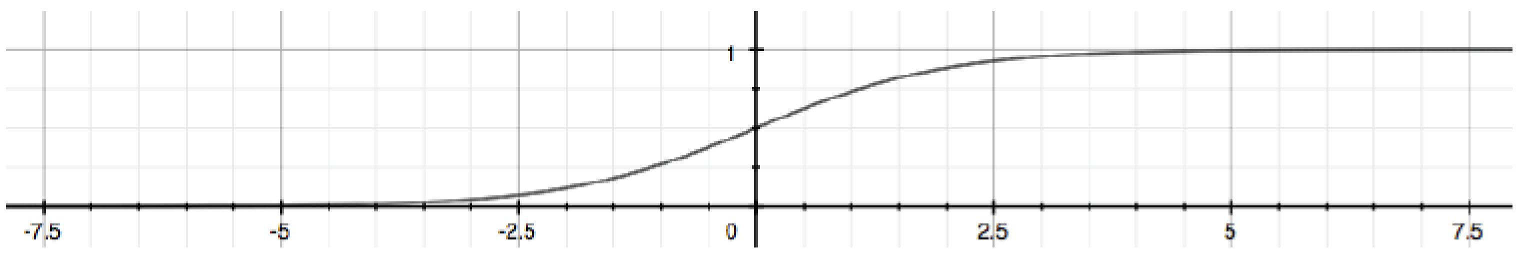
- Sigmoid function
  Logistic function

HYPOTHESIS INTERPRETATION

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

$$z = \theta^T x$$

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



hθ(x) = P(y = 1|x; θ)

"probability that y = 1, given x, parameterized by θ"

HYPOTHESIS OUTPUT

$h_\theta(x)$ = estimated probability that y = 1 on input x

Example: If $x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{tumorSize} \end{bmatrix}$

$$h_\theta(x) = 0.7$$

$h\theta(x) = P(y = 1|x; \theta)$

Tell patient that 70% chance of tumor being malignant

HYPOTHESIS INTERPRETATION:

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

$$z = \theta^T x$$

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

"probability that y = 1, given x, parameterized by θ"

$$\boxed{\begin{aligned} P(y=0|x;\theta) + P(y=1|x;\theta) = 1 \\ P(y=0|x;\theta) = 1 - P(y=1|x;\theta) \end{aligned}}$$

HYPOTHESIS INTERPRETATION:

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

$$z = \theta^T x$$

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

The probability that the prediction is 0 is just the complement of our probability that it is 1

if probability that it is 1 is 70%, then the probability that it is 0 is 30%).

$$\boxed{\begin{aligned} P(y=0|x;\theta) + P(y=1|x;\theta) = 1 \\ P(y=0|x;\theta) = 1 - P(y=1|x;\theta) \end{aligned}}$$

HOW TO MAKE A PREDICTION?

- Suppose θ_0 = -10.65 and θ_1 = 0.0055. What is the probability that a patient has a malignant tumor with the 1,000 tumor size?

HOW TO MAKE A PREDICTION?

- Suppose θ_0 = -10.65 and θ_1 = 0.0055. What is **the probability** that a patient has a malignant tumor with the 1,000 tumor size?

P(malignant = yes | tumor size= 1000; θ)

- Where θ is

$$\theta = \begin{matrix} -10.65 \\ 0.0055 \end{matrix}$$

DECISION BOUNDARY

ANOTHER WAY TO LOOK AT LOGISTIC REGRESSION HYPOTHESIS

CLASSIFICATION WITH LOGISTIC REGRESSION

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

$$z = \theta^T x$$

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

predict $y = 1$ if $h_\theta(x) \geq 0.5$

predict $y = 0$ if $h_\theta(x) < 0.5$

CLASSIFICATION WITH LOGISTIC REGRESSION

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

$$z = \theta^T x$$

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

$$h_\theta(x) \geq 0.5 \rightarrow y = 1$$
$$h_\theta(x) < 0.5 \rightarrow y = 0$$

Can we draw any direct relation between $\boldsymbol{\theta^T x}$ and output ?

