Machine Learning

Apple

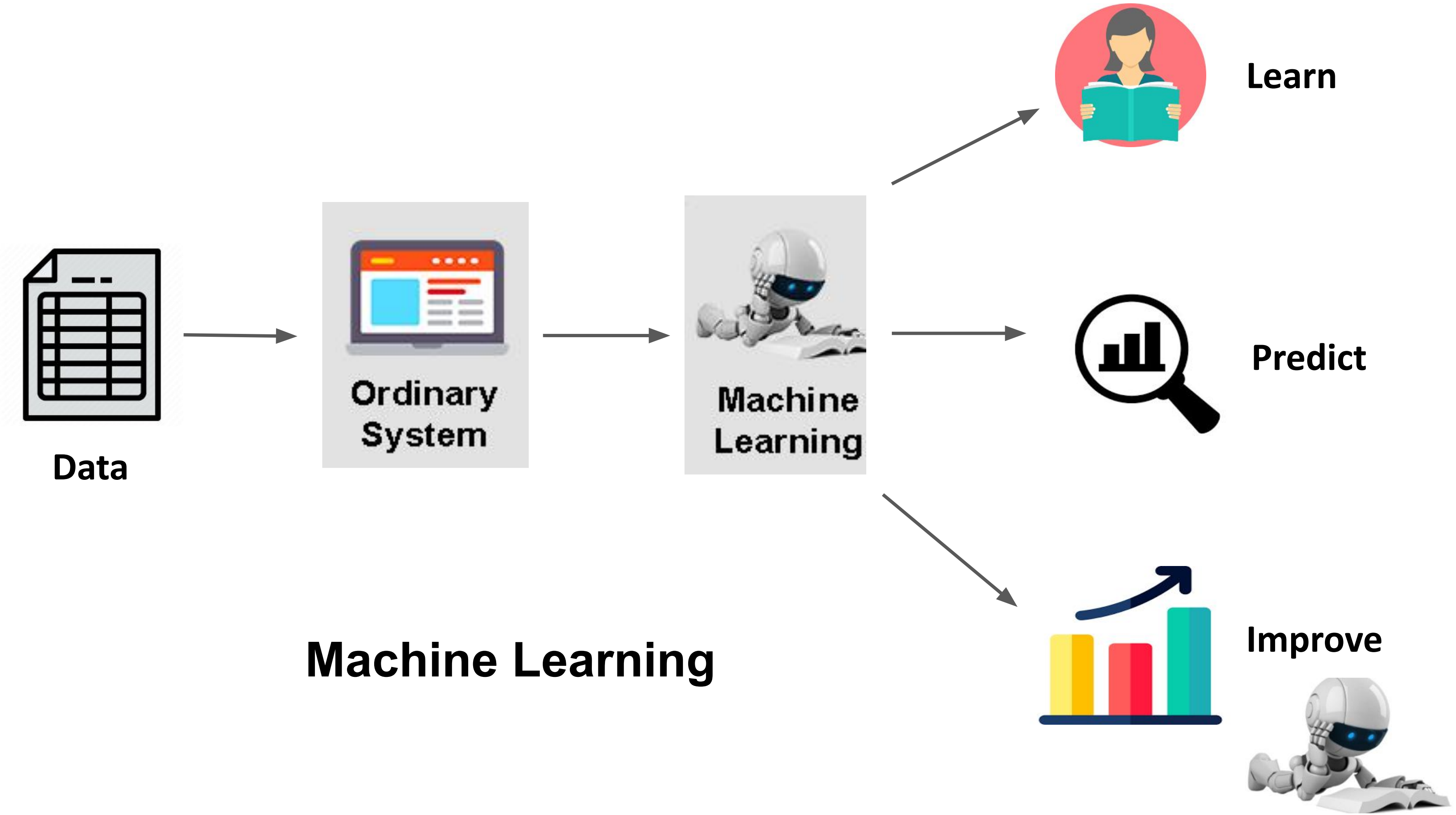
Human

Learn from experience

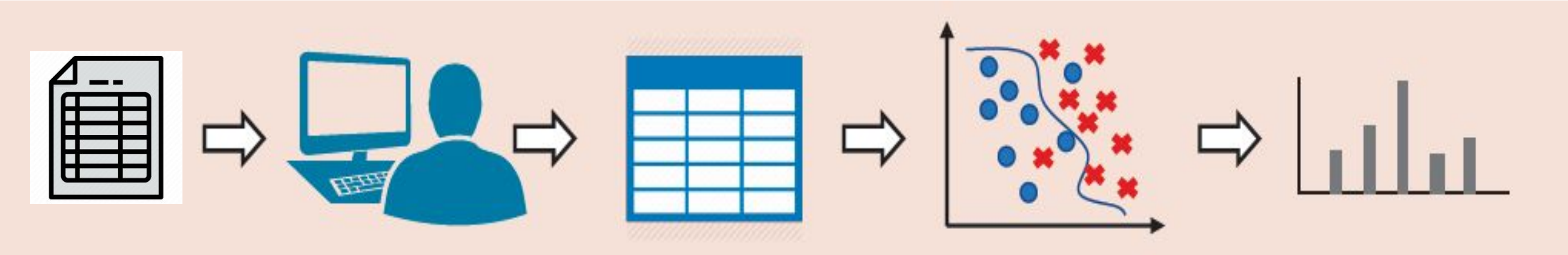
Machine

Learn from past data

Machine Learning


Data
Ordinary System
Machine Learning
Learn
Predict
Improve
Machine Learning

How Machine Learning ?



Data   Feature Extraction   Feature Selection   ML Algorithms   Output

Why Machine Learning?

