Introduction To Machine Learning

Presented by: Tan Li Tung

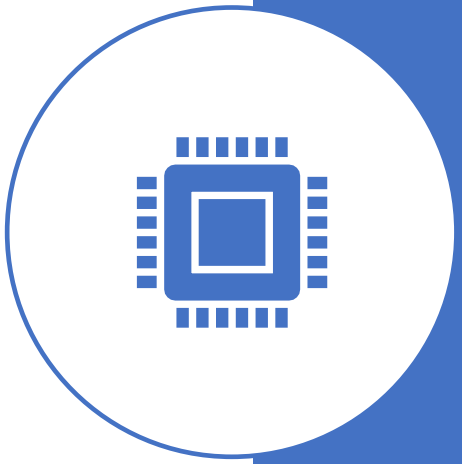
Overview

- Introduction
- Applications of Machine Learning
- Types of Machine Learning
- Regression vs Classification Problem
- Train, Validation, Test Dataset
- Supervised Machine Learning Workflow

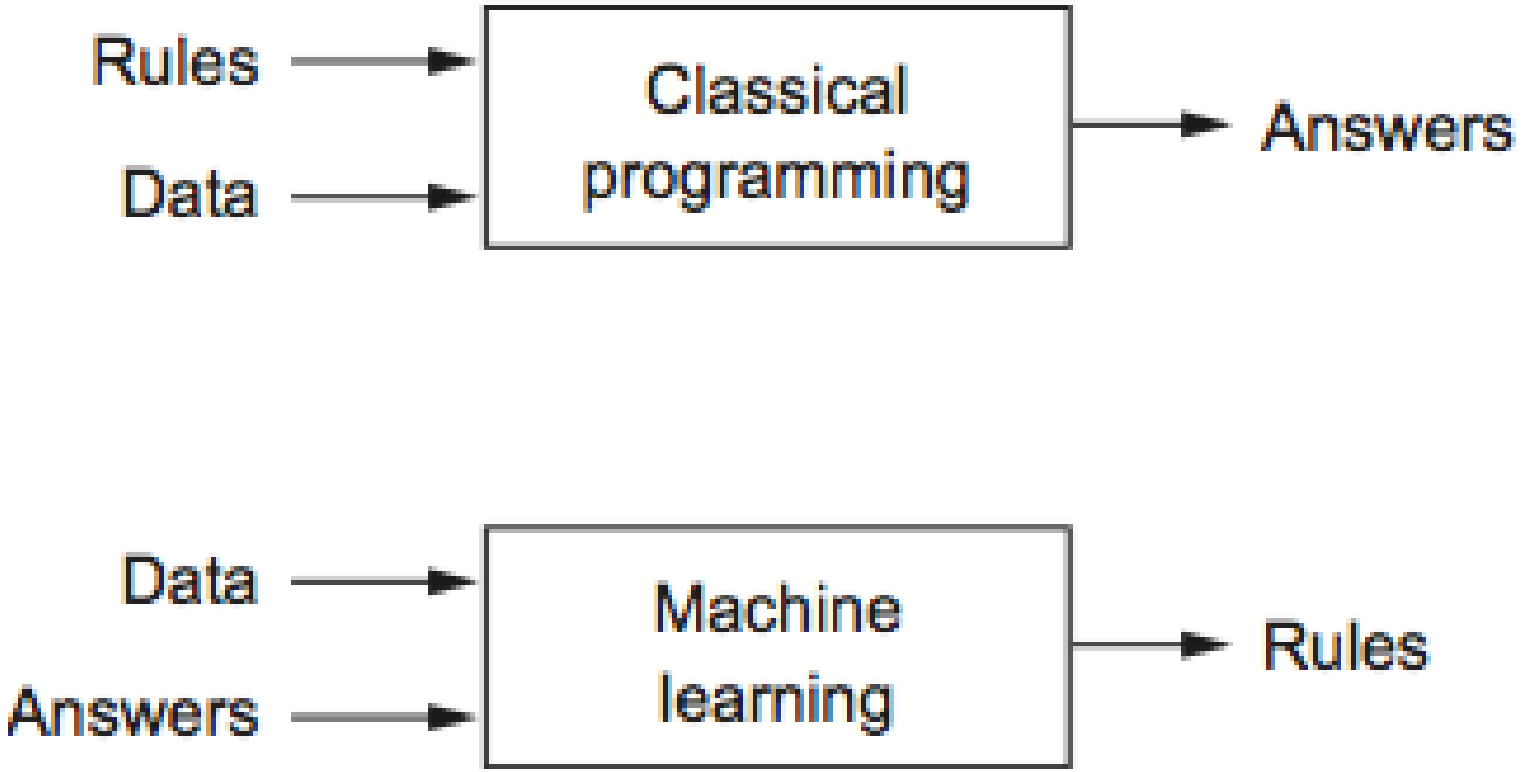
What is Machine Learning?

