UNIT 4:
Supervised Learning

Prepared by Nima Dema

Regression

Question time

1. What is supervised learning? Why do we called it a supervised learning?

1. Given a problem statement, how do you decide whether it is a classification problem or a regression problem?

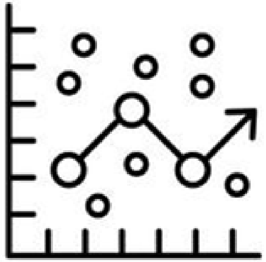
Regression vs Classification problem

- Supervised learning is used whenever we want to predict a certain outcome from a given input, and we have examples of input/output pairs.

- Our goal is to make accurate predictions for new, never-before-seen data. Supervised learning often requires human effort to build the training set.

Two major types of supervised learning:

- **Classification**: A classification problem is when the output variable is a category, such as "red" or "blue" or "disease" and "no disease".

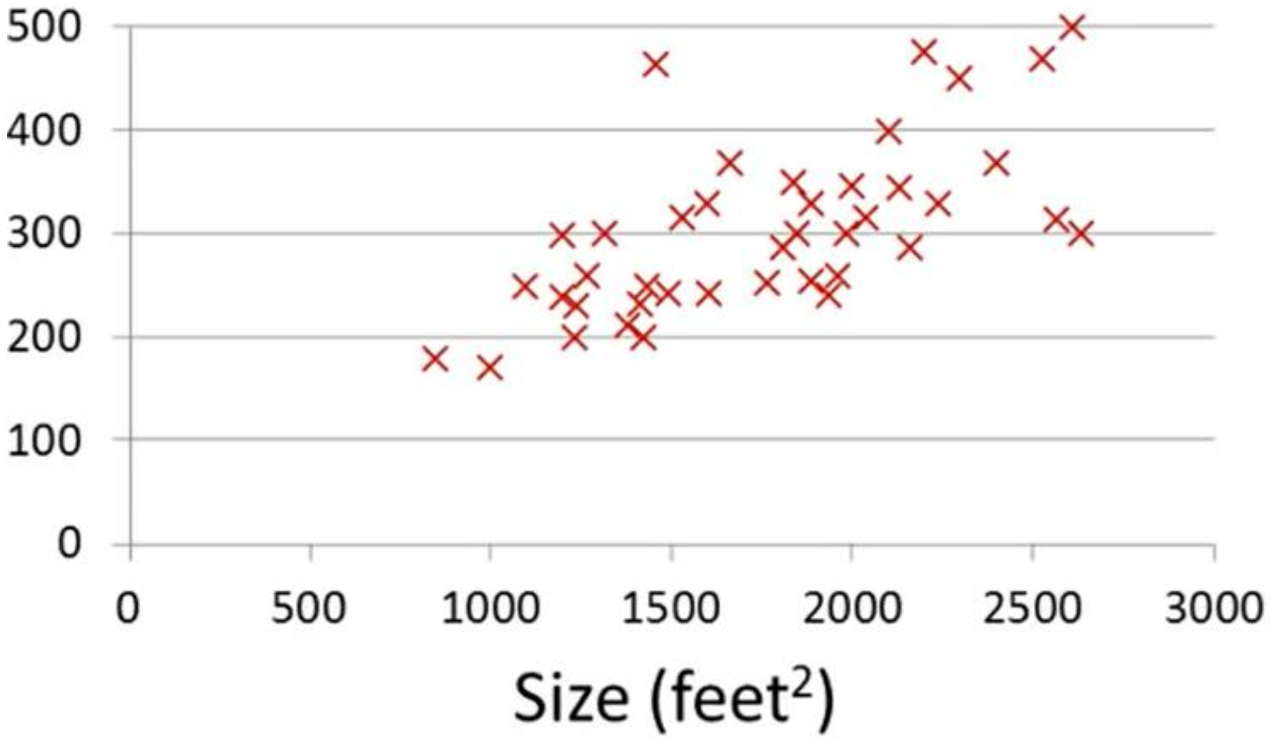


- **Regression**: A regression problem is when the output variable is a real value, such as "dollars" or "weight".

Regression

Regression searches for relationships among variables.

1. Linear Regression:



Linear Regression – Simple Linear Regression

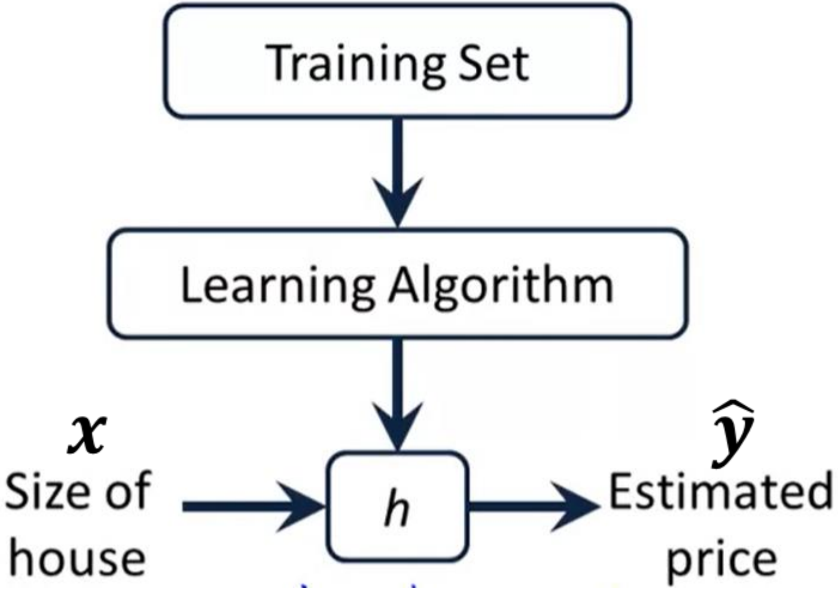
Housing Price prediction problem.

Notation:

- n: number of training examples
- X's: input variables/features
- y's: output variable/target variable

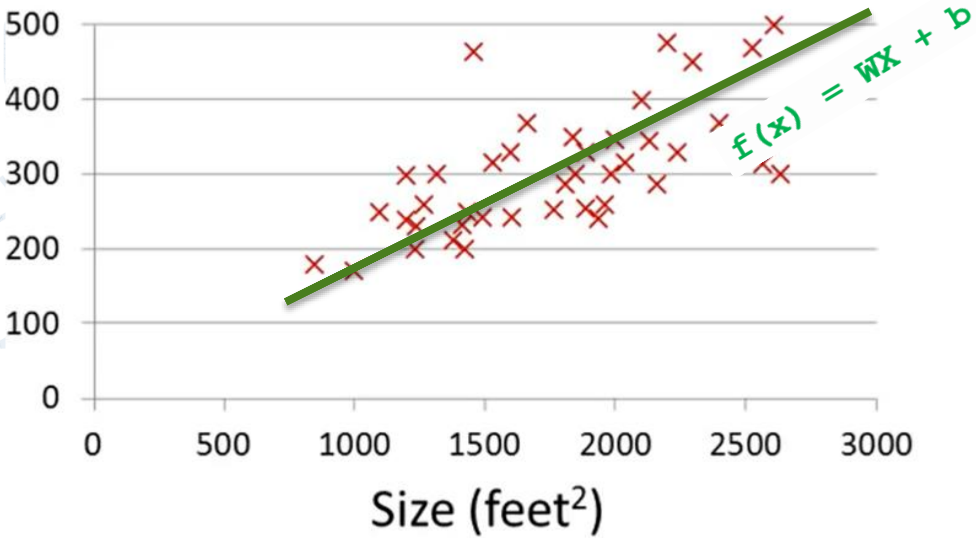
Size in feet^2 (x)	Price ($) in 1000's (y)
2104	460
1416	232
1534	315
852	178
...	...

Linear Regression – Simple Linear Regression



Representation of **h**

$$h \Rightarrow f(x) = y\_pred = wx + b$$



Linear Regression – Simple Linear Regression

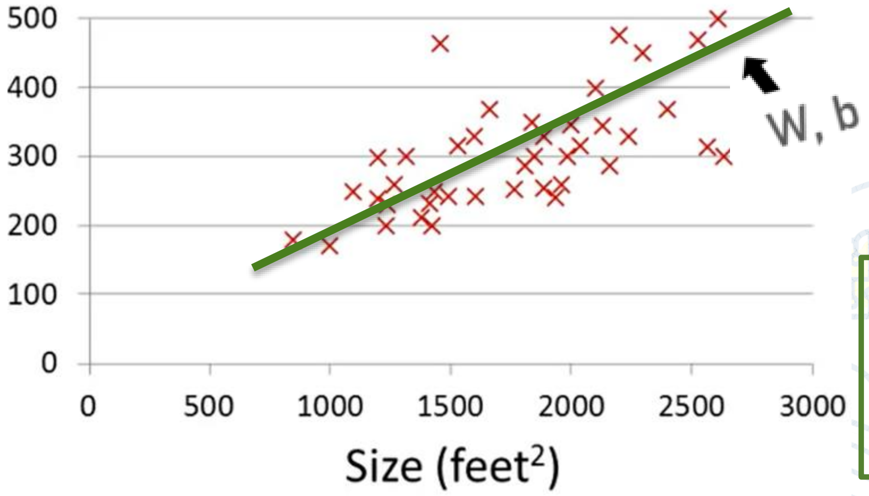
Training Set:

$$\boxed{f(x) = w\,x + b}$$

Size in feet^2 (x)	Price (\$) in 1000's (y)
2104	460
1416	232
1534	315
852	178
...	...

- **w** and **b** are the parameters or coefficients or weights
- How to choose **w** and **b**?

Linear Regression – Simple Linear Regression

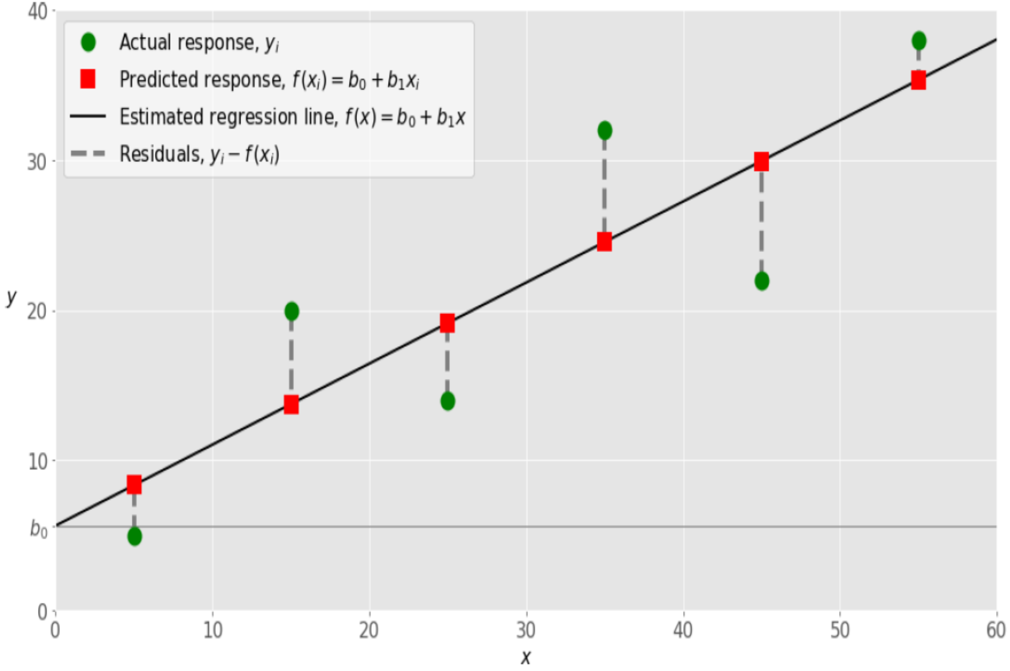


Minimize $\underset{w,\,b}{}$ $\frac{1}{n}\sum_{i=1}^{n}\left(f(x^i) - y^i\right)^2$

$$J(w,b) = MSE = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(f(x^i) - y^i\right)^2}_{\text{Mean Squared Error/ Cost Function/ Squared Error Function}}$$

Idea/Goal: Choose w and b so that f(x) is close to y for our training examples (x,y)

Linear Regression – Simple Linear Regression



$$\frac{1}{n}\sum_{i=1}^{n}\left(f\left(x^i\right) - y^i\right)^2$$

Find Parameters using Normal Equation

$$= \left(\mathbf{X}^T \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}^T \cdot \mathbf{y}$$

Normal Equation to learn parameters

Examples: $n = 4.$

x_0	Size (feet²) x_1	Number of bedrooms x_2	Number of floors x_3	Age of home (years) x_4	Price ($1000) y
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix} \qquad y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

$m \times (n+1)$ -dimensional vector

$$\boxed{\theta = (X^T X)^{-1} X^T y} \leftarrow \quad \text{Where } \theta = [w,b]$$

GYALPOZHING
COLLEGE OF INFORMATION TECHNOLOGY

Multiple Linear Regression

	Size in feet^2 x_1	Number of bedrooms x_2	Number of floors x_3	Age of home in years x_4	Price ($) in $1000's
	2104	5	1	45	460
$i=2$	1416	3	2	40	232
	1534	3	2	30	315
	852	2	1	36	178
	...	...	...	...	...