- Rapid increment in the production of data
- Solving complex problems, which are difficult for a human
- Decision making in various sector including finance
- Finding hidden patterns and extracting useful information from data

Definition:

Field of Study that gives computers the ability to learn without being explicitly programmed.

Formal Defn:

A computer Program is said to learn from experience E w.r.t some task T and some performance measures P, if its performance on T as measured by P improves with experience E

Machine Learning Algorithms:

- Supervised Learning
- Unsupervised learning

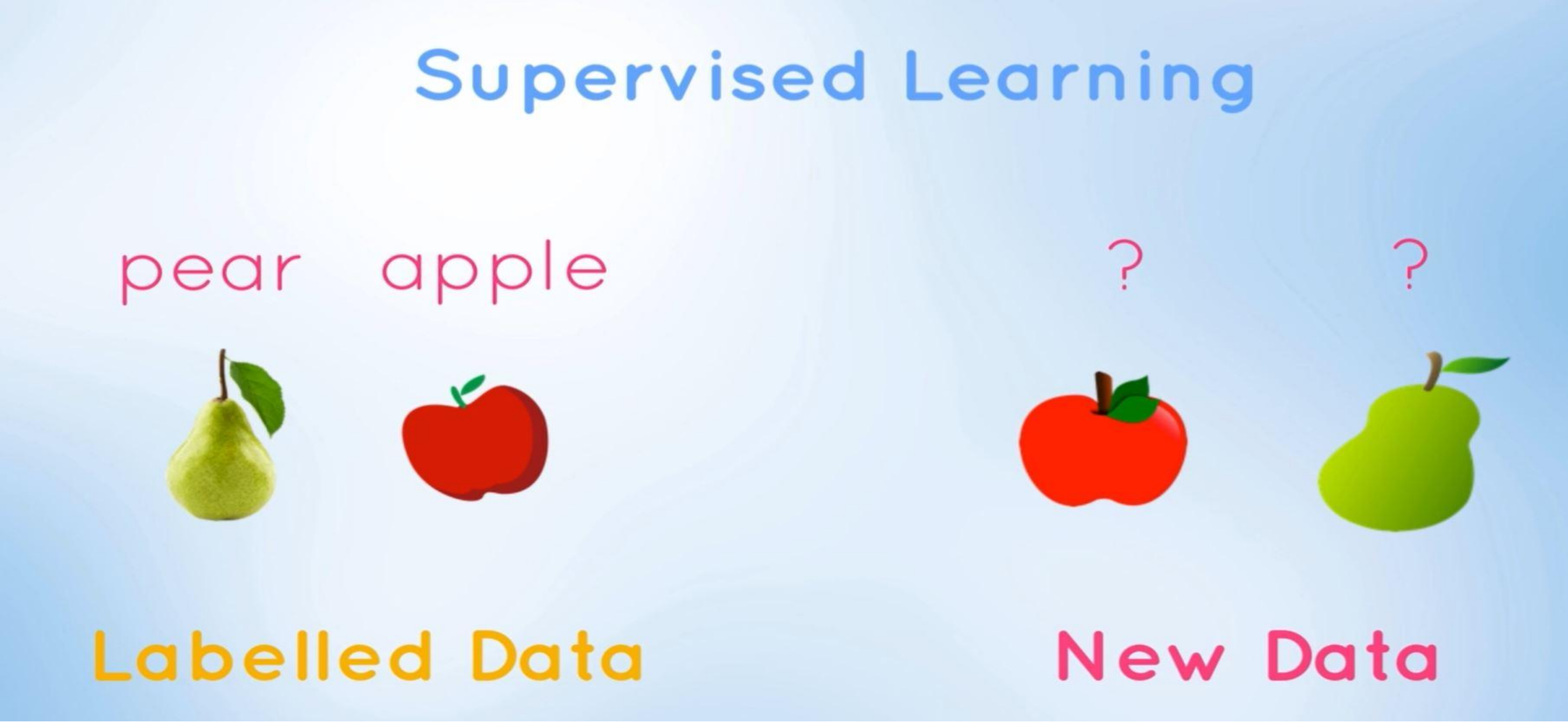
Others: Reinforcement Learning, Recommender Systems