CLASSIFICATION WITH LOGISTIC REGRESSION

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

$$z = \theta^T x$$

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

$$h_\theta(x) \geq 0.5 \rightarrow y = 1$$
$$h_\theta(x) < 0.5 \rightarrow y = 0$$

The way our logistic function **g(z)** behaves is that when its input is greater than or equal to zero, its output is greater than or equal to 0.5

CLASSIFICATION WITH LOGISTIC REGRESSION

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

$$z = \theta^T x$$

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

$$h_\theta(x) \geq 0.5 \rightarrow y = 1$$
$$h_\theta(x) < 0.5 \rightarrow y = 0$$

$$g(z) \geq 0.5$$
$$when\ z \geq 0$$

CLASSIFICATION WITH LOGISTIC REGRESSION

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

$$z = \theta^T x$$

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

$$h_\theta(x) \geq 0.5 \rightarrow y = 1$$
$$h_\theta(x) < 0.5 \rightarrow y = 0$$

$$g(z) \geq 0.5$$
$$when\ z \geq 0$$

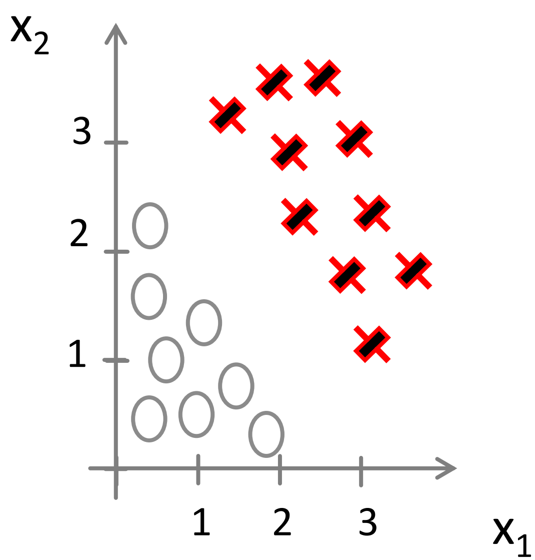
It means logistic regression predict y = 1 when $\theta^T x$ is greater than or equal to 0.5

CLASSIFICATION

- $\boldsymbol{\theta^T x}$ represent a line so we can draw a line using this equation.
- The values those are greater than the line ($\boldsymbol{\theta^T x \geq 0}$) should be classified as Positive Class.
- The values those are less than the line ($\boldsymbol{\theta^T x < 0}$) should be classified as Negative Class.

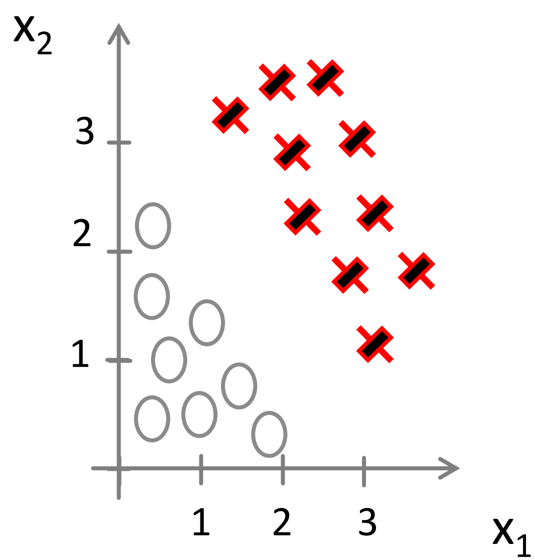
DECISION BOUNDARY

- The line $\theta^T x$ is also called decision boundary because this line help us to classify the example into positive or negative class.



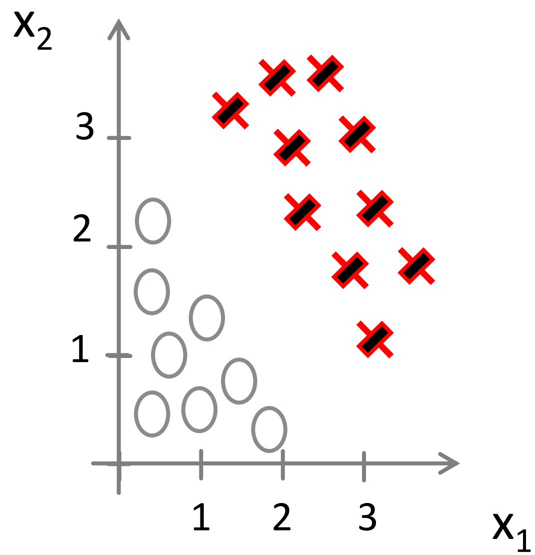
DECISION BOUNDARY: EXAMPLE

- Let $\theta_0 = -3, \theta_1 = 1, and\ \theta_2 = 1$
- Draw the decision boundary at figure while hypothesis is $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$



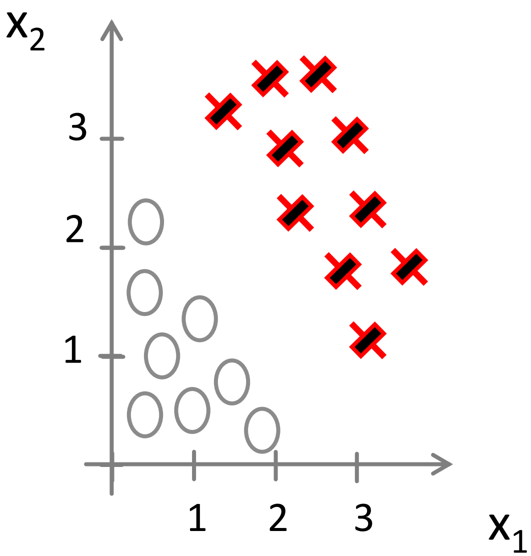
DECISION BOUNDARY: EXAMPLE

- Let $\theta_0 = -3, \theta_1 = 1, and\ \theta_2 = 1$
- Draw the decision boundary at figure while hypothesis is $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$



Predict $y = 1$ if $-3 + x_1 + x_2 \geq 0$

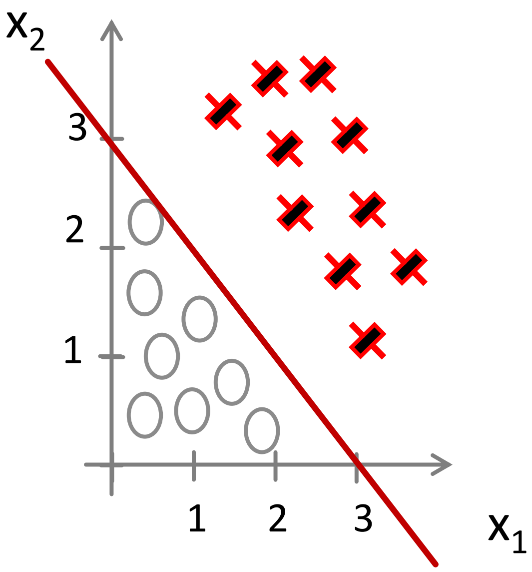
DECISION BOUNDARY ANOTHER EXAMPLE



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

Predict $y = 1$ if $-3 + x_1 + x_2 \geq 0$

DECISION BOUNDARY

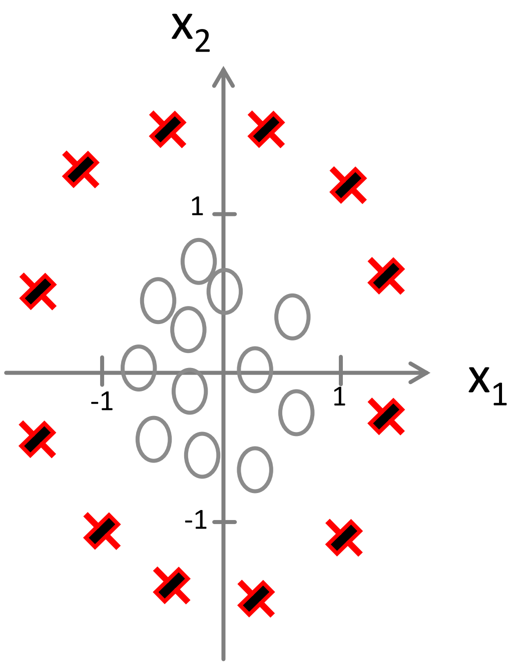


$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

Predict $y = 1$ if $-3 + x_1 + x_2 \geq 0$

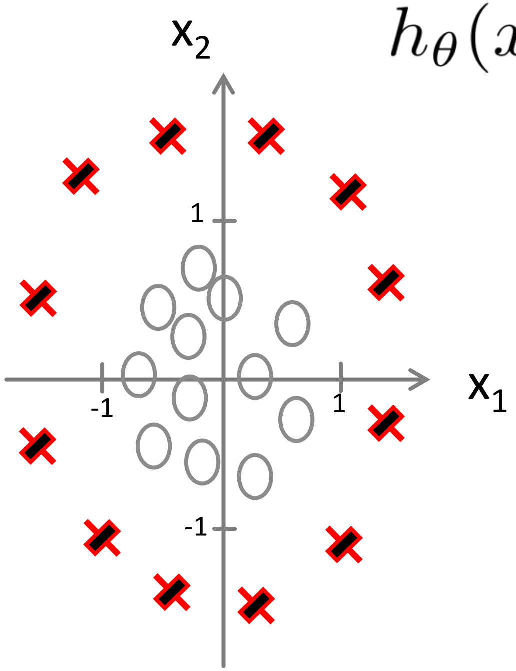
$$x_1 + x_2 \geq 3$$

COMPLEX DECISION BOUNDARY



- For non linear type of data, we can use polynomial Hypothesis instead of linear one.

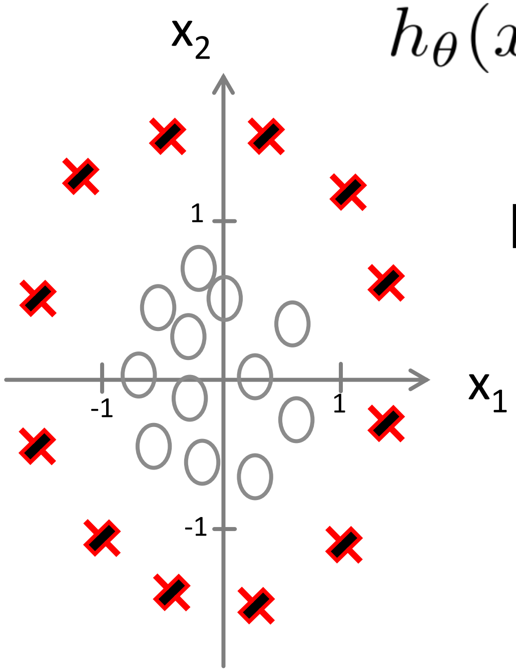
COMPLEX DECISION BOUNDARY



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

- For non linear type of data, we can use polynomial Hypothesis instead of linear one.

COMPLEX DECISION BOUNDARY



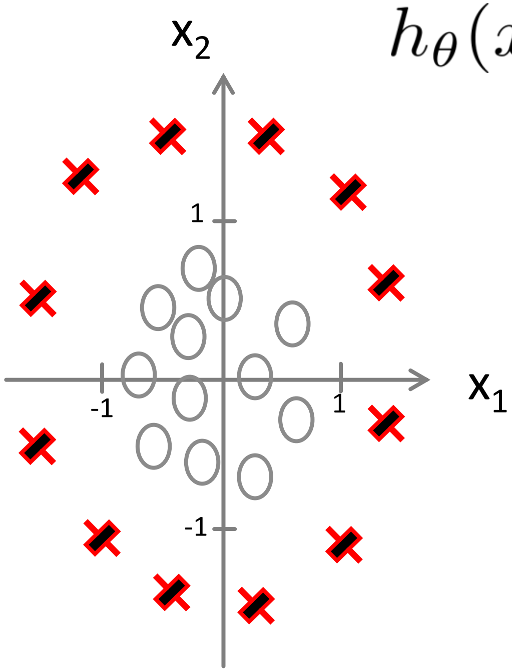
$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

Let $\theta_0 = -1, \theta_1 = 0, \theta_2 = 0, \theta_3 = 1, \theta_4 = 1$

Predict $y = 1$ if $-1 + x_1^2 + x_2^2 \geq 0$

- Draw the decision boundary

COMPLEX DECISION BOUNDARY



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

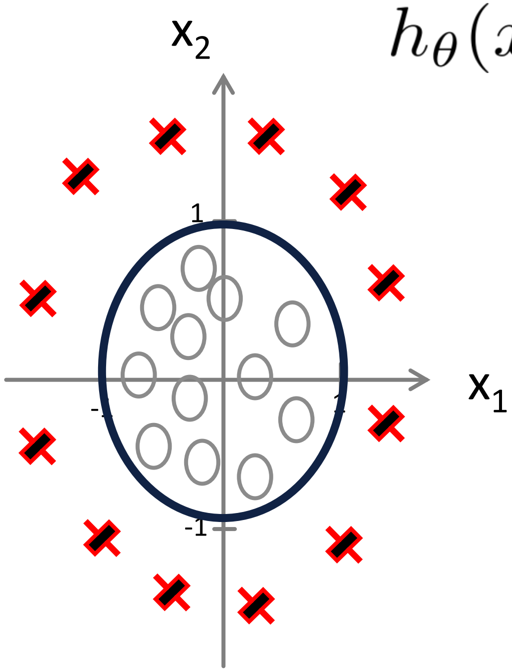
Let $\theta_0 = -1, \theta_1 = 0, \theta_2 = 0, \theta_3 = 1, \theta_4 = 1$

Predict y = 1 if $-1 + x_1^2 + x_2^2 \geq 0$

This is equation of circle $x_1{}^2 + x_2{}^2 \geq 1$

- Draw the decision boundary

COMPLEX DECISION BOUNDARY



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

Let $\theta_0 = -1, \theta_1 = 0, \theta_2 = 0, \theta_3 = 1, \theta_4 = 1$

Predict y = 1 if $-1 + x_1^2 + x_2^2 \geq 0$

This is equation of circle $x_1^{\,2} + x_2^{\,2} \geq 1$

- Possible decision boundary

COST FUNCTION

LOGISTIC REGRESSION

SELECTING PARAMETERS

Training set: $$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(m)}, y^{(m)})\}$$

m examples $$x \in \begin{bmatrix} x_0 \\ x_1 \\ \dots \\ x_n \end{bmatrix} \qquad x_0 = 1, y \in \{0, 1\}$$

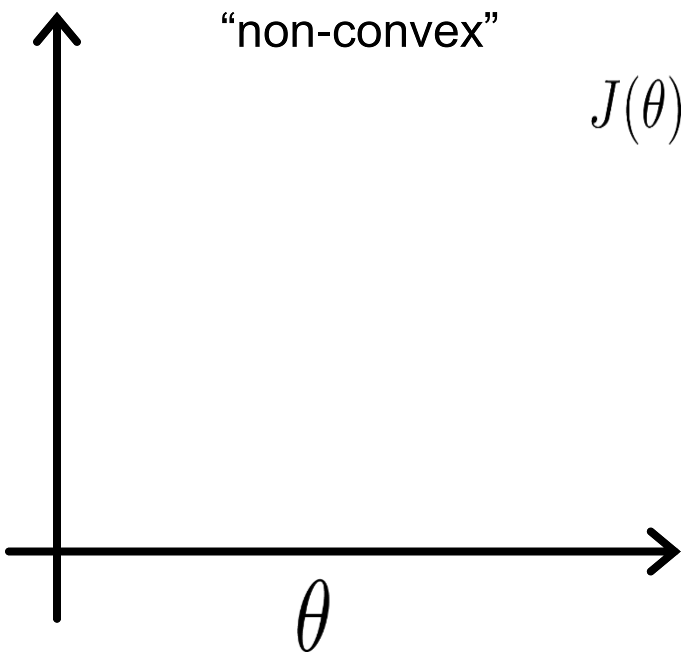
$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

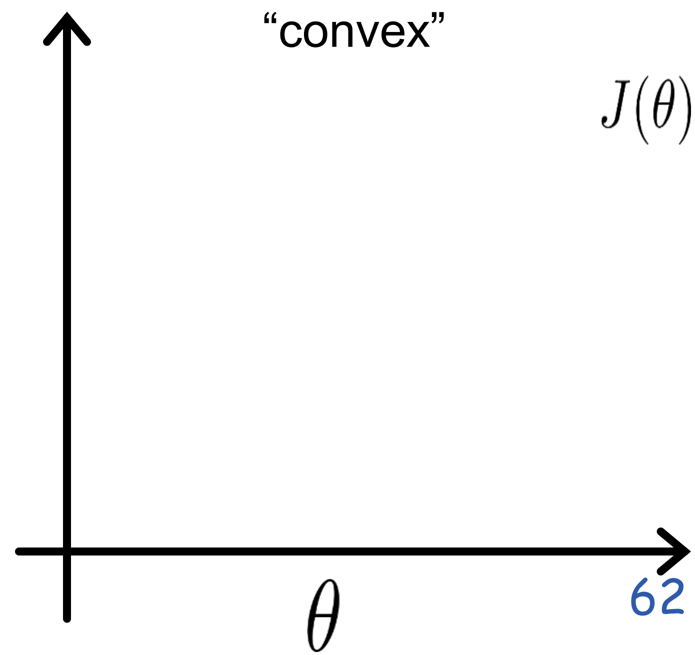
How to choose parameters θ ?

Cost function

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

$$\mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)}) = \frac{1}{2} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

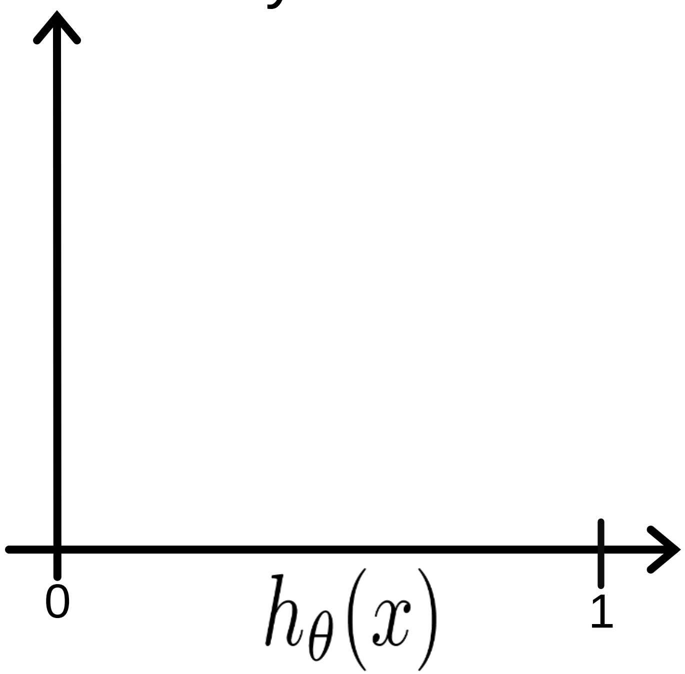




Logistic regression cost function

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

If y = 1



$\text{Cost} = 0 \text{ if } y = 1, h_\theta(x) = 1$

But as $h_\theta(x) \rightarrow 0$

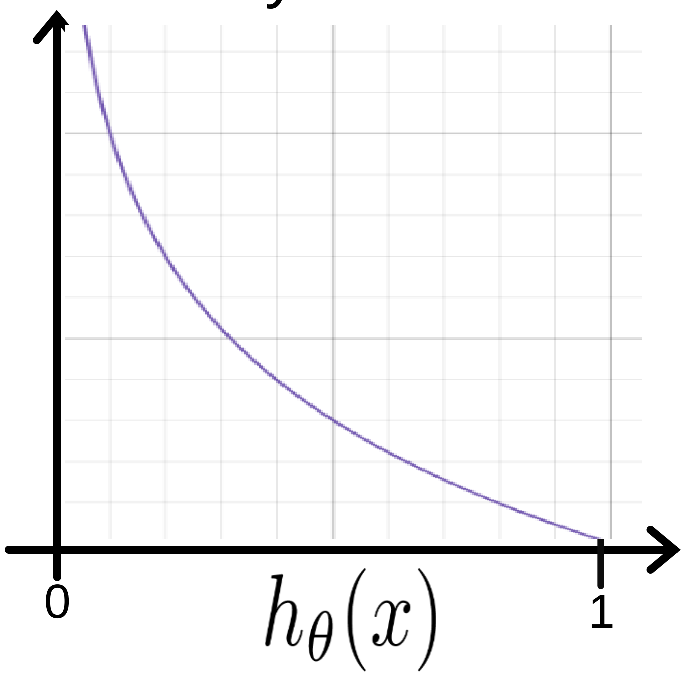
$Cost \rightarrow \infty$

Captures intuition that if $h_\theta(x) = 0$, (predict $P(y = 1|x; \theta) = 0$), but $y = 1$, we'll penalize learning algorithm by a very large cost.

Logistic regression cost function

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

If y = 1



$\text{Cost} = 0 \text{ if } y = 1, h_\theta(x) = 1$

But as $h_\theta(x) \to 0$

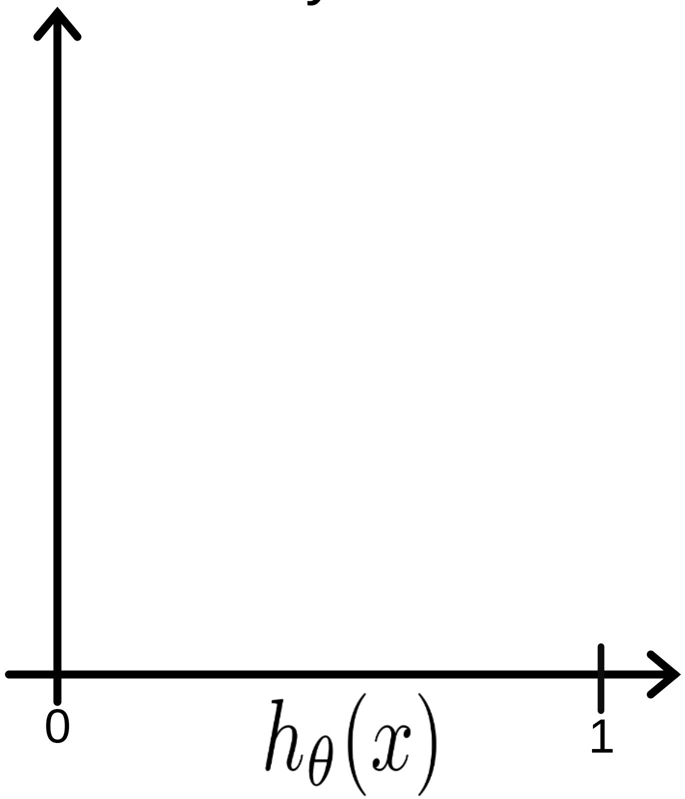
$Cost \to \infty$

Captures intuition that if $h_\theta(x) = 0$, (predict $P(y = 1|x; \theta) = 0$), but $y = 1$, we'll penalize learning algorithm by a very large cost.