Multiple Linear Regression Model

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

Function-> $\boxed{f(x) = w_1x_1+w_2x_2+w_3x_3+w_4x_4+b}$

Parameters : $w_1,\ w_2\ \dots\ w_r$ where r is the number of features.

Cost Function: $J(w_1,\ w_2\ \dots\ w_r, b)$

For multiple linear regression, goal is to choose $w_1,w_2 \dots w_r$ such that the cost function or the MSE is minimum

Classification

Classification Problem

Problem

- Email: Is this email spam?
- Online Transactions: Fraudulent?
- Is the tumor Malignant?

Answer (y)

YES NO
YES NO
YES NO

y can only be one of two values

"binary classification"

Class = Category

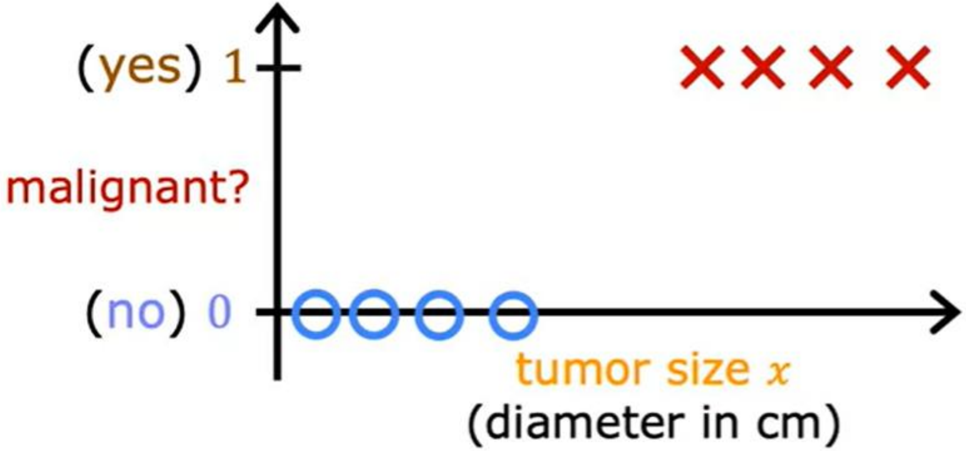
True False

1 0

"positive class"
!= "good"
Presence

"negative class"
!="bad"
Absence

Binary Classification

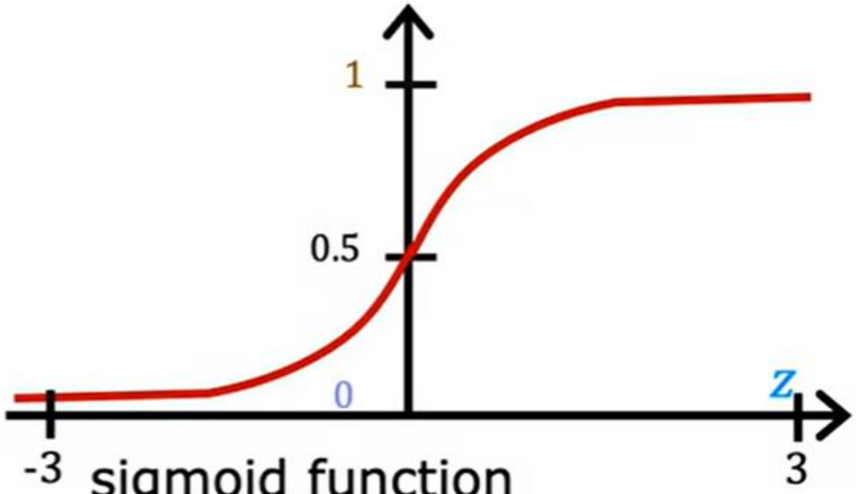


- Supposed you use Linear Regression for Classification problem with threshold of 0.5,
- If f(x) > 0.5, y = 1
- If f(x) < 0.5, y = 0

- Linear Regression may not get the best result.
- Classification should result y = 0 or y = 1 but f(x) can result value greater than 1 or even smaller than 0.