Supervised Learning

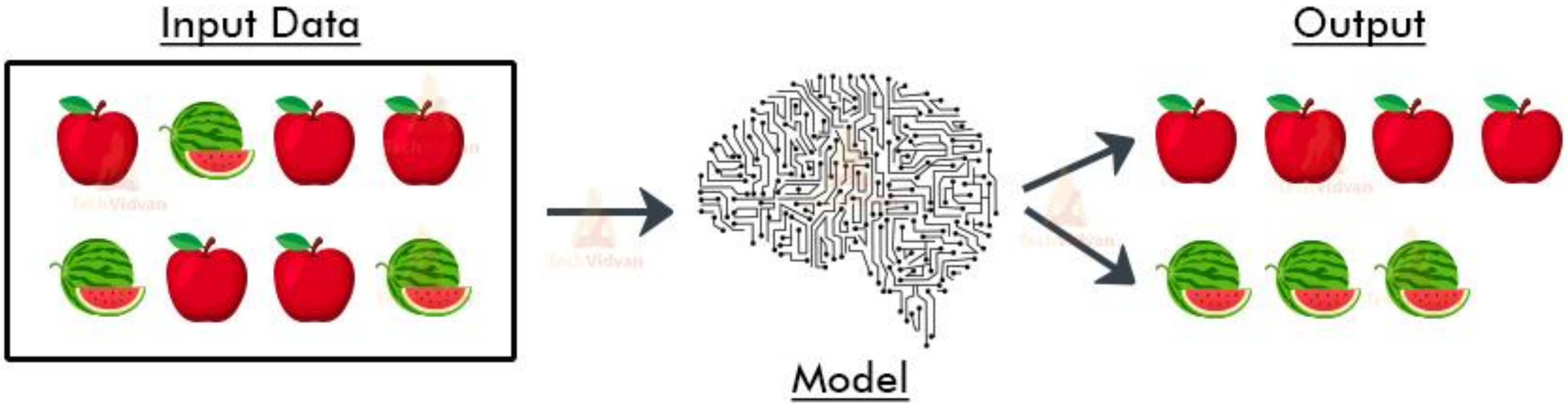
	1	2	3		?	?	?	
input	1	2	3		1	2	3	input

Training Data

New Data

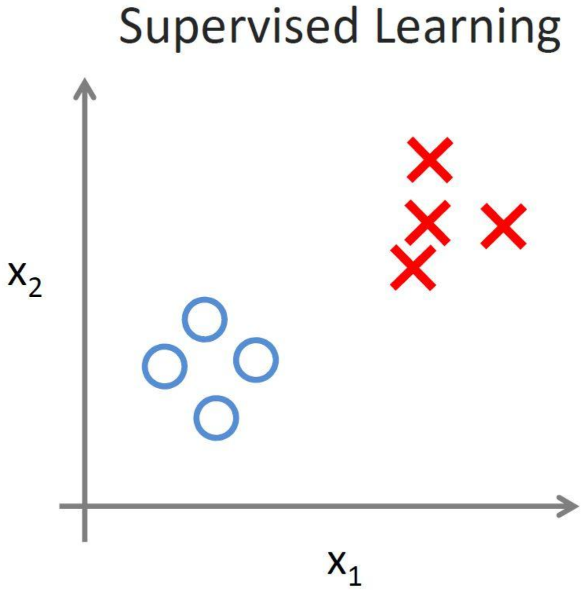

Supervised Learning
pear
apple
?
?
Labelled Data
New Data

Unsupervised Learning in ML



Supervised Learning:

- Learns an Input and O/P map.
    - Classification: categorical O/P
    - Regression: Continuous O/P

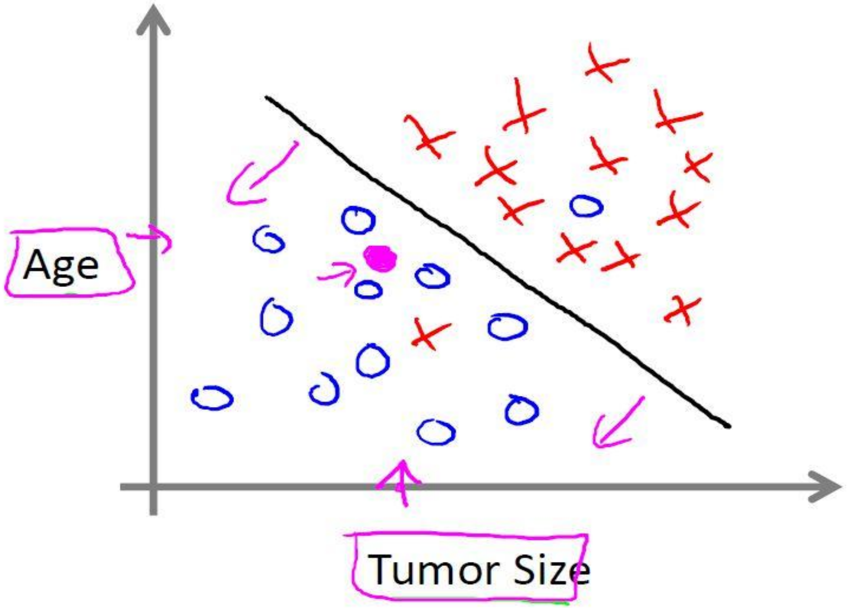


Regression:

Given huge dataset, when the machine is given several inputs it finds the right answer accordingly.

This terminology is also Known as Regression

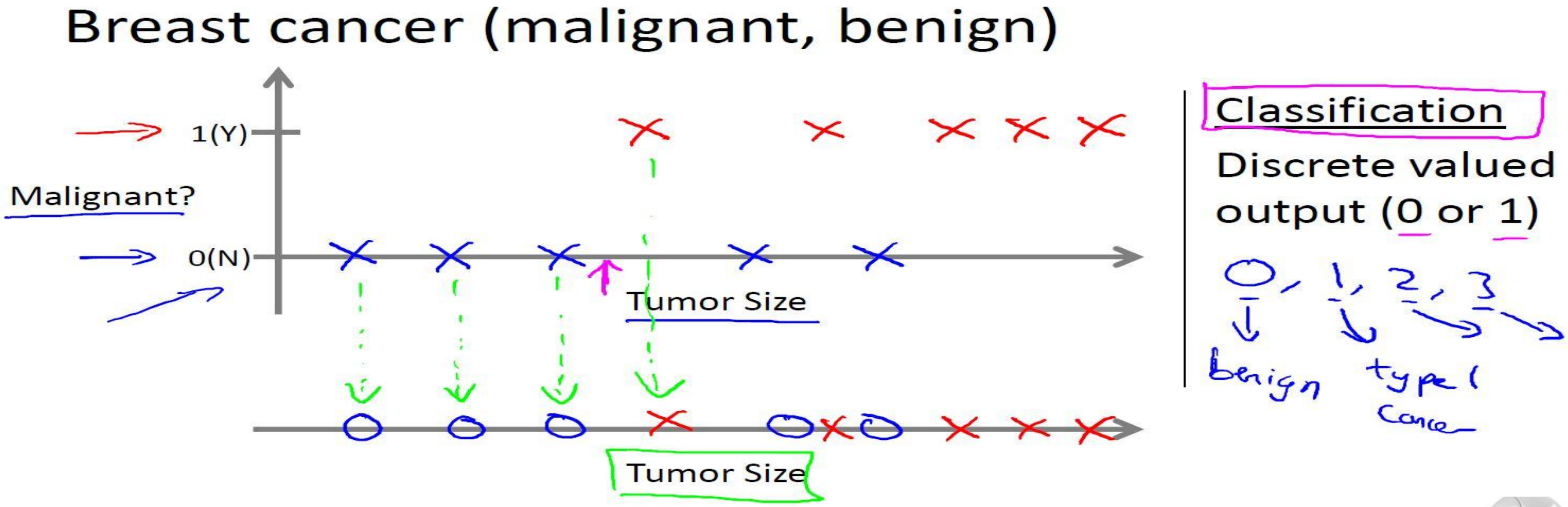
Regression: Predict Continuous valued output (price)



- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

...

Classification:



Identify:

For infinite feature vectors we need some mathematical tool for access.(e.g in vector machine)

#identical - Regression problem

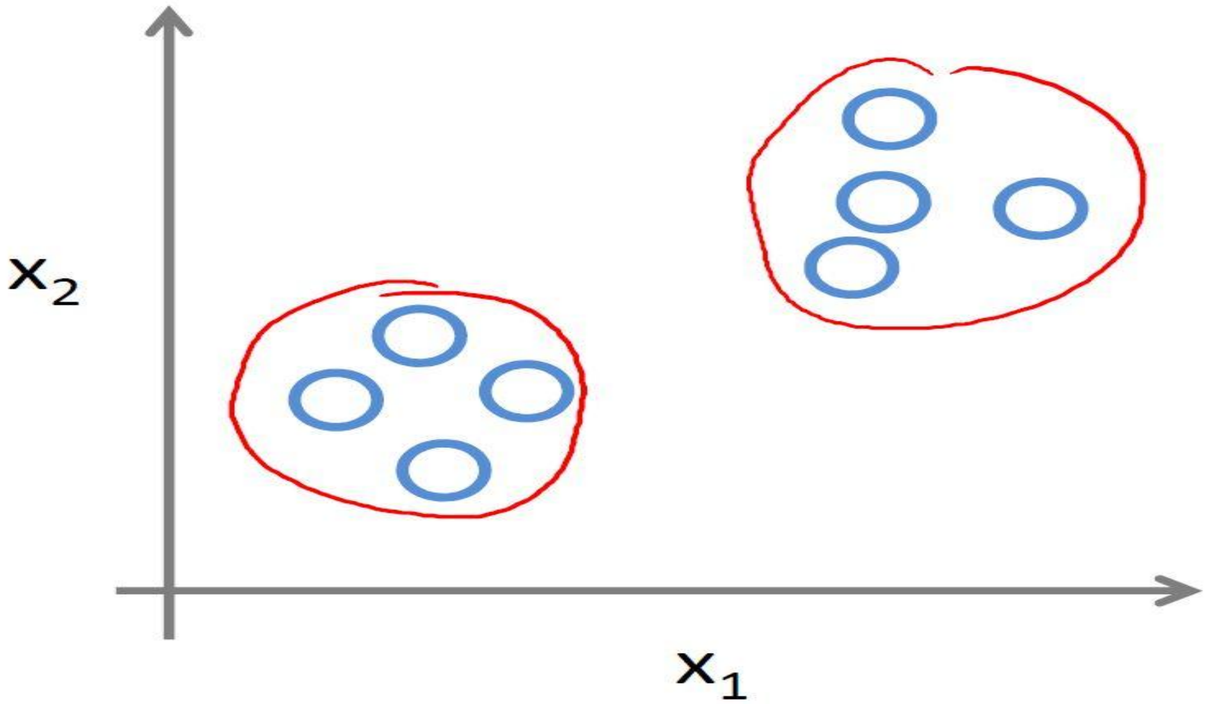
#unique - Classification

Unsupervised Learning:

The goal is not really to discover some O/P but here the goal is to discover some patterns in data.

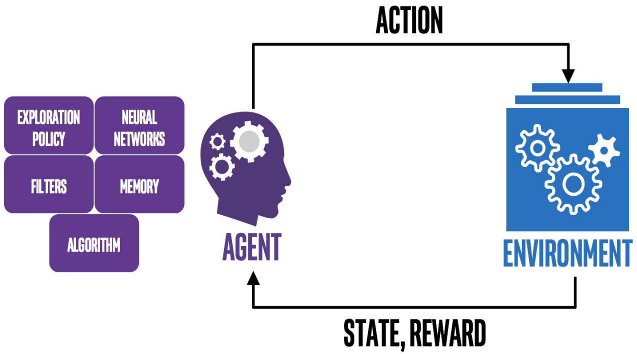
- Discover patterns in the data.
    - **Clustering: Cohesive Grouping** - Categories of customers in a shop
        - College student
        - IT professionals etc.
    - **Association: Frequent Co-occurrence** [finding frequent co-occurrence of items in the dataset given to me. If I see 'A' then it is very likely to see B in my shop.]