Logistic regression cost function

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

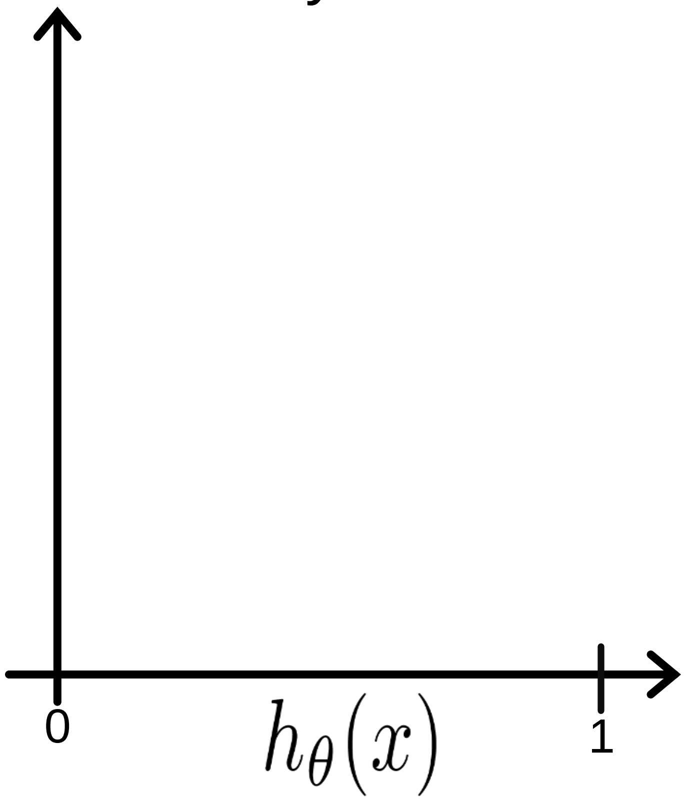
If y = 0



Logistic regression cost function

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

If y = 0



FINDING THE BEST PARAMETERS

SIMPLE COST FUNCTION GRADIENT DESCENT

LOGISTIC REGRESSION COST FUNCTION

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

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

Note: $y = 0$ or 1 always

- Can we write this in term of single function so we can easily use t. ?

LOGISTIC REGRESSION COST FUNCTION

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

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

Note: $y = 0$ or 1 always

- Can we write two conditional statements in term of one statement ?

LOGISTIC REGRESSION COST FUNCTION

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

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

Note: $y = 0$ or 1 always

$$\mathrm{Cost}(h_\theta(x), y) = -y \log(h_\theta(x)) - (1 - y)\log(1 - h_\theta(x))$$

LOGISTIC REGRESSION COST FUNCTION

Hypothesis $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$

parameters : θ

Cost Function

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

$$= -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)}\log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))]$$

Goal: $\min_\theta J(\theta)$

GRADIENT DESCENT

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_\theta(x^{(i)}))]$$

Want $\min_\theta J(\theta)$:

Repeat $\{$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

$\}$ (simultaneously update all θ_j)

PARTIAL DERIVATIVE OF SIGMOID

- Before finding the partial derivative of $J(\theta)$,first we find derivative of sigmoid function which is required to find partial derivative of $J(\theta)$

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

$$= \frac{-(1+e^{-x})'}{(1+e^{-x})^2}$$

$$= \frac{-1' - (e^{-x})'}{(1+e^{-x})^2}$$

$$= \frac{0 - (-x)'(e^{-x})}{(1+e^{-x})^2}$$

$$= \frac{-(-1)(e^{-x})}{(1+e^{-x})^2}$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

FINDING PARTIAL OF SIGMOID

- Before finding the partial derivative of $J(\theta)$,first we find derivative of sigmoid function which is required to find partial derivative of $J(\theta)$