Binary Classification – Logistic Regression

Want outputs between 0 and 1



sigmoid function

logistic function

outputs between 0 and 1

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

f(x) = g(z) = g(wx+b)

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

"logistic regression"

Interpretation of Logistic Regression output

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

f(x) = estimated probability that y = 1 on input x

Example:

x is the ""tumor size"

Y is 0 (not malignant)

or 1 (malignant)

If f(x) = 0.7

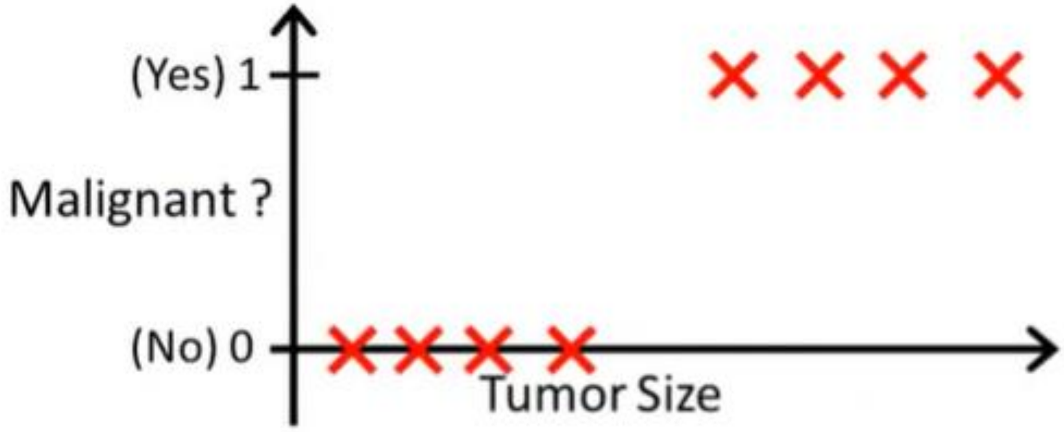
Tell patient that 70% chance of tumor being
malignant (70% chance that y is 1)

What is the chance of patient's tumor being not malignant? Or what is the chance of y = 0?

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

Decision Boundary

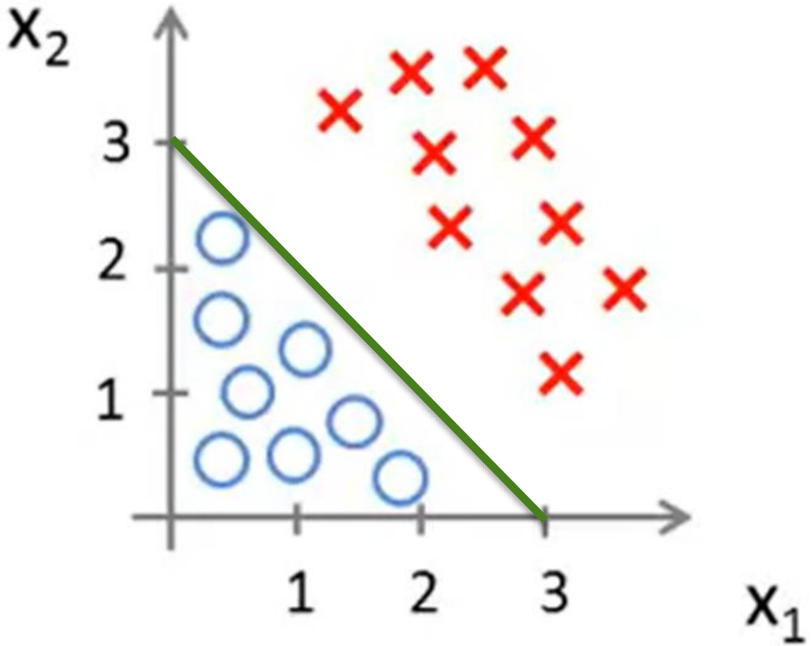
- The **decision boundary** is the line that separates the area where y = 0 and where y = 1. It is created by our function or model.