Unsupervised Learning



Reinforcement Learning:

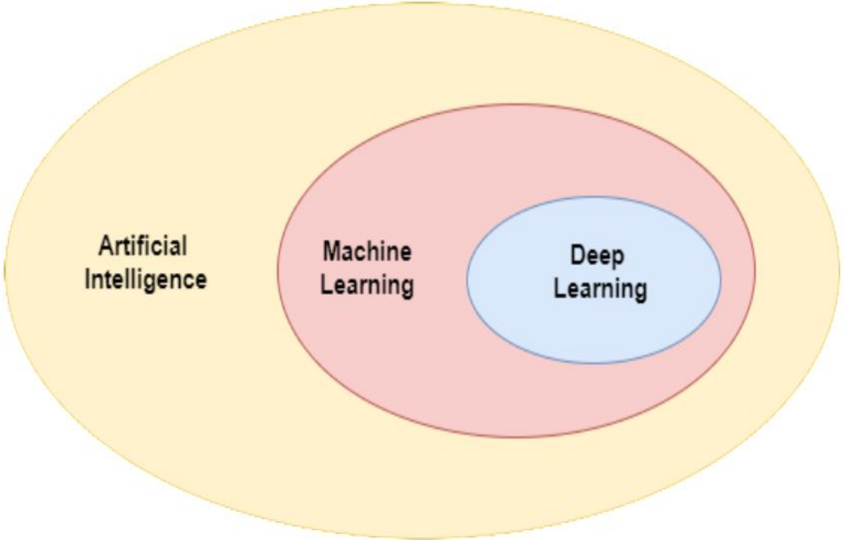
Learning to control behaviour of a system.

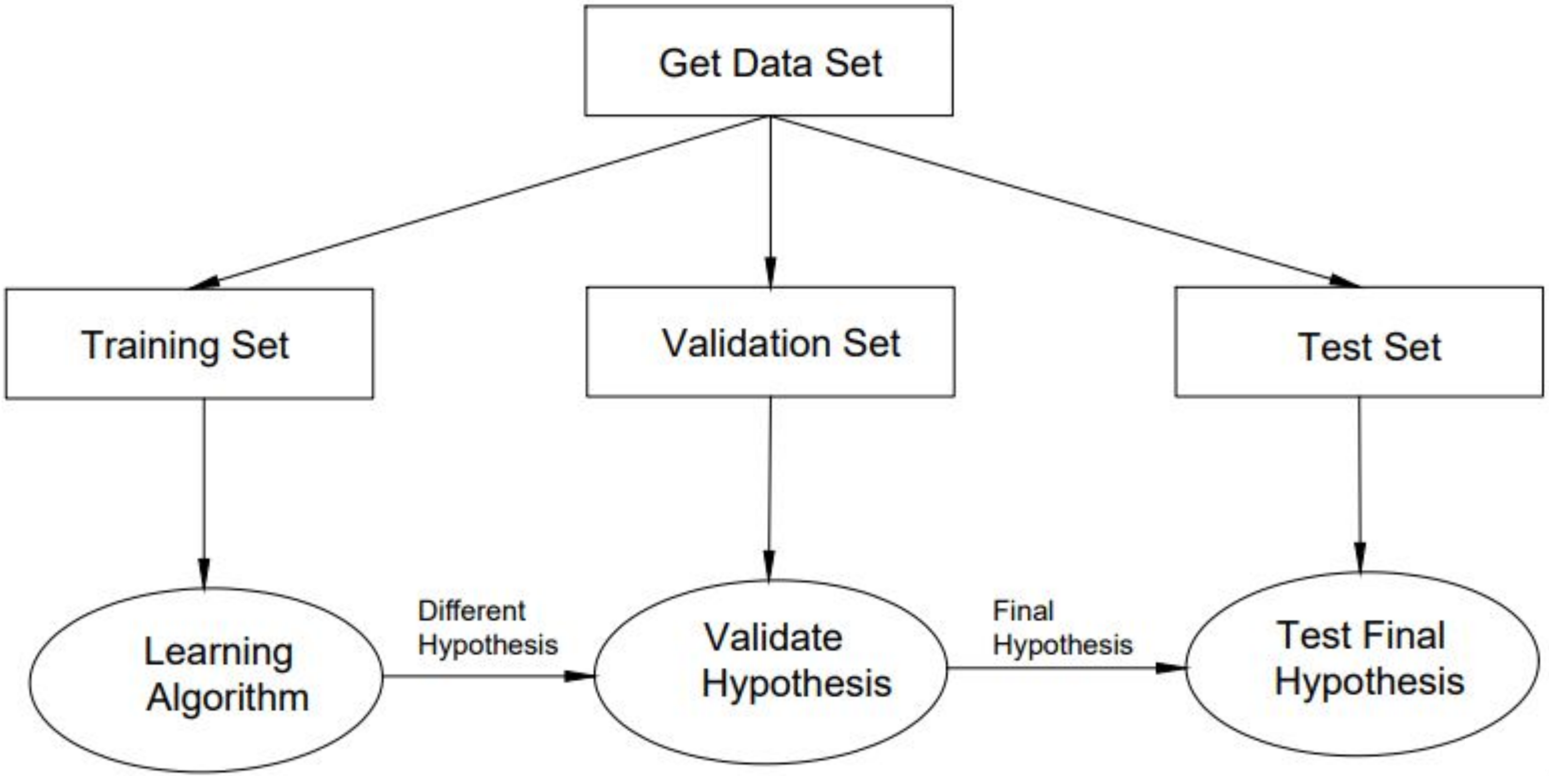


How did you learn to cycle?

- Neither of SV nor USV.
- Some trial or error.
- Falling down Hurts.

This type of learning is a method to learn to control the system through the trial and error and minimal feedback is essential.


Artificial Intelligence
Machine Learning
Deep Learning


Get Data Set
Training Set
Validation Set
Test Set
Learning Algorithm
Different Hypothesis
Validate Hypothesis
Final Hypothesis
Test Final Hypothesis

Dimensionality Reduction

- PCA
- Encoders

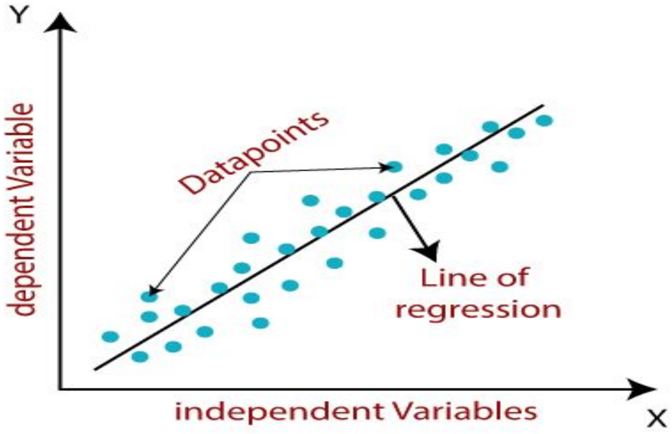
Regression -

- Linear Regression
- Multiple Regression
- Polynomial Regression
- Support Vector Regression
- Decision Tree Regression
- Random Forest Regression

Linear regression -

Algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables.

Linear regression makes predictions for continuous/real or numeric variables



House Pricing Prediction -

Training set

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

m = 47

- Notation:
  - m = Number of training examples
  - x = Input variable / features
  - y = Output variable / target variable
  - (x, y) = One training example
  - $(x^{(i)}, y^{(i)})$ = i^{th} training example

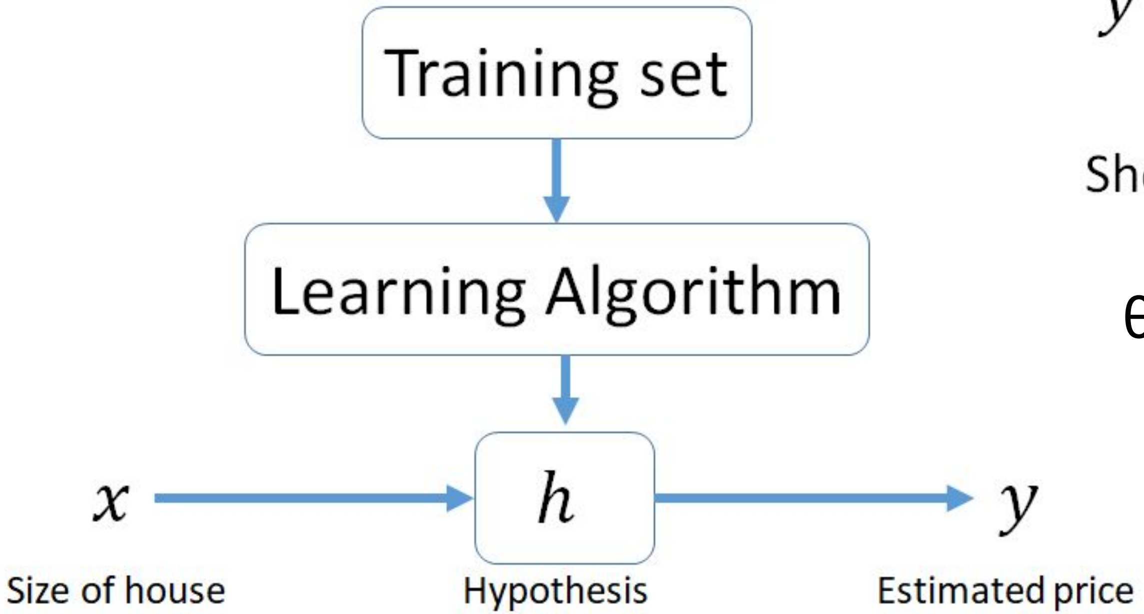
Examples:

$x^{(1)} = 2104$

$x^{(2)} = 1416$

$y^{(1)} = 460$

Model representation



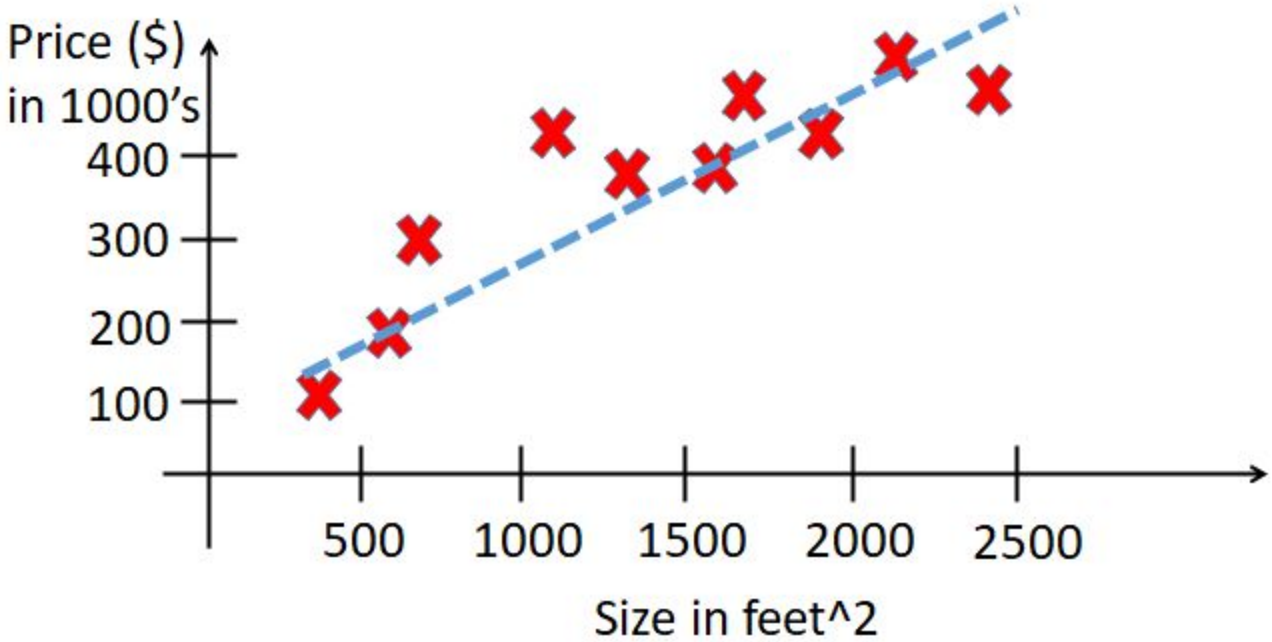
$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

Shorthand $h(x)$

θ_0, θ_1: parameters/weights

How to choose θ_i's?