- Machine learning is a method used to iteratively learn from data without being explicitly programmed by humans

- It allow computers to discover hidden and useful insights

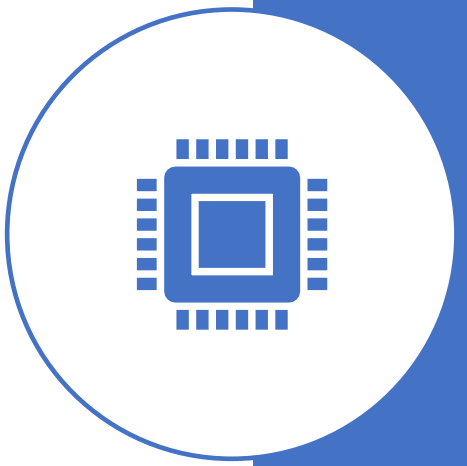


The New Program Paradigm



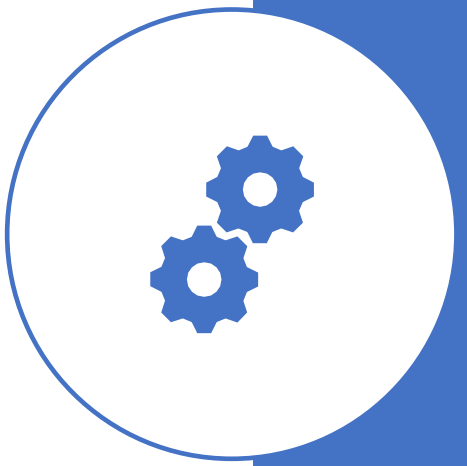
When to Use Machine Learning?

- A pattern exists
- We cannot pin down the pattern mathematically
- We have data and hopefully **LOTS** of data



Applications of Machine Learning

- Products recommendations
- Fraud detection
- Email spam filtering
- Customer segmentation
- Maps
- Image recognition
- Search engine
- Financial market analysis
- And a lot more!



Supervised Learning

- Makes machine learn explicitly
- Data with clear defined output is given
- Direct feedback is given
- Predicts outcome/ future
- Resolve classification and regression problem

Features

Target

PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Survived
1	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25		S	0
2	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C	1
3	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.925		S	1
4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123	S	1
5	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05		S	0
6	3	Moran, Mr. James	male		0	0	330877	8.4583		Q	0
7	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S	0
8	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075		S	0
9	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333		S	1

Unsupervised Learning

- Machine understand the data (identifies patterns/ structures)
- Evaluation is qualitative or indirect
- Does not predict or find anything specific

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850	United Kingdom
536370	22728	ALARM CLOCK BAKELIKE PINK	24	12/1/2010 8:45	3.75	12583	France
536370	22727	ALARM CLOCK BAKELIKE RED	24	12/1/2010 8:45	3.75	12583	France
536370	22726	ALARM CLOCK BAKELIKE GREEN	12	12/1/2010 8:45	3.75	12583	France
536370	21724	PANDA AND BUNNIES STICKER SHEET	12	12/1/2010 8:45	0.85	12583	France
536370	21883	STARS GIFT TAPE	24	12/1/2010 8:45	0.65	12583	France
536370	10002	INFLATABLE POLITICAL GLOBE	48	12/1/2010 8:45	0.85	12583	France