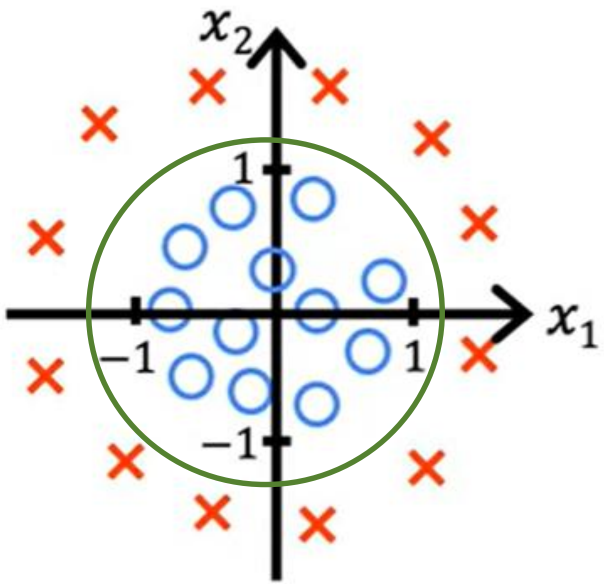


For example, b = 5, w1 = -1

$$Y = 1 \text{ if } w_1x_1+b \geq 0$$
$$\Rightarrow -1(x_1) + 5 \geq 0$$
$$-x_1 \geq -5$$
$$x_1 \geq 5$$

Decision Boundary





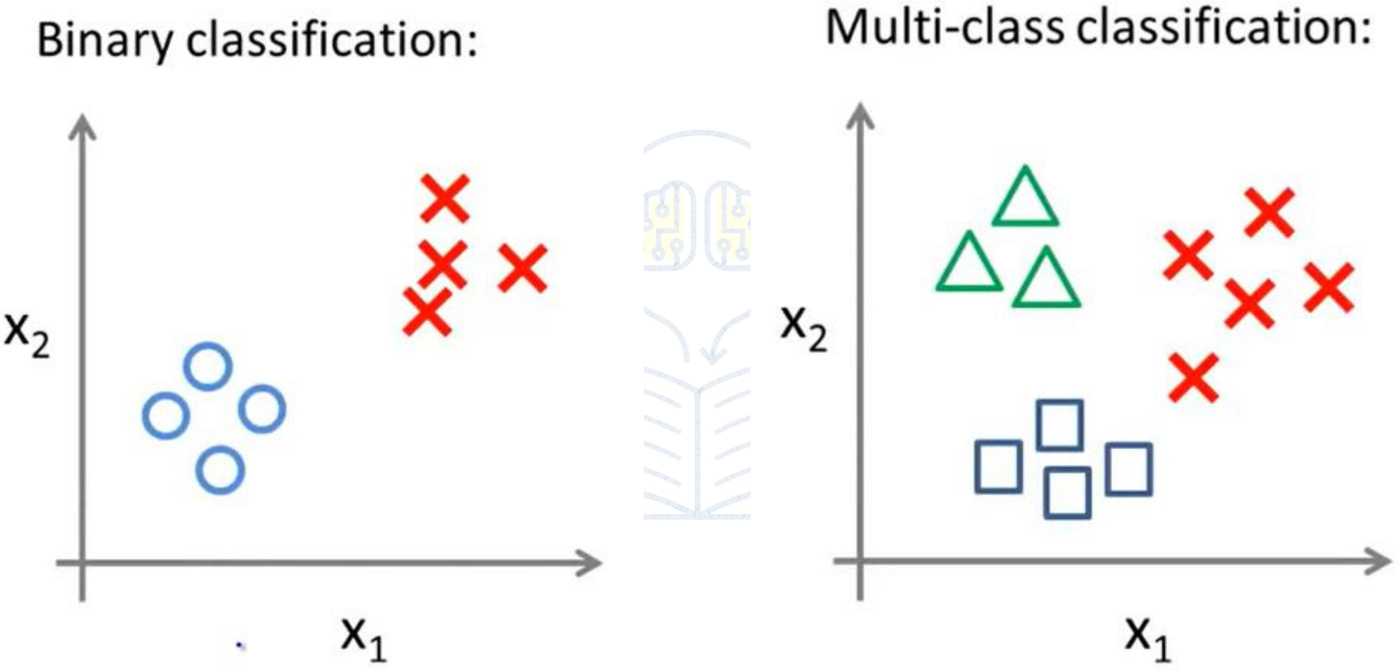
Multiclass Classification – Logistic Regression

- Email foldering/tagging: Work, friends, Family, Hobby

- Medical diagrams: Not ill, cold, Flu

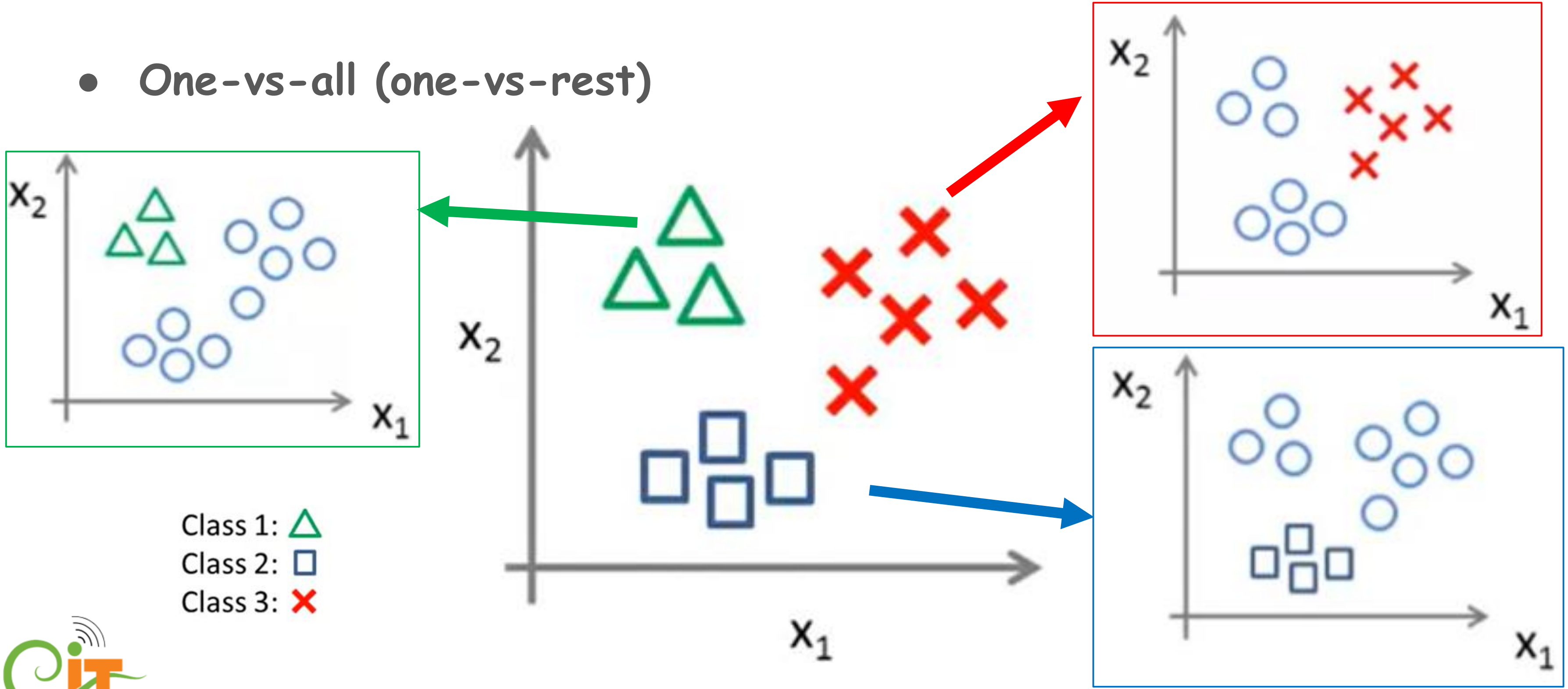
- Weather: Sunny, Cloudy, Rain, Snow

Multiclass Classification – Logistic Regression



Multiclass Classification – Logistic Regression

- *One-vs-all (one-vs-rest)*



Class Exercise

-Instructions.

1. Open Practical 5 Feature engineering lab notebook. Use your alldf Dataframe to create Logistic Regression Model.

1. Make Prediction for few example.

1. Use accuracy score model evaluation metrics to check performance of the model.

THANK YOU ☺