House pricing prediction



Cost function

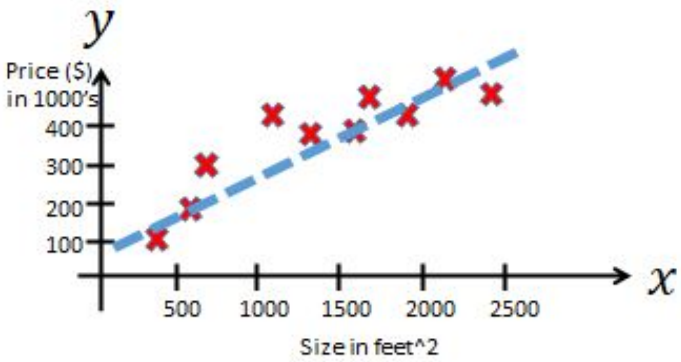
(Root Mean Squared Error (RMSE))

- Idea:

  Choose θ_0, θ_1 so that $h_\theta(x)$ is close to y for our training example (x, y)

$$\underset{\theta_0, \theta_1}{\text{minimize}} \ \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$h_\theta\left(x^{(i)}\right) = \theta_0 + \theta_1 x^{(i)}$$



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

$$\underset{\theta_0, \theta_1}{\text{minimize}} \ \boxed{J(\theta_0, \theta_1)}$$ Cost function

- **Hypothesis:**

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

- **Parameters:**

$$\theta_0, \theta_1$$

- **Cost function:**

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

- **Goal:**

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

Simplified

- **Hypothesis:**

$$h_\theta(x) = \theta_1 x \qquad \theta_0 = 0$$

- **Parameters:**

$$\theta_1$$

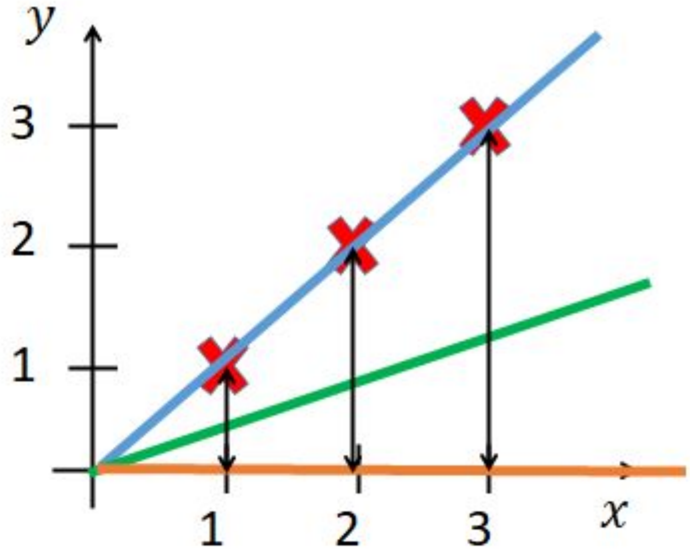
- **Cost function:**

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

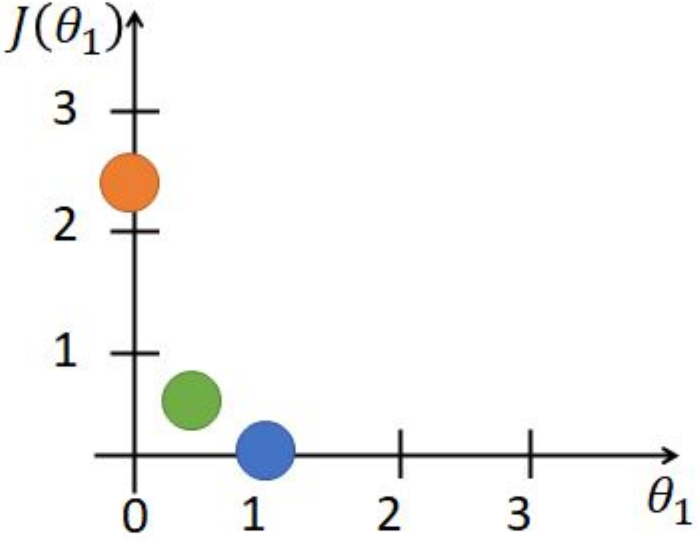
- **Goal:**

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

$h_\theta(x)$, function of x



$J(\theta_1)$, function of θ_1



- **Hypothesis:** $h_\theta(x) = \theta_0 + \theta_1 x$
- **Parameters:** θ_0, θ_1
- **Cost function:** $J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$
- **Goal**: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$?