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

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

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

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

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

$$= \sigma(x)(1-\sigma(x))$$

FINDING PARTIAL OF SIGMOID

- Writing partial derivative of sigmoid in more useful form

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

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

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

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

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

$$= \sigma(x)(1-\sigma(x))$$

PARTIAL DERIVATIVE OF COST FUNCTION

$$\frac{\partial}{\partial\theta_j}J(\theta) = \frac{\partial}{\partial\theta_j}\frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}log(h_\theta(x^{(i)})) + (1-y^{(i)})log(1-h_\theta(x^{(i)}))\right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\frac{\partial}{\partial\theta_j}log(h_\theta(x^{(i)})) + (1-y^{(i)})\frac{\partial}{\partial\theta_j}log(1-h_\theta(x^{(i)}))\right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[\frac{y^{(i)}\frac{\partial}{\partial\theta_j}h_\theta(x^{(i)})}{h_\theta(x^{(i)})} + \frac{(1-y^{(i)})\frac{\partial}{\partial\theta_j}(1-h_\theta(x^{(i)}))}{1-h_\theta(x^{(i)})}\right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[\frac{y^{(i)}\frac{\partial}{\partial\theta_j}\sigma(\theta^T x^{(i)})}{h_\theta(x^{(i)})} + \frac{(1-y^{(i)})\frac{\partial}{\partial\theta_j}(1-\sigma(\theta^T x^{(i)}))}{1-h_\theta(x^{(i)})}\right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[\frac{y^{(i)}\sigma(\theta^T x^{(i)})(1-\sigma(\theta^T x^{(i)}))\frac{\partial}{\partial\theta_j}\theta^T x^{(i)}}{h_\theta(x^{(i)})} + \frac{-(1-y^{(i)})\sigma(\theta^T x^{(i)})(1-\sigma(\theta^T x^{(i)}))\frac{\partial}{\partial\theta_j}\theta^T x^{(i)}}{1-h_\theta(x^{(i)})}\right]$$

PARTIAL DERIVATIVE OF COST FUNCTION

$$\frac{\partial}{\partial\theta_j}J(\theta) = \frac{\partial}{\partial\theta_j}\frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}log(h_\theta(x^{(i)})) + (1-y^{(i)})log(1-h_\theta(x^{(i)}))\right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[\frac{y^{(i)}h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))\frac{\partial}{\partial\theta_j}\theta^T x^{(i)}}{h_\theta(x^{(i)})} - \frac{(1-y^{(i)})h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))\frac{\partial}{\partial\theta_j}\theta^T x^{(i)}}{1-h_\theta(x^{(i)})}\right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}(1-h_\theta(x^{(i)}))x_j^{(i)} - (1-y^{(i)})h_\theta(x^{(i)})x_j^{(i)}\right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}(1-h_\theta(x^{(i)})) - (1-y^{(i)})h_\theta(x^{(i)})\right]x_j^{(i)}$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} - h_\theta(x^{(i)})\right]x_j^{(i)}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left[h_\theta(x^{(i)}) - y^{(i)}\right]x_j^{(i)}$$

GRADIENT DESCENT

$$
\begin{aligned}
&Repeat\ \{\\
&\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)\\
&\}
\end{aligned}
$$

- Partial Derivative of the cost function is

$$= \frac{1}{m} \sum_{i=1}^{m} \left[h_\theta(x^{(i)}) - y^{(i)} \right] x_j^{(i)}$$

GRADIENT DESCENT

$$Repeat \; \{$$
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$
$$\}$$

Partial Derivative of the cost function

$$= \frac{1}{m} \sum_{i=1}^{m} \left[h_\theta(x^{(i)}) - y^{(i)} \right] x_j^{(i)}$$

- Finally, the Gradient Descent is

$$Repeat \; \{$$
$$\theta_j := \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$
$$\}$$

GRADIENT DESCENT

$$Repeat \; \{$$
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$
$$\}$$

Partial Derivative of the cost function

$$= \frac{1}{m} \sum_{i=1}^{m} \left[h_\theta(x^{(i)}) - y^{(i)} \right] x_j^{(i)}$$

- Finally, the Gradient Descent is

$$Repeat \; \{$$
$$\theta_j := \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$
$$\}$$

GRADIENT DESCENT

$$Repeat\ \{$$
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$
$$\}$$

Partial Derivative of the cost function

$$= \frac{1}{m} \sum_{i=1}^{m} \left[h_\theta(x^{(i)}) - y^{(i)} \right] x_j^{(i)}$$

- Finally, the Gradient Descent is

$$Repeat\ \{$$
$$\theta_j := \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$
$$\}$$

SAME as of Linear Regression

GRADIENT DESCENT

$$Repeat\ \{$$
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$
$$\}$$

Partial Derivative of the cost function

$$= \frac{1}{m} \sum_{i=1}^{m} \left[h_\theta(x^{(i)}) - y^{(i)} \right] x_j^{(i)}$$

- Finally, the Gradient Descent is

$$Repeat\ \{$$
$$\theta_j := \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$
$$\}$$

Only h(θ) is different

GRADIENT DESCENT

$$Repeat\ \{$$
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$
$$\}$$

Partial Derivative of the cost function

$$= \frac{1}{m} \sum_{i=1}^{m} \left[h_\theta(x^{(i)}) - y^{(i)} \right] x_j^{(i)}$$

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

- Finally, the Gradient Descent is

$$Repeat\ \{$$
$$\theta_j := \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$
$$\}$$

Only h(θ) is different

PROGRAMMING ASSIGNMENT 09

- Convert the linear regression program into Logistic Regression Program for Binary Class Classification Problems.