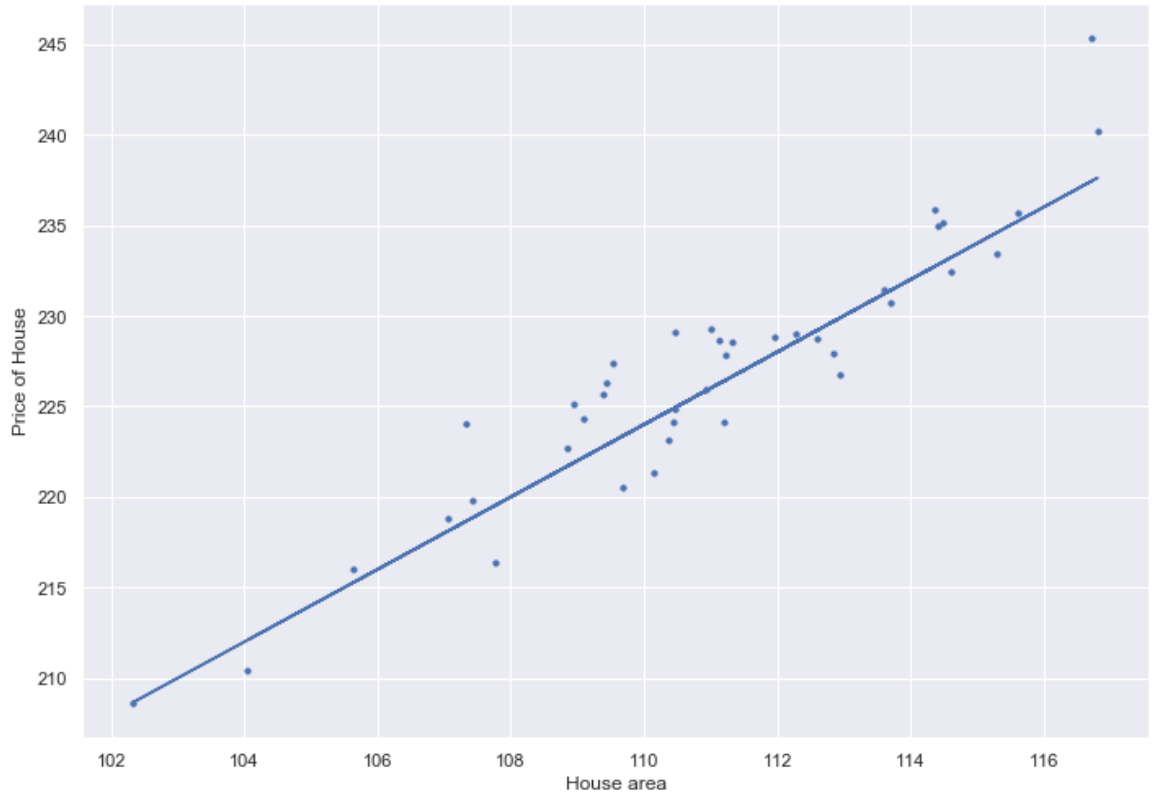
Reinforcement Learning

- An approach to AI
- Reward based learning
- Learn from positive and negative reinforcement
- Machine learns how to act in certain environment
- To maximize reward or minimize punishment


1780
HIGH
1780

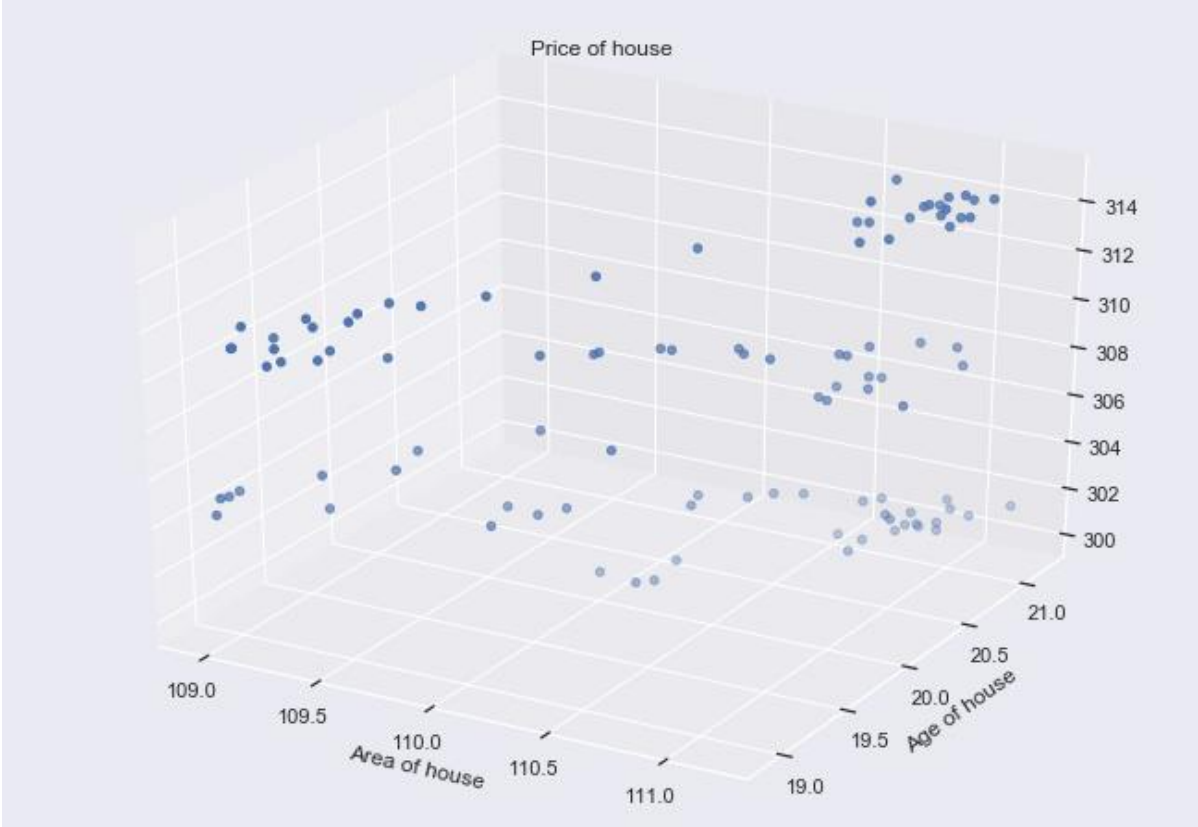
Linear Regression

Feature	Area of house
Label	Price of house
Goal/ Aim	We want to predict the price of a house from the given area



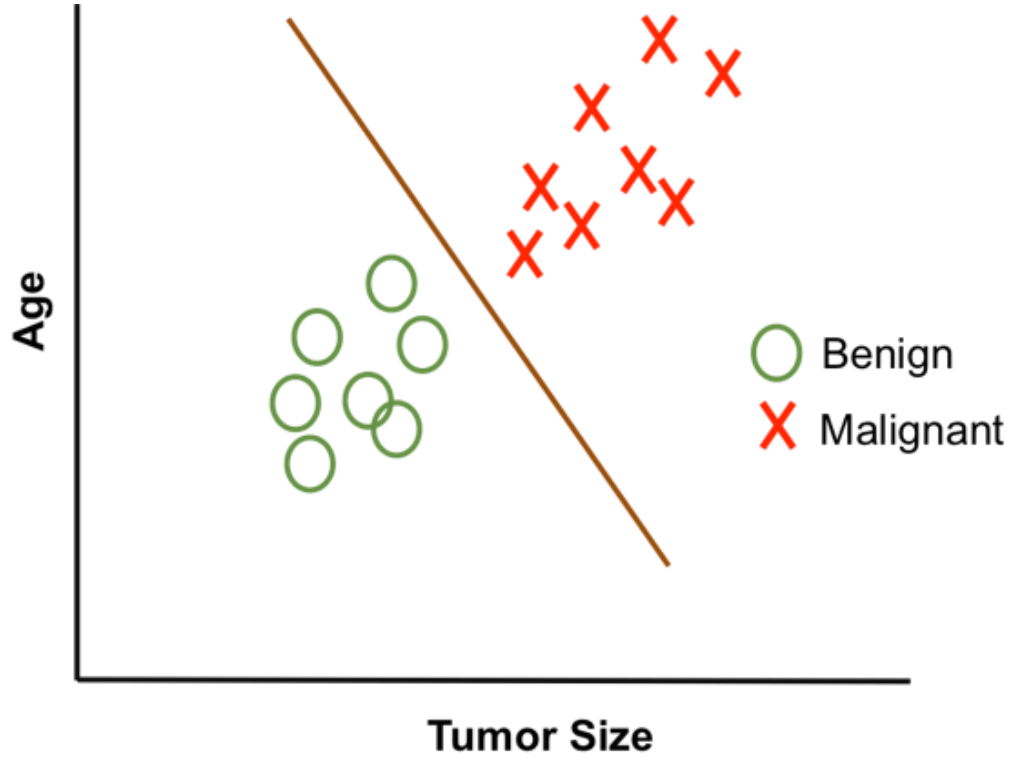
Multiple Linear Regression

Feature	Area of house, Age of house
Label	Price of house
Goal/ Aim	We want to predict the price of a house from the given area and age of house



Classification

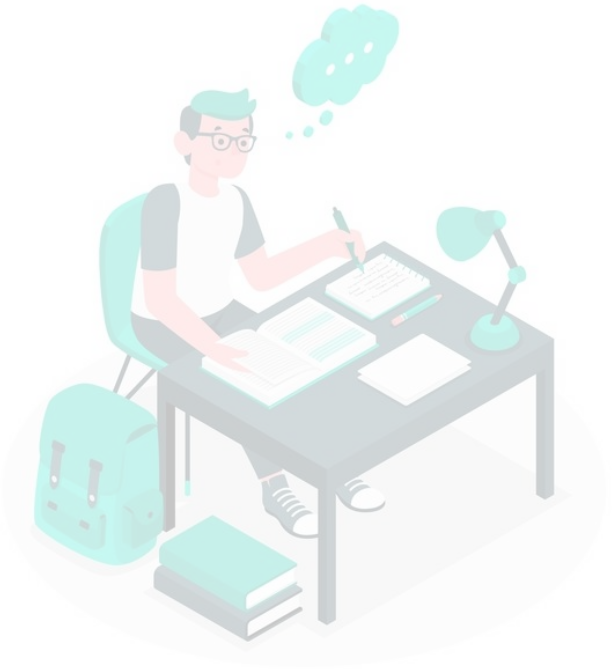
Feature	Tumor Age and Tumor Size
Label	Tumor (Benign or Malignant)
Goal/ Aim	We want to predict whether a tumor is benign or malignant from the given age and tumor size



Train, Validation and Test Dataset

Learn

(Train)



Practice by doing Exercise

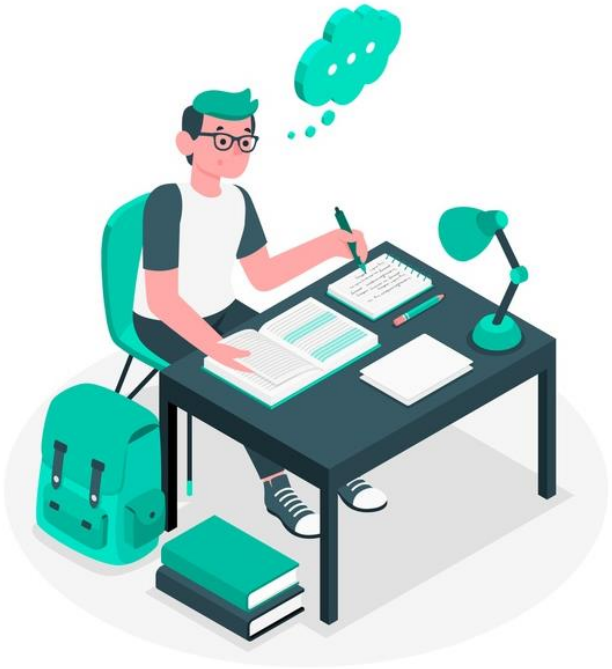
(Validate)

Take Exam

(Test)

Train, Validation and Test Dataset

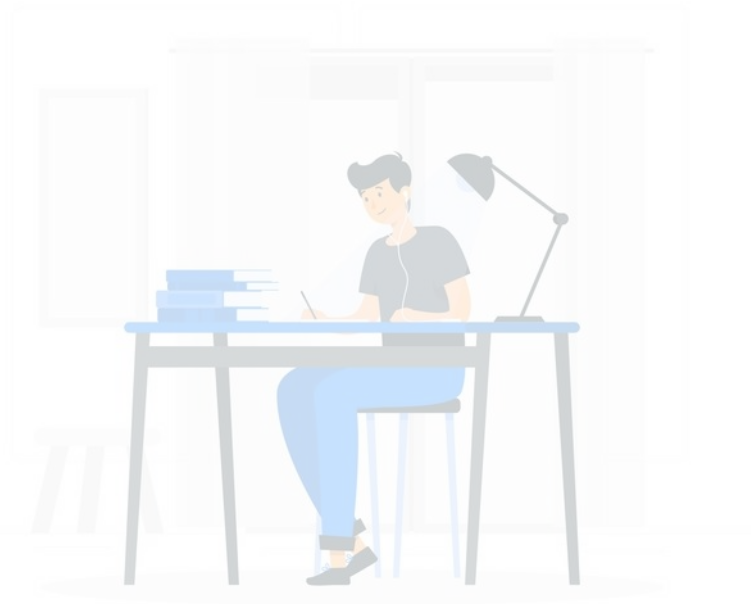
Learn
(Train)



Practice by doing Exercise
(Validate)

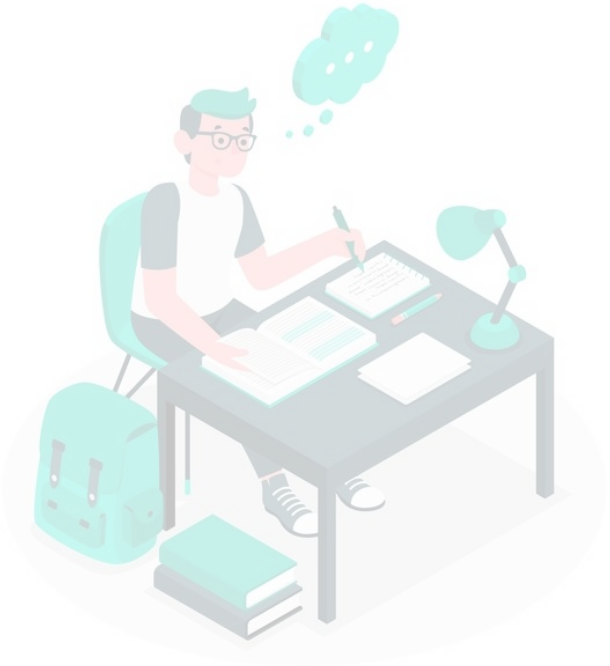
Take Exam
(Test)

Train, Validation and Test Dataset



Learn

(Train)



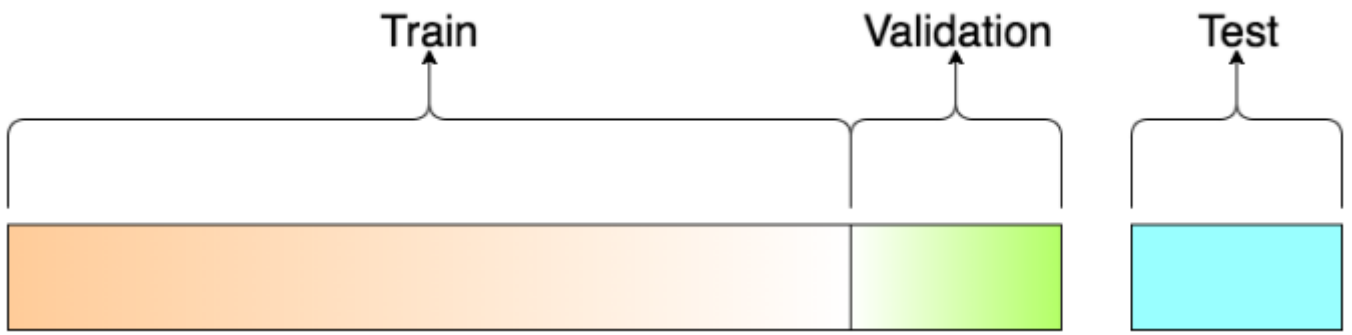
Practice by doing Exercise

(Validate)

Take Exam

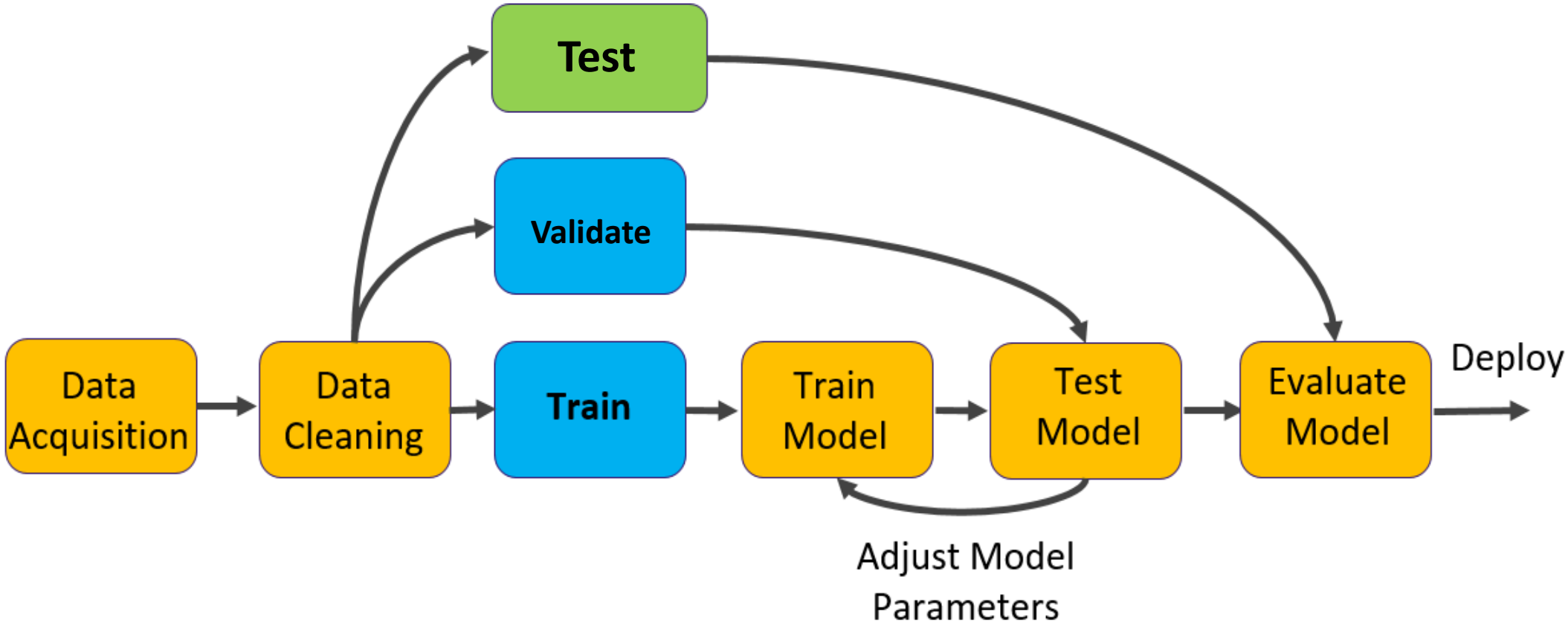
(Test)

Train, Validation and Test Dataset



	Purpose	Yield	Used for Model training	Used for Parameter tuning
Train Data	To learn patterns from the data.	A model that makes near-expected predictions	Yes	Yes
Validation Data	To understand model behaviour and generalizibility on unseen data.	Insights on how to tune your model.	No	Yes
Test Data	To understand how the model would perform in real world scenario.	A completely unbiased estimate of model performance.	No	No

Supervised Machine Learning Workflow



Hands on!

Normalization and Standardization

- **Normalization**

Scale the numeric values in between the range of 0 and 1.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

- **Standardization**

Scale the numeric values such that it have a mean of 0 and variance of 1.

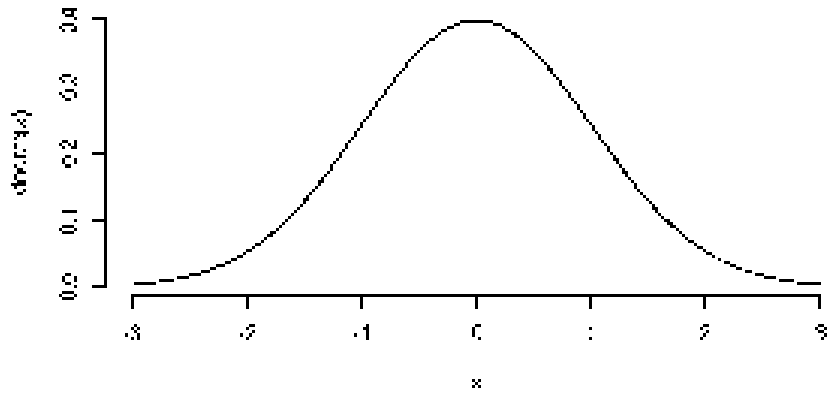
$$X' = \frac{X - \mu}{\sigma}$$

When to Normalize and Standardize?

Generally (but not limited to):

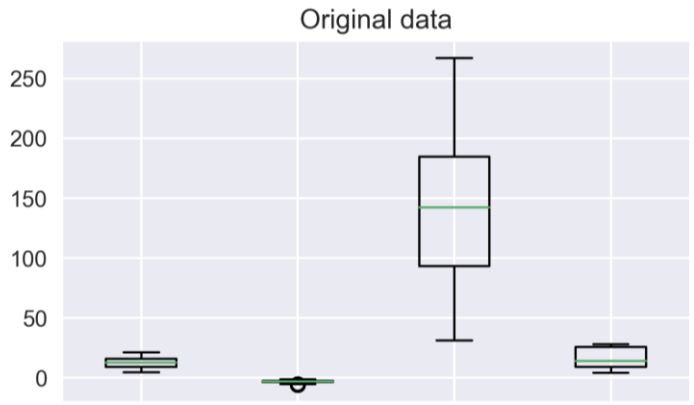
- Normalization is used when data is not Gaussian distributed.

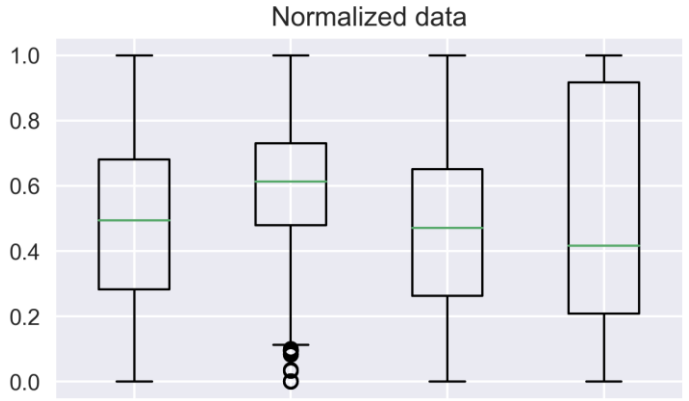
- Standardization, on the other hand, is used when the data follows a Gaussian distribution.

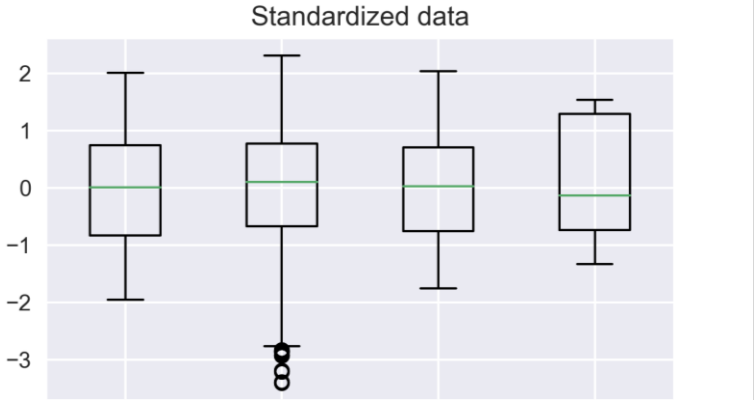


When to Normalize and Standardize?

You can always fit your model with raw, normalized and standardized data then compare the performance to obtain the best outcome!







Metrices for Model Evaluation (Regression)

Mean squared error	$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$
Root mean squared error	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$
Mean absolute error	$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} \lvert e_t \rvert$

Questions?

Thank you