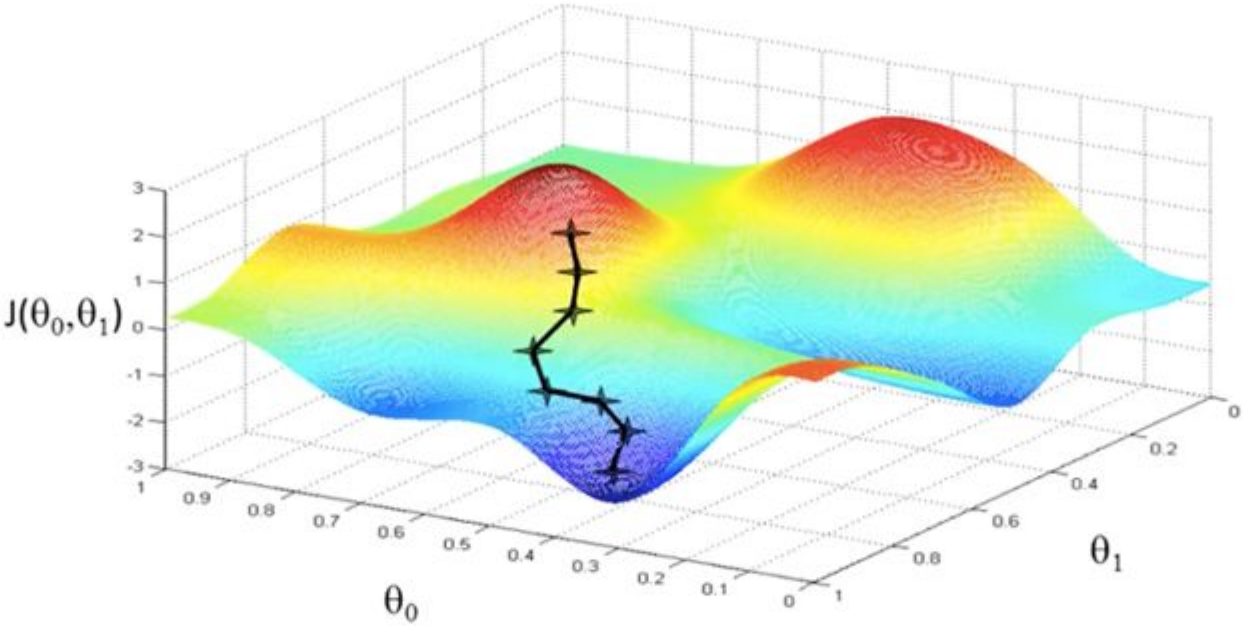
Gradient descent

Have some function $J(\theta_0, \theta_1)$

Want $\underset{\theta_0, \theta_1}{\text{argmin}} \; J(\theta_0, \theta_1)$

Outline:

- Start with some θ_0, θ_1
- Keep changing θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$ until we hopefully end up at minimum



$J(\theta_0,\theta_1)$
3
2
1
0
-1
-2
-3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
θ_0
0
0.2
0.4
0.6
0.8
1
θ_1

Data Preprocessing -

- Getting the dataset
- Importing libraries
- Importing datasets
- Finding Missing Data
- Encoding Categorical Data
- Splitting dataset into training and test set
- Feature scaling

Multiple Linear Regression -

Regression algorithms which models the linear relationship between a single dependent continuous variable and more than one independent variable.

Hypothesis

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

Now: $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$

Multiple features (input variables)

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

Notation:

n = Number of features

$x^{(i)}$ = Input features of i^{th} training example

$x_j^{(i)}$ = Value of feature j in i^{th} training example

$x_3^{(2)} = ?$

$x_3^{(4)} = ?$

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

- For convenience of notation, define $x_0 = 1$ ($x_0^{(i)} = 1$ for all examples)
- $\boldsymbol{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in R^{n+1}$ $\quad\quad$ $\boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \in R^{n+1}$
- $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$
  $= \boldsymbol{\theta}^\top \boldsymbol{x}$

House prices prediction

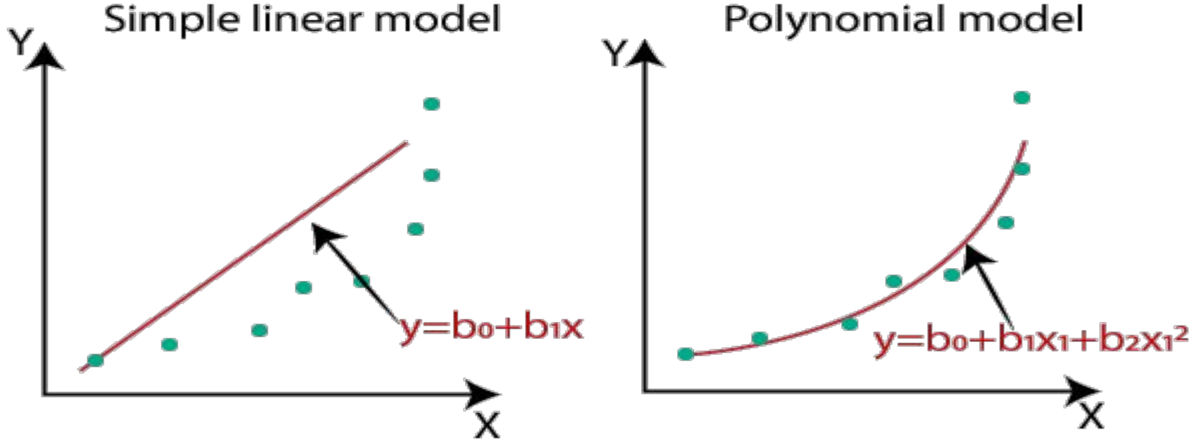
- $h_\theta(x) = \theta_0 + \theta_1 \times \text{frontage} + \theta_2 \times \text{depth}$

- Area $x = \text{frontage} \times \text{depth}$

- $h_\theta(x) = \theta_0 + \theta_1 x$

Polynomial Regression -

Regression algorithm that models the relationship between a dependent(y) and independent variable(x) as nth degree polynomial



Coding content:

1. Linear Regression with two datasets
2. Linear Regression with data preprocessing.