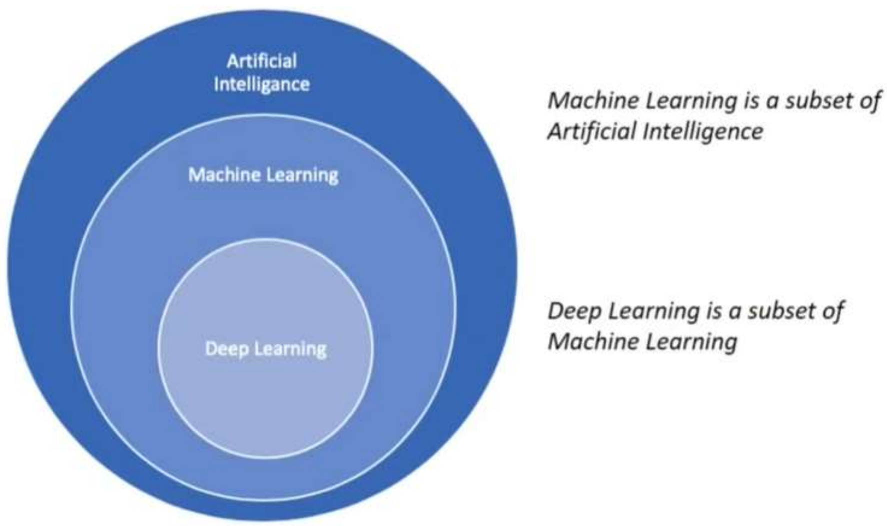
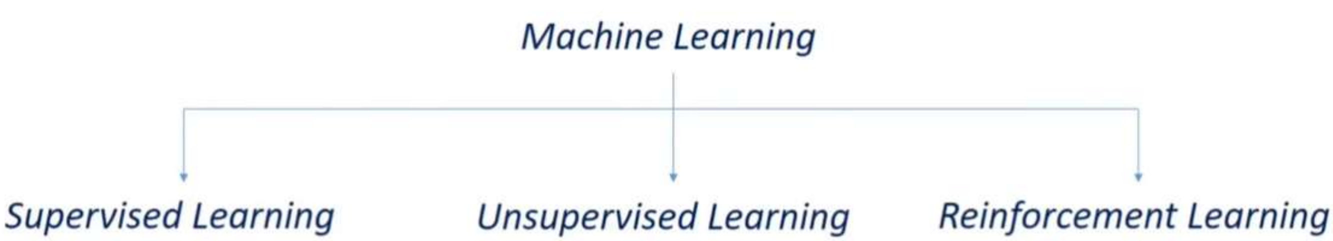
16 March 2025 11:23

AI vs ML vs DL

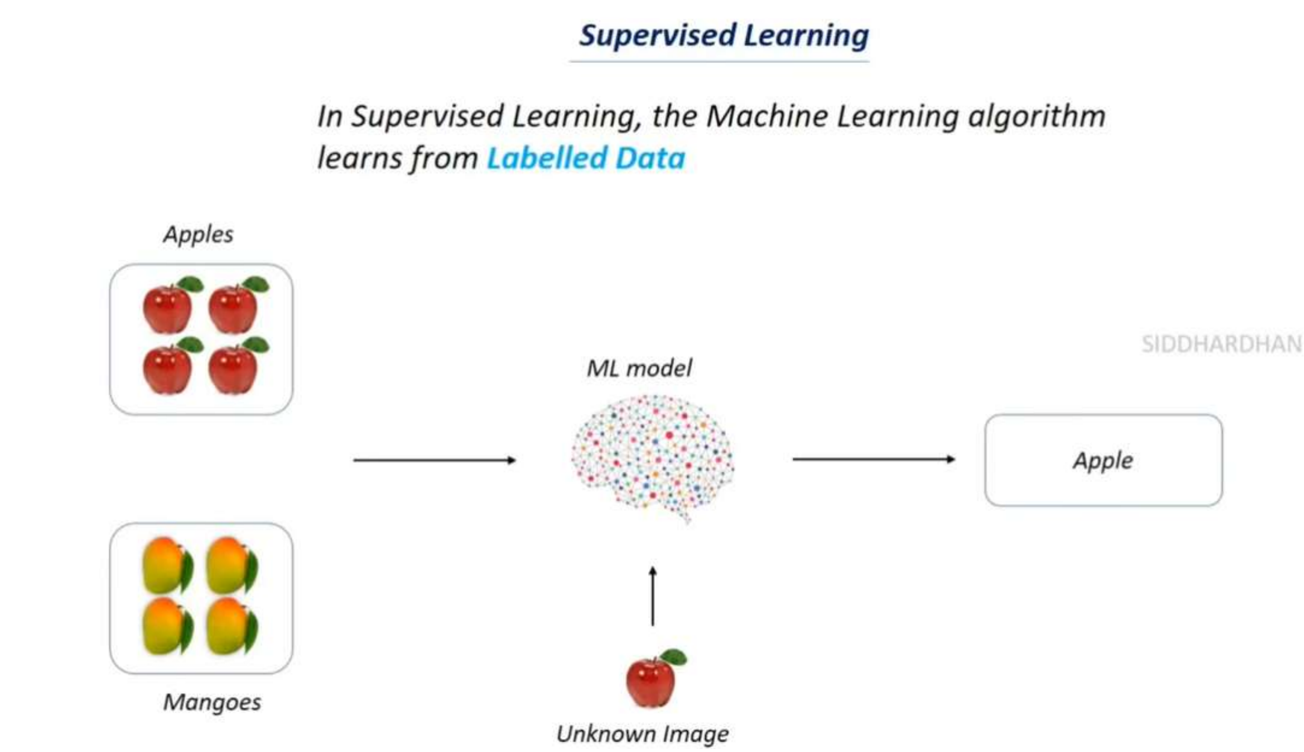


AI - To be built intelligent machines
ML - Technique used to implement AI with data to learn from
DL - Uses artitficial neural network to learn from data

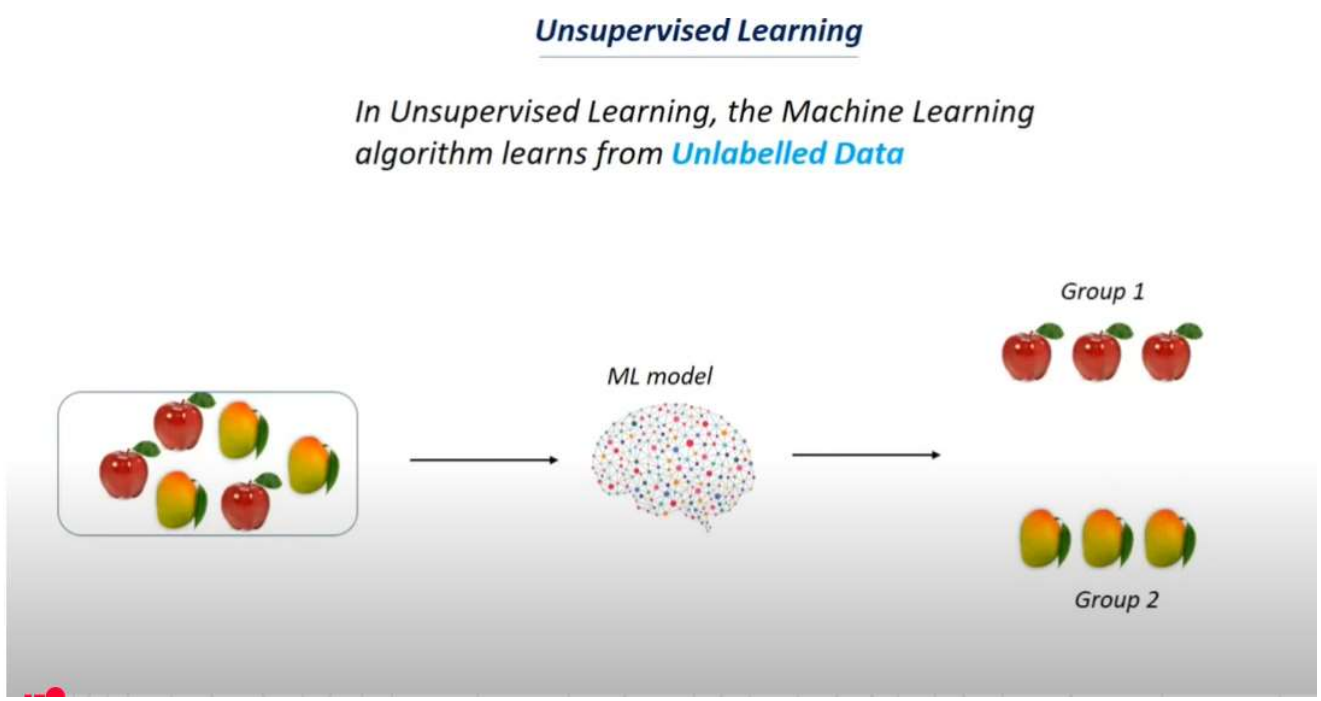
Types of Machine Learning



Supervised Learning : learns from labelled data



Unsupervised Learning : learns from Unlabelled data



Reinforcement Learning : intelligent agents takes action in an environement

Reinforcement Learning

Reinforcement Learning is an area of Machine Learning concerned with how intelligent agents take actions in an environment to maximize its rewards.

1. *Environment*
2. *Agent*
3. *Action*
4. *Reward*

Agent

Environment

Types of Supervised Learning

Types of Supervised Learning

Supervised Learning

Classification

Classification is about predicting a class or discrete values
Eg: Male or Female; True or False

Regression

Regression is about predicting a quantity or continuous values
Eg: Salary; age; Price.

Types of Unsupervised Learning

Types of Unsupervised Learning

Unsupervised Learning

Clustering

Clustering is an unsupervised task which involves grouping the similar data points.

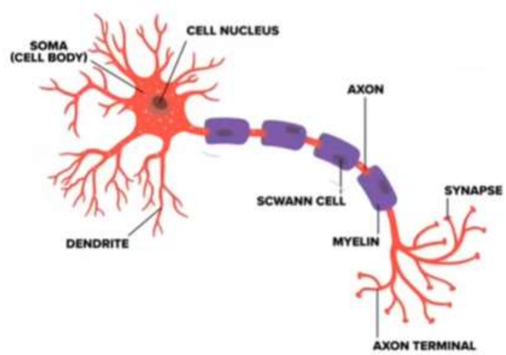
Association

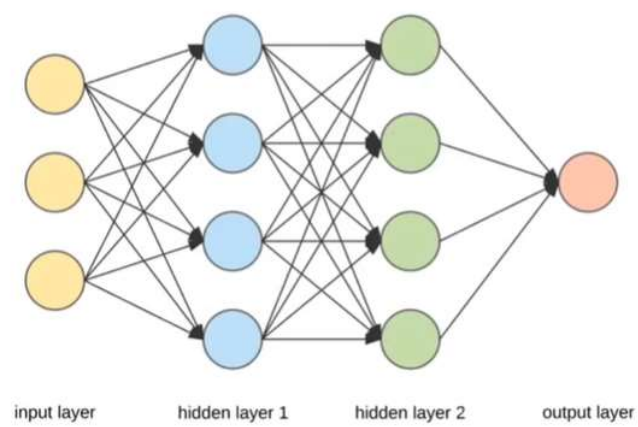
Association is an unsupervised task that is used to find important relationship between data points

Deep Learning

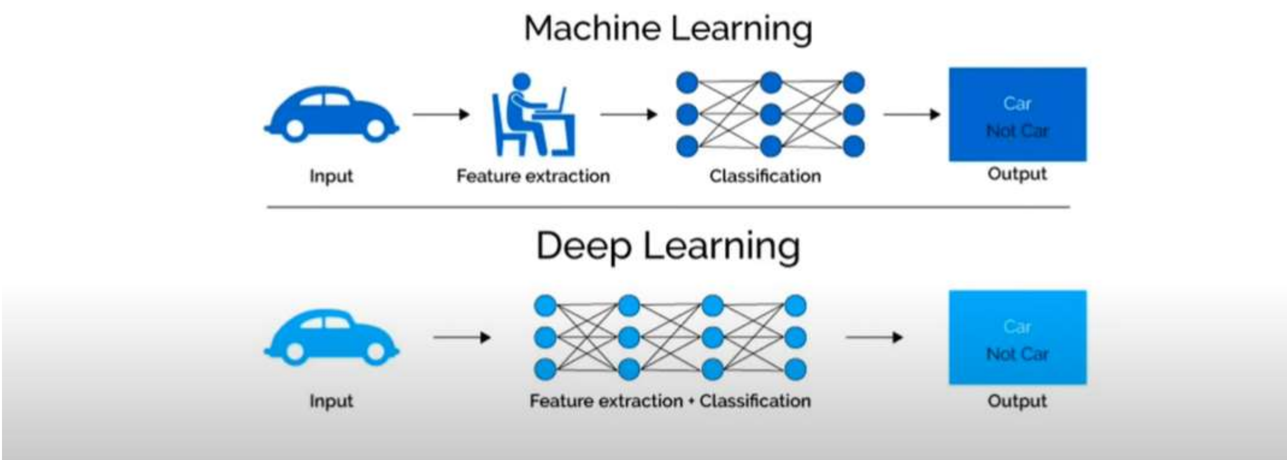
Deep Learning

Deep Learning is a subfield of Machine Learning that uses Artificial Neural Networks to learn from the data.





Machine Learning vs Deep Learning



No need to do feature extraction manually in Deep Learning

Different data types:

Immutable objects - whose value cannot be changed ex: int, float, bool, str, tuple
Mutable objects - whose value can be changed , ex: list, dictionary, set

List should be mentioned in squared bracket.
.append() function is used to add element in the list, only one element at a time.
List allows duplicate value.

Lists:

- ***Syntax***: *[]*
- ***Ordered***: *Yes (Maintains the insertion order)*
- ***Mutable***: *Yes (Elements can be changed)*
- ***Duplicates***: *Yes (Allows duplicate elements)*
- ***Example***: *[1, 2, 3, 4]*

Tuples:

- ***Syntax***: *()*
- ***Ordered***: *Yes (Maintains the insertion order)*
- ***Mutable***: *No (Elements cannot be changed once assigned)*
- ***Duplicates***: *Yes (Allows duplicate elements)*
- ***Example***: *(1, 2, 3, 4)*

Sets:

- ***Syntax***: *{} or set()*
- ***Ordered***: *No (No guarantee of order in Python versions before 3.7)*
- ***Mutable***: *Yes (Elements can be added or removed)*
- ***Duplicates***: *No (Does not allow duplicate elements)*
- ***Example***: *{1, 2, 3, 4}*

Dictionaries:

- ***Syntax***: *{key: value}*
- ***Ordered***: *Yes (Maintains the insertion order starting from Python 3.7)*
- ***Mutable***: *Yes (Keys and values can be changed)*
- ***Duplicates***: *No (Keys must be unique, but values can be duplicated)*
- ***Example***: *{'a': 1, 'b': 2, 'c': 3}*

From <https://www.geeksforgeeks.org/differences-and-applications-of-list-tuple-set-and-dictionary-in-python/>

Numpy

They allow several mathematical operations
They are faster

Correalation

If one value increases and other also increases with it is positively correlated
If one value decreases and other also decrease with it is negatively correlated

Matplotlib

```
plt.plot(x,y)
plt.xlabel("value")
plt.ylabel("value")
plt.title("Parabola")
plt.show()
```

```
languages = ["French", "English", "Latin", "German", "Hindi"]
people = [40,80,20,10,100]
plt.bar(languages, people, color = "orange")
plt.xlabel("languages")
plt.ylabel("people")
plt.title("Diversity and Inclusion")
plt.show()
```

```
x = [10,20,30,40]
y = ["green","blue","red","violet"]
plt.pie( x , labels = y, startangle= 90)
plt.title("Pie chart")
plt.show()
```

```
 x = np.linspace(0, 10, 30)
 y = np.sin(x)
 plt.scatter(x, y)
 plt.xlabel("values")
 plt.ylabel("sin values")
 plt.title("sin scatter plot")
 plt.show()
```

Seaborn

```
sns.set_theme()
```

```
iris = sns.load_dataset("iris")    #load dataset
```

```
sns.scatterplot(data = iris, x = "sepal_length", y ="petal_length", hue
="species")
```

```
sns.countplot(data = titanic , x = "class")
```

```
sns.barplot(data = titanic , x = "sex", y = "survived", hue = "class")
```

```
sns.displot(data = titanic , x = "fare")
```

Handling Missing values

There are two methods to handle it

1. Imputation
2. Dropping

Impute with Central Tendencies:

1. Mean
2. Median
3. Mode

When to use mean median mode

Create a distribution plot, if you have outliers and the chart is skewed you cannot use mean, it will increase the mean so in that case use median or mode

```
df['salary'].fillna(df['salary'].median(), inplace = True)
```

Drop only if dataset is very large

```
placement_dataset = dataset.dropna( how = "any")
```

Data Standardization

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import sklearn.datasets
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
```

Standarization should be near to 1
If not then use standard scaler to make it near to 1

```
X = df
Y = datasets.target
```

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2,
random_state=3)

datasets.data.std()
```

Create variable

```
scaler = StandardScaler()
```

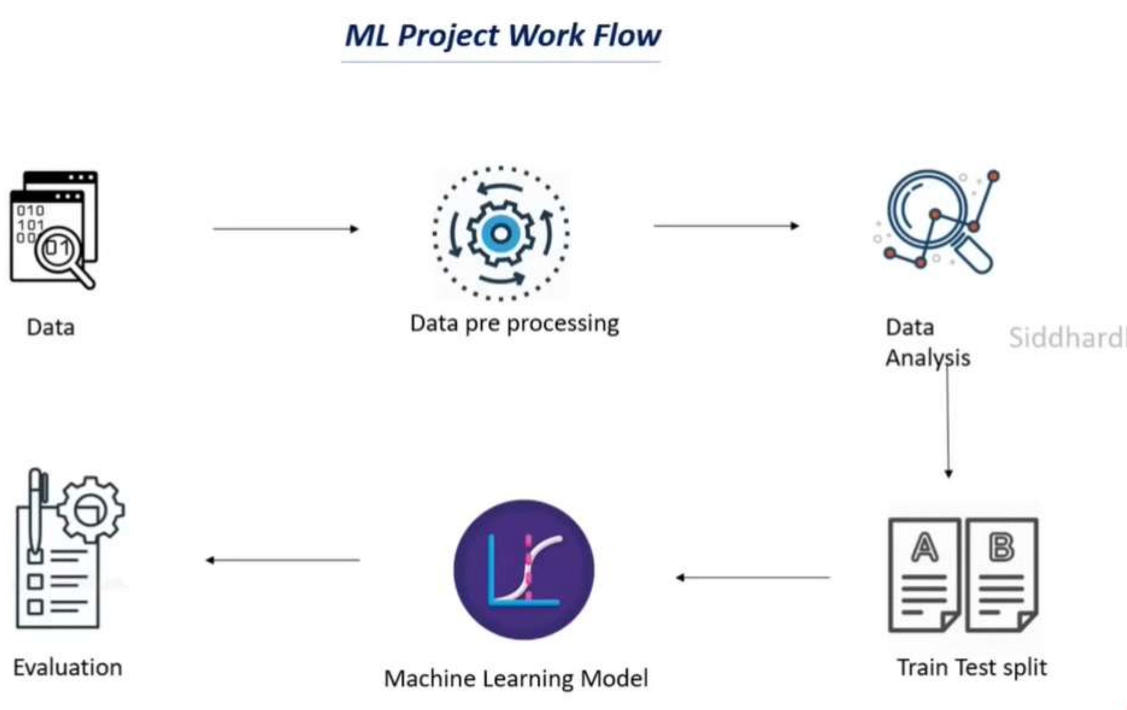
First fit the data then later transform

```
scaler.fit(X_train)

X_train_standardized = scaler.transform(X_train)

print(X_train_standardized.std())

X_test_standardized = scaler.transform(X_test)

print(X_test_standardized.std())
```

Now std should be near to 1

Label Encoding (Used in Classification)
Convert labels into numeric form

```
#import the lib
import pandas as pd
from sklearn.preprocessing import LabelEncoder

#create a dataframe
iris = pd.read_csv("/content/iris_data.csv")

#to check no. of labels present in df
iris["Species"].value_counts()

#create a variable
label_encode = LabelEncoder()

#to transform labels into no.
label = label_encode.fit_transform(iris["Species"])

# add column in table
iris["Target"] = label

iris.head()

iris["Target"].value_counts()
```



Imbalanced Dataset

Suppose if your data have 1000 diabetes and only 100 non- diabetes it will make the prediction go wrong , so before that we need to handle the imbalanced dataset to make correct prediction

You can do undersampling here to match the data

```
legit_sample = legit.sample(n = 264)
```

Feature Extraction:

Since machine cannot understand textual data, we convert the text into vector values

Feature Extraction

The mapping from textual data to real valued vectors is called feature extraction.

Bag Of Words (BOW): list of unique words in the text corpus

Term Frequency-Inverse Document Frequency (TF-IDF):
To count the number of times each word appears in a document

Tf-idf Vectorizer

Term Frequency (TF) = (Number of times term t appears in a document)/(Number of terms in the document)

Inverse Document Frequency (IDF) = log(N/n), where, N is the number of documents and n is the number of documents a term t has appeared in.

The IDF value of a rare word is high, whereas the IDF of a frequent word is low.

Tf-idf

```
[21] # convert the textual data to Feature Vectors
     vectorizer = TfidfVectorizer()

[22] vectorizer.fit(X)

     X = vectorizer.transform(X)

     print(X)

  (0, 15686)	0.28485063562728646
  (0, 13473)	0.2565896679337957
  (0, 8909)	0.3635963806326075
  (0, 8630)	0.29212514087043684
  (0, 7692)	0.24785219520671603
  (0, 7005)	0.21874169089359144
  (0, 4973)	0.233316966909351
  (0, 3792)	0.2705332480845492
```

The number are generated from = TF * IDF

Numerical dataset Pre-processing

Same as Data Standardiztion to bring numbers in -1 to +1 range

Textual Data Pre-preprocessing

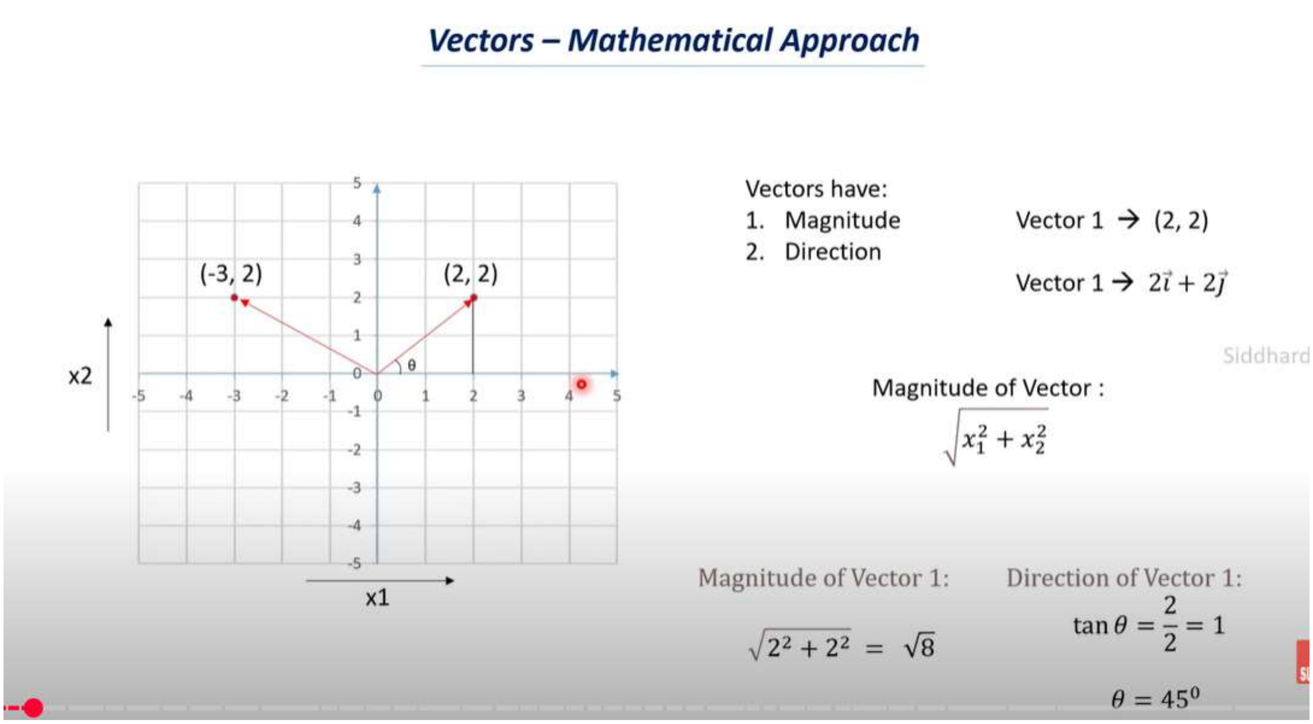
Refer to ML2 colab notebook

Stemming is reducing the word to root word, ex: enjoying, enjoybale -- enjoy

Mathermatics for machine learning

Physics approach:

Scalar have only magnitude
Vectors have magnitude and direction



Vectors – Computer Science Approach

Scalar

24

Vector

$\begin{bmatrix} 2 & -8 & 7 \end{bmatrix}$

row

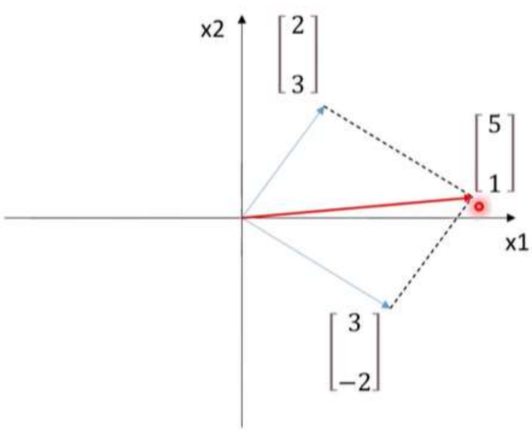
or column $\begin{bmatrix} -6 \\ -4 \\ 27 \end{bmatrix}$

Work Experience & Salary

$\begin{bmatrix} 5 \\ ₹\,5{,}00{,}000 \end{bmatrix}$

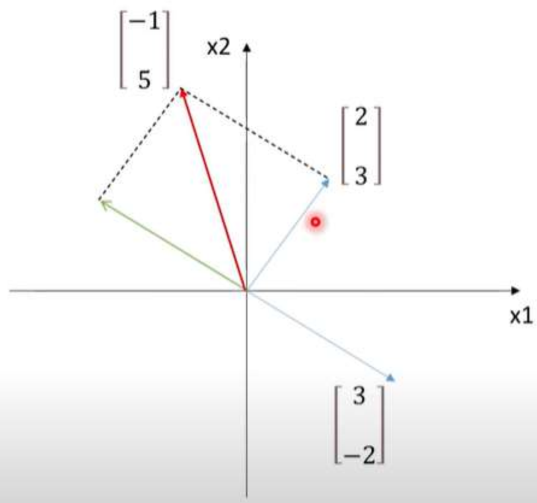
$\begin{bmatrix} 10 \\ ₹\,10{,}00{,}000 \end{bmatrix}$

Vector Addition



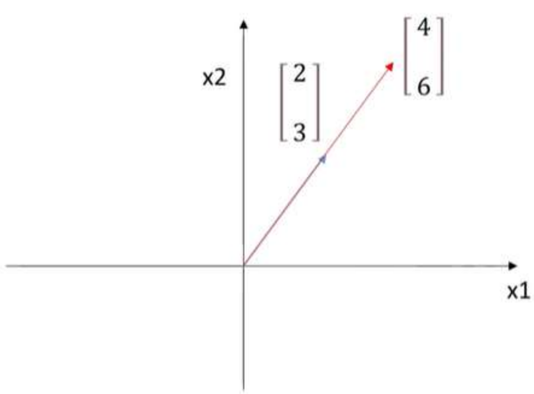
$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 2+3 \\ 3+(-2) \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

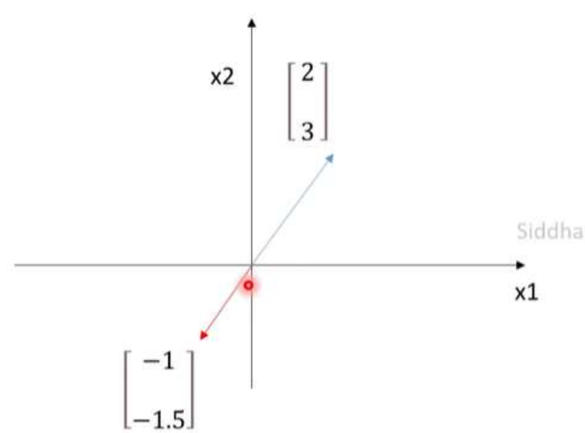
Vector Subtraction



$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 2-3 \\ 3-(-2) \end{bmatrix} = \begin{bmatrix} -1 \\ 5 \end{bmatrix}$$

Multiplying a vector by a Scalar

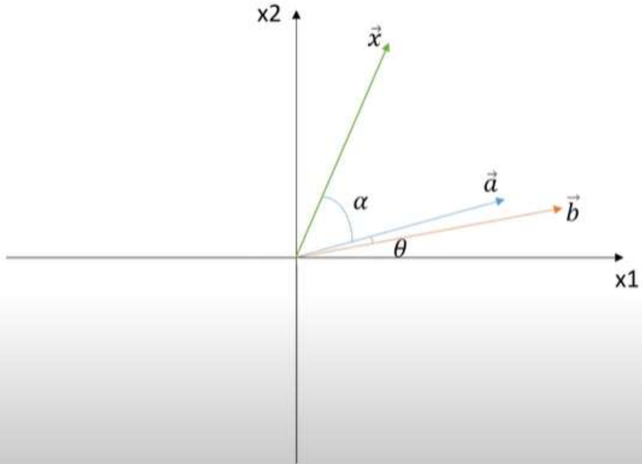




$$2 \times \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

$$-0.5 \times \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -1 \\ -1.5 \end{bmatrix}$$

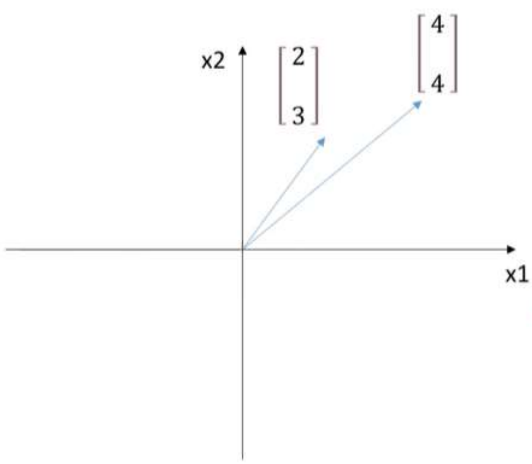
Angle Between 2 Vectors



Inference:

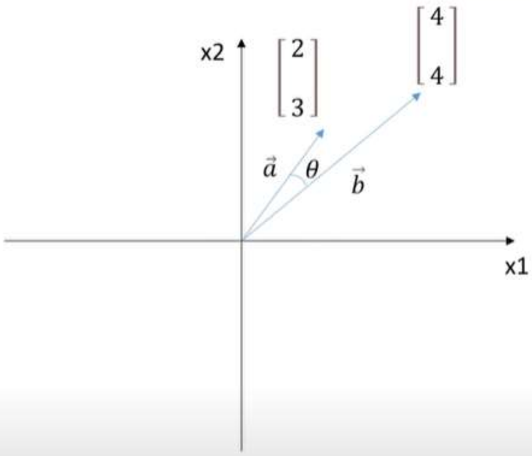
- ✓ If the angle between 2 vectors is small, then the 2 vectors are similar.
- ✓ If the angle between 2 vectors is large, then the 2 vectors are very different.

Dot Product of 2 Vectors



$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} \bullet \begin{bmatrix} 4 \\ 4 \end{bmatrix} = (2 \times 4) + (3 \times 4) = 20$$

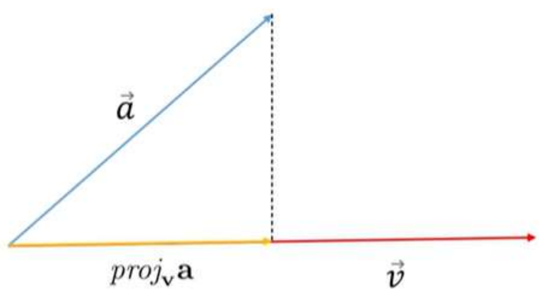
Cross Product of 2 Vectors



$$\vec{a} x \vec{b} = |\vec{a}||\vec{b}| \sin\theta$$

$$\overline{a} \times \overline{b} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 2 & 3 & 0 \\ 4 & 4 & 0 \end{vmatrix} = \boldsymbol{i}(3\cdot 0 - 0\cdot 4) - \boldsymbol{j}(2\cdot 0 - 0\cdot 4) + \boldsymbol{k}(2\cdot 4 - 3\cdot 4) = \boldsymbol{i}(0 - 0) - \boldsymbol{j}(0 - 0) + \boldsymbol{k}(8 - 12) = \{0; 0; -4\}$$

Projection of Vector



$$proj_{\mathbf{v}}\mathbf{a} = \frac{\mathbf{a} \cdot \mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v}$$

Scalars; Vectors; Matrix

Scalar

24

Vector

$\begin{bmatrix} 2 & -8 & 7 \end{bmatrix}$

row

or

column $\begin{bmatrix} -6 \\ -4 \\ 27 \end{bmatrix}$

Matrix

$$\begin{bmatrix} 1 & 4 & 5 \\ 2 & 7 & 3 \\ 8 & -3 & 1 \end{bmatrix}$$

Multiplying 2 Matrices

Rule : The number of columns in the First matrix should be equal to the number of rows in the Second Matrix

The resultant matrix will have the same number of rows as the first matrix & the same number of columns as the Second Matrix

$$\begin{bmatrix} 2 & 3 \\ 10 & 5 \end{bmatrix} \times \begin{bmatrix} 10 & 5 \\ 20 & 4 \end{bmatrix}$$

2 x 2   2 x 2

Can be multiplied.
Resultant matrix will have the shape 2 x 2

$$\begin{bmatrix} 2 & 1 \\ 4 & 2 \\ 6 & 3 \end{bmatrix} \times \begin{bmatrix} 5 & 2 \\ 3 & 6 \\ 2 & 5 \end{bmatrix}$$

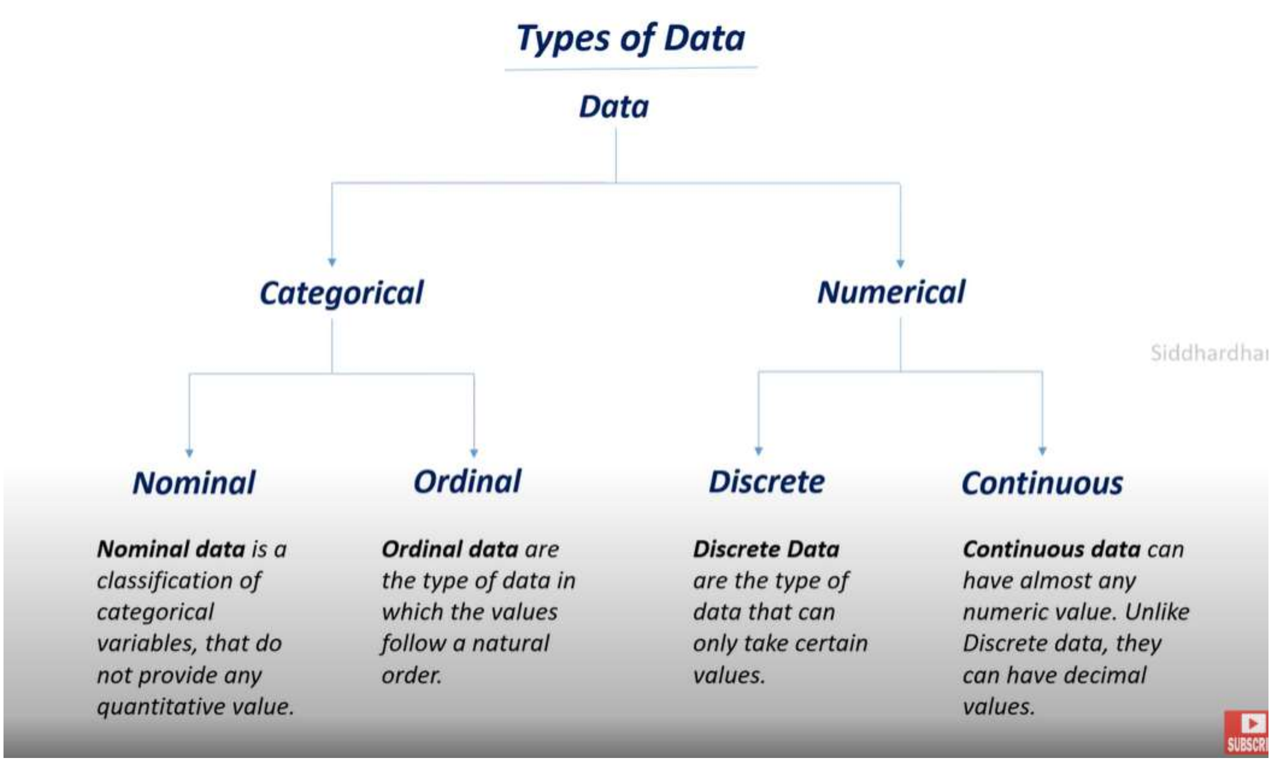
3 x 2   3 x 2

Cannot be multiplied.

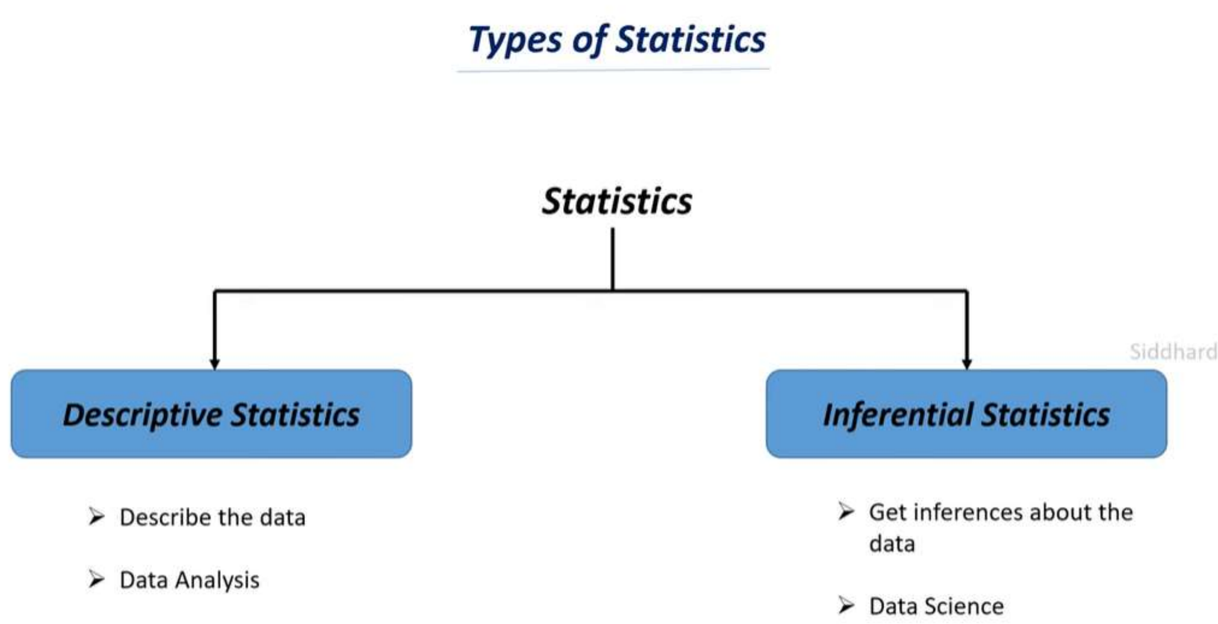
Statistics for machine learning:

Need: Helps us to extract information and knowledge from data

Applications: Six Sigma in manufacturing plant, businesses, weather forecast, clinical trial of medicines



Nominal =male, female
Ordinal = bad ,average, good
Discrete = cannot have decimal value
Continuous = can have decimal value



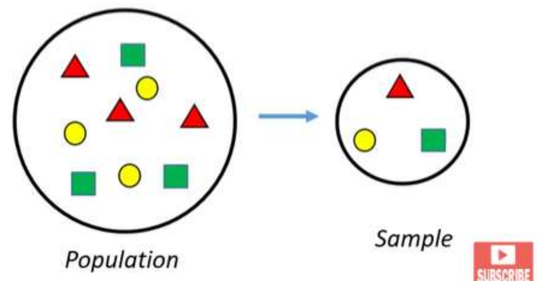
Types of Statistics

1. Descriptive Statistics:

***Descriptive statistics** are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures.*

2. Inferential Statistics:

***Inferential statistics** takes data from a sample and makes inferences and predictions about the larger population from which the sample was drawn.*



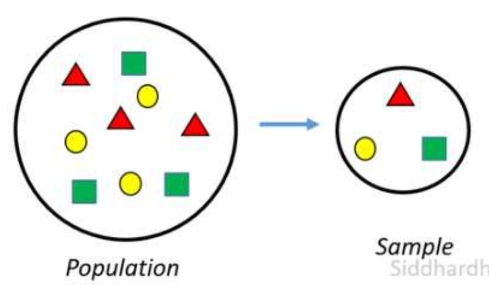
Descriptive Statistics

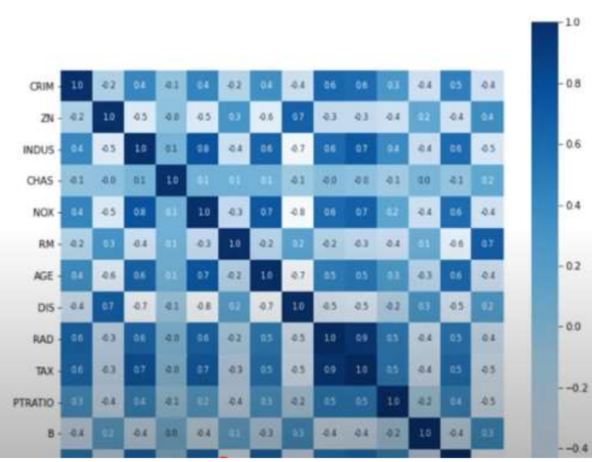
2 important measures of Descriptive Statistics:

1. Measure of Central Tendencies (Mean, Median, Mode)
2. Measure of Variability (Range, Standard Deviation, Variance)

Inferential Statistics

***Inferential statistics** takes data from a sample and makes inferences and predictions about the larger population from which the sample was drawn.*





Correlation of House Price Data

Types of Statistical Studies

Statistical Study

Sample Study

A **sample study** is a study which is carried out on a sample which represents the total population.

Observational Study

An **observational study** is a study where we simply collect and analyze data. We won't inject any changes. We just observe the correlation in the data.

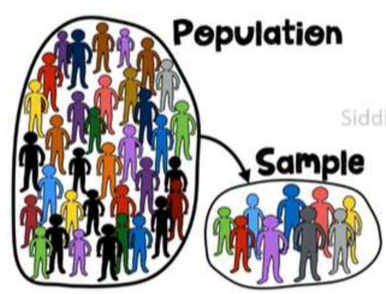
Experimental Study

An **experimental study** is a **study** in which conditions are controlled and manipulated by the experimenter.

Types of Sampling Techniques

Sampling Techniques:

- Simple Random Sampling
- Systematic Sampling
- Stratified Random Sampling
- Cluster Sampling



Simple Random Sampling

In **Simple Random Sampling**, the sample is randomly picked from a larger population. Hence, all the individual datapoints has an equal probability to be selected as sample data.

Example: Employee survey in a company

Pros:

1. No sample Bias
2. Balanced Sample
3. Simple Method of sampling
4. Requires less domain knowledge

Cons:

1. Population size should be high
2. Cannot represent the population well sometimes

Systematic Sampling

In **Systematic Sampling**, the sample is picked from the population at regular intervals. This type of sampling is carried out if the population is homogeneous and the data points are uniformly distributed

Example: Selecting every 10^{th} member from a population of 10,000

Pros:

1. Quick & easy
2. Less bias
3. Even distribution of data

Cons:

1. Data manipulation risk
2. Requires randomness in data
3. Population should not have patterns.

Stratified Random Sampling

In **Stratified Random Sampling**, the population is subdivided into smaller groups called **Strata.** Samples are obtained randomly from all these strata.

Example: Smartphone sales in all the states

Pros:

1. Finds important characteristics in the population
2. High precision can be obtained if the differences in the strata is high

Cons:

1. Cannot be performed on populations that cannot be classified into groups.
2. Overlapping data points

Cluster Sampling

Cluster Sampling is carried out on population that has inherent groups. This population is subdivided into **clusters** and then random clusters are taken as sample.

Example: Smartphone sales in randomly selected states

Pros:

1. Requires only fewer resources
2. Reduced Variability
3. Advantages of both Random sampling and Stratified Sampling

Cons:

1. Cannot be performed on populations without natural groups
2. Overlapping data points
3. Can't provide a general insight for the entire population

Central Tendencies

Mean

Mean *or arithmetic mean is the sum of values divided by the number of values.*

$$M = \frac{\Sigma x}{N}$$

Heights

160
172
165
168
174

$$\frac{160+172+165+168+174}{5}$$

Mean = 167.8

Median

The ***median*** *is the* ***middle*** *value in the list of numbers. To find the median, the numbers have to be listed in numerical order from smallest to largest.*

160 165 168 172 174

160 165 $\boxed{168 \quad 172}$ 174 176

$$\frac{168+172}{2} = 170$$

Median = 170

Mode

The ***mode*** *is the value that occurs most often. If no number in the list is repeated, then there is no mode for the list.*

Heights

160
172
160
168
174

Mode = 160

Central Tendencies in Data Pre-Processing

Central Tendencies are very useful in **handling the missing values** in a dataset

Mean : Missing values in a dataset can be replaced with **mean** value, if the data is uniformly distributed.

Median : Missing values in a dataset can be replaced with **median** value, if the data is skewed.

Mode : Missing values in a dataset can be replaced with **mode** value, if the data is skewed. Missing categorical values can also be replaced with **mode** value.

Measure of Variability

Measure of Variability

Range

*The **range** of a set of data is the difference between the largest and smallest values. It can give a rough idea about the distribution of our dataset.*

Variance

***Variance** is a measure of how far each number in the set is from the mean and therefore from every other number in the dataset.*

Standard Deviation

Standard Deviation is the square root of Variance. Standard deviation looks at how spread out a group of numbers is from the mean.

Range ; Variance ; Standard Deviation

-5, 0, 5, 10, 15,

$$Mean = \frac{-5+0+5+10+15}{5} = 5$$

Range = 15 – (-5) = 20

$$Variance = \frac{(-5-5)^2 + (0-5)^2 + (5-5)^2 + (10-5)^2 + (15-5)^2}{5}$$

$$Variance = 50$$

$$Standard\ Deviation = 7.1$$

3, 4, 5, 6, 7

$$Mean = \frac{3+4+5+6+7}{5} = 5$$

Range = 7 – 3 = 4

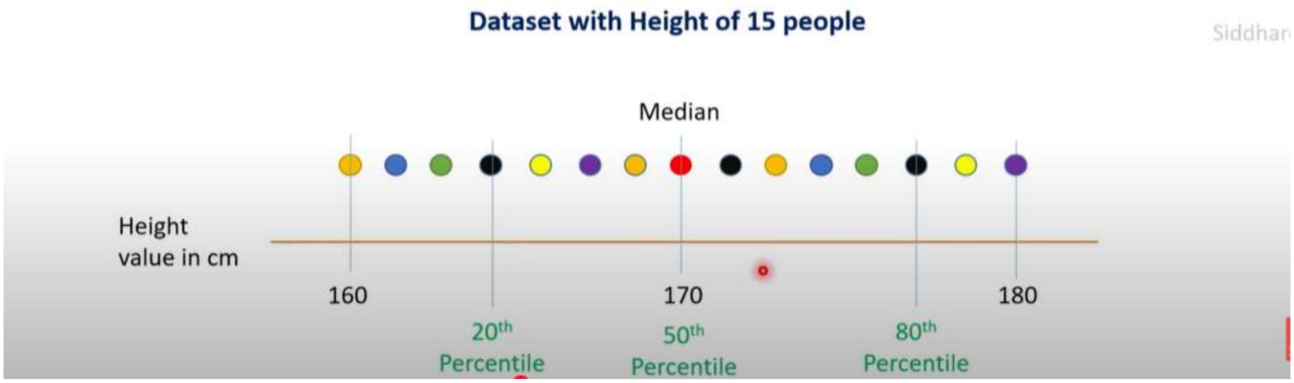
$$Variance = \frac{(3-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2}{5}$$

$$Variance = 2$$

$$Standard\ Deviation = 1.4$$

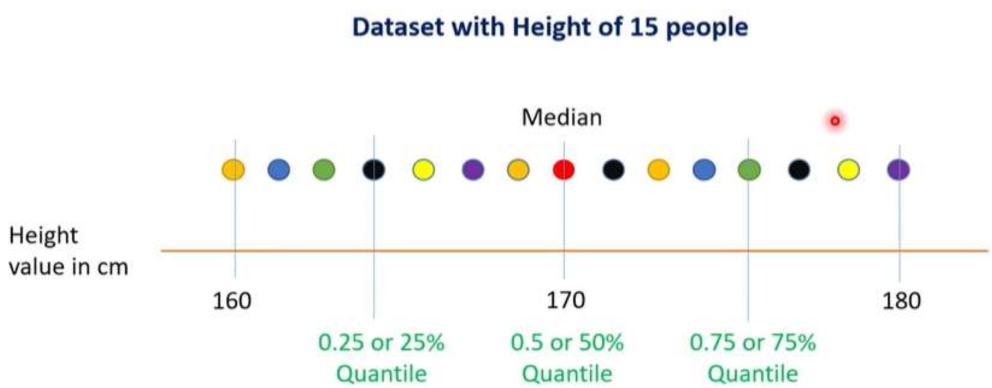
Percentiles

Percentile is a value on a scale of 100 that indicates the percent of a distribution that is equal to or below it.



Quantiles

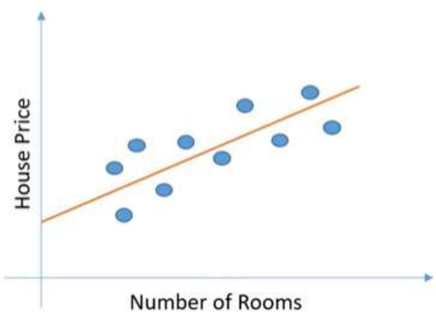
Quantile is a measure that tells how many values in a dataset are above or below a certain limit. It divides the members of the dataset into equally-sized subgroups.



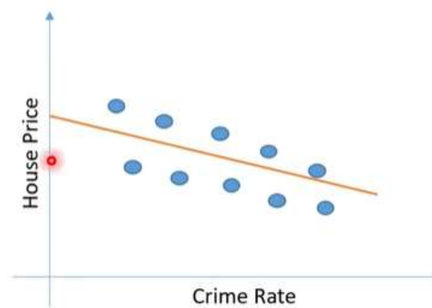
Correlation

Correlation is a measure that determines the extent to which two variables are related to each other in a dataset. But it doesn't mean that one event is the cause of the other event.

Positive Correlation

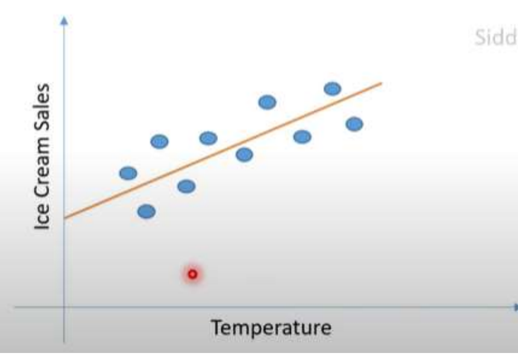


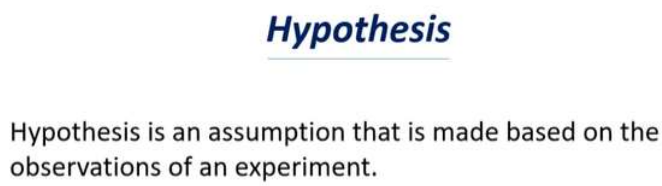
Negative Correlation

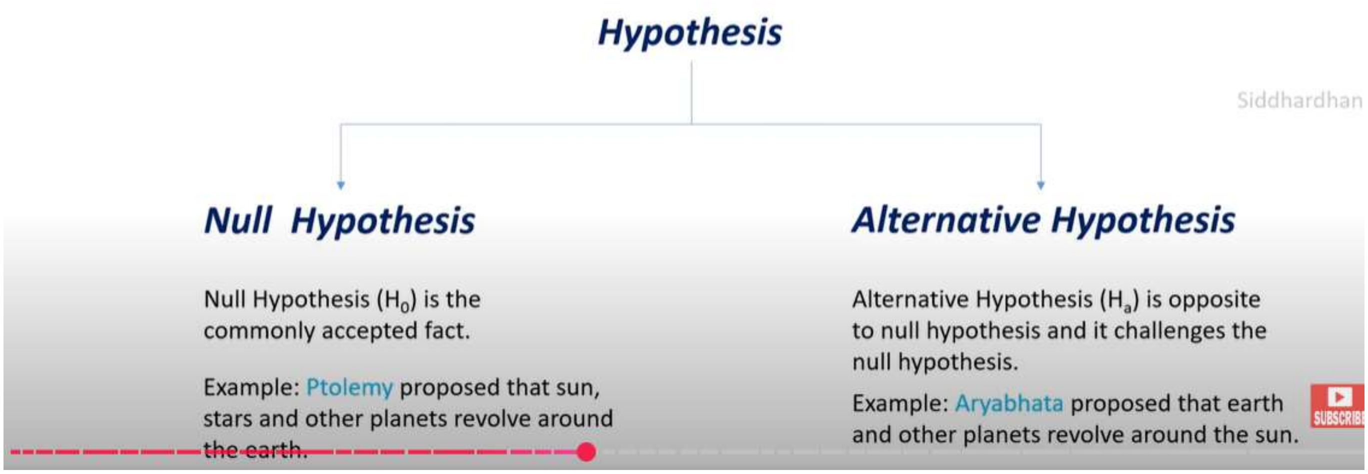


Causation

In statistics, Causation means that one event causes another event to occur. Thus, there is a cause and effect relationship between the two variables in a dataset.

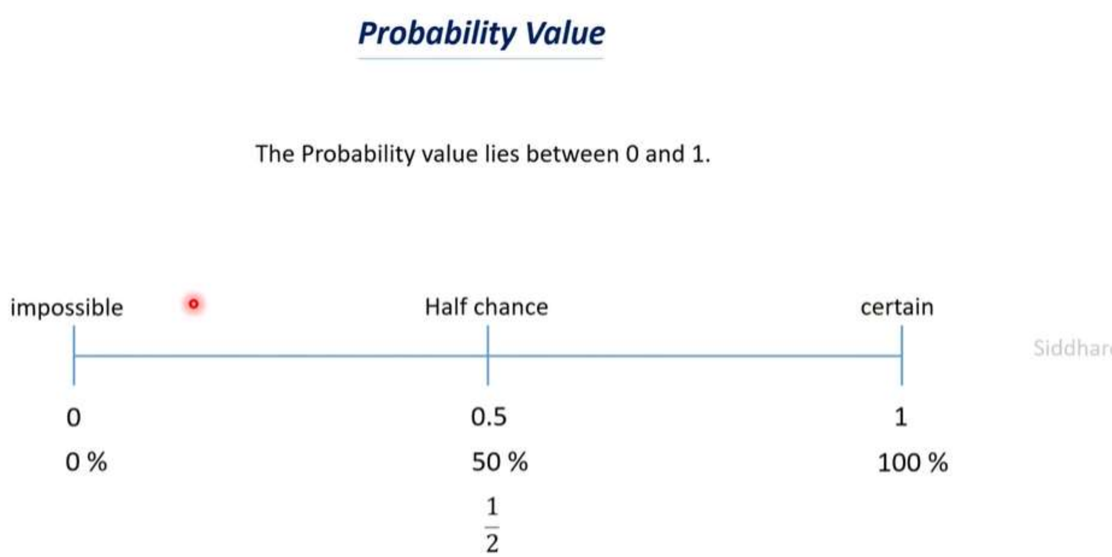






Probability in Machine Learning

Definition : Likelihood of given event to occur



Basics of Probability

$$\text{Probability of an event to occur} = \frac{\text{Number of ways an event can occur}}{\text{Total number of outcomes}}$$

Head Tail

(H, T)

Possible Outcomes

$P(H) = \frac{1}{2}$ $P(T) = \frac{1}{2}$

(1, 2, 3, 4, 5, 6)

Possible Outcomes

$P(5) = \frac{1}{6}$ $P(even) = \frac{3}{6}$

$P(5) = 0.16$ $P(even) = 0.5$

Random Variables

A Random Variable is a numerical description of the outcomes of Random events.

In other words, a random variable maps the outcomes of random events to numerical values.

Consider Tossing a Coin

Random Variable

$$X = \begin{cases} 0 & \leftarrow \\ 1 & \leftarrow \end{cases}$$

Head

Tail

Y = Weight of a random person in a class

P (Weight of a random person in a class is less than 60 kg)

P (Y < 60)

Random Variables

Discrete

A discrete random variable takes only discrete or distinct values.

Examples: Coin toss, Colour of the ball.

Continuous

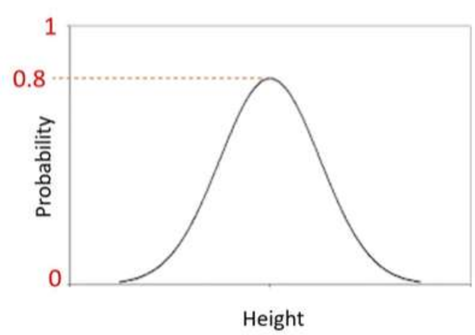
A continuous random variable can take any value in a given range.

Examples: weight of a random person in a class.

Normal Distribution

A **normal distribution** is an arrangement of a data set in which most of the data points lie in the middle of the range and the rest taper off symmetrically toward either extreme.

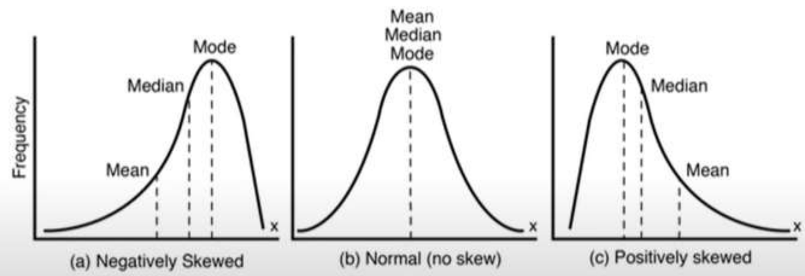
Normal Distribution is also known as **Gaussian Distribution**.



Bell Shaped Curve

Skewness

A data is considered **skewed** when the distribution curve appears distorted or skewed either to the left or to the right, in a statistical distribution.



Poisson Distribution

Poisson Distribution is a probability distribution that measures how many times an event is likely to occur within a specified period of time.

Poisson distribution is used to understand independent events that occur at a constant rate within a given interval of time.

Examples of Poisson Distribution

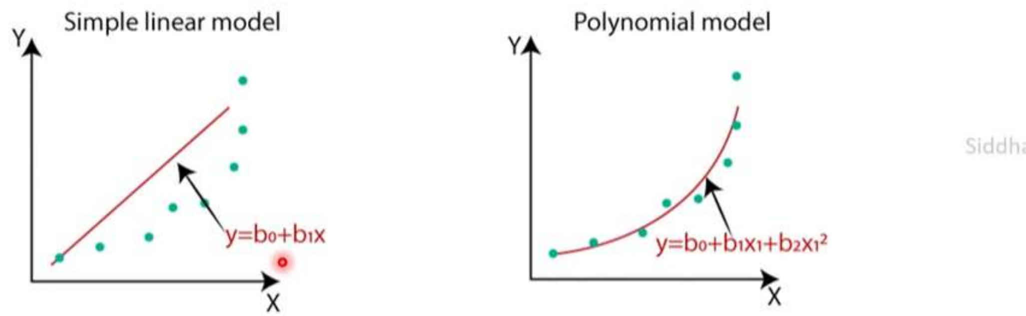
- Number of accidents occurring in a city from 6 pm to 10 pm
- Number of Patients arriving in an Emergency Room between 10 pm to 12 pm
- How many views does your blog gets in a day

Machine Learning Model

A **Machine Learning Model** is a function that tries to find the relationship between the Features and the Target variable.

It tries to find the pattern in the data, understand the data and trains on the data. Based on this learning, a Machine Learning Model makes Predictions and recognize patterns.

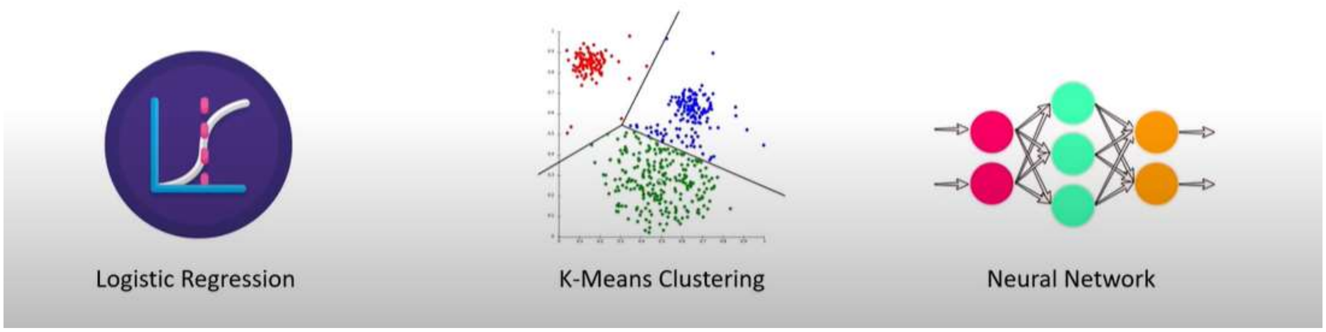
Machine Learning Model



We cannot have a linear relationship between the variables all the time.

Model Selection

Model Selection in Machine Learning is the process of choosing the best suited model for a particular problem. Selecting a model depends on various factors such as the dataset, task, nature of the model, etc.



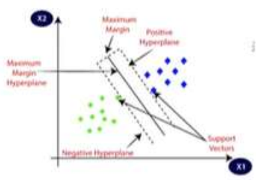
Model Selection

Models can be selected based on :

1. Type of Data available:
    a. Images & Videos – CNN
    b. Text data or Speech data – RNN
    c. Numerical data – SVM, Logistic Regression, Decision trees, etc.

2. Based on the task we need to carry out:
    a. Classification tasks – SVM, Logistic Regression, Decision trees, etc.
    b. Regression tasks – Linear regression, Random Forest, Polynomial regression, etc.
    c. Clustering tasks – K-Means Clustering, Hierarchical Clustering

Cross Validation



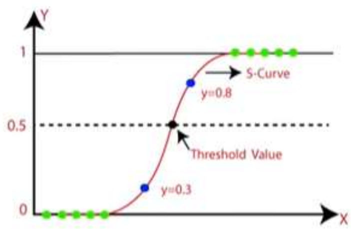
Support Vector Machine

	Dataset					*Accuracy*
Iteration 1	Train	Train	Train	Train	Test	88%
Iteration 2	Train	Train	Train	Test	Train	83%
Iteration 3	Train	Train	Test	Train	Train	86%
Iteration 4	Train	Test	Train	Train	Train	81%
Iteration 5	Test	Train	Train	Train	Train	84%

$$\text{Mean Accuracy} = \frac{88 + 83 + 86 + 81 + 84}{5} = 84.4\ \%$$

6:42:02 / 11:41:33 • Model Selection & Cross Validation

Cross Validation



Logistic Regression

	Dataset					Accuracy
Iteration 1	Train	Train	Train	Train	Test	90%
Iteration 2	Train	Train	Train	Test	Train	88%
Iteration 3	Train	Train	Test	Train	Train	86%
Iteration 4	Train	Test	Train	Train	Train	91%
Iteration 5	Test	Train	Train	Train	Train	85%

Mean Accuracy = $\frac{90 + 88 + 86 + 91 + 85}{}$ = 88 %

Cross Validation

✓ *Accuracy score for SVM = 84.4 %*

✓ *Accuracy score for Logistic Regression = 88 %*

Choose the best model with higher accuracy

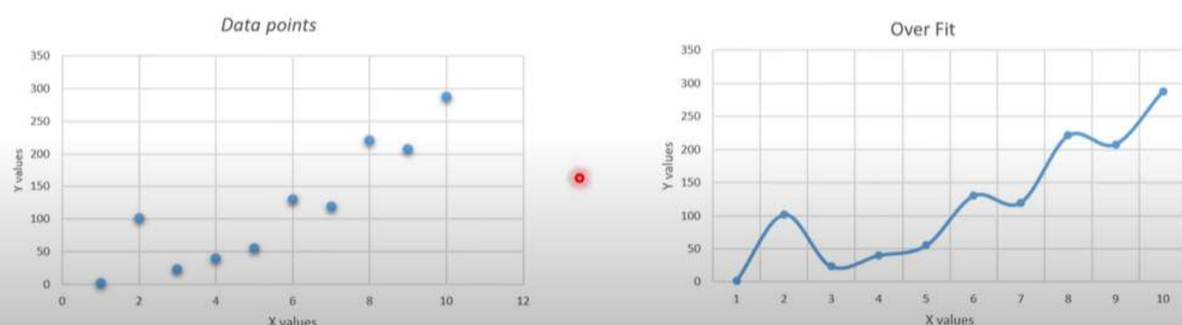
Overfitting

Overfitting refers to a model that models the training data too well. Overfitting happens when a model learns the detail and noise in the training dataset to the extent that it negatively impacts the performance of the model.

Sign that the model has Overfitted : High Training data Accuracy & very low Test data Accuracy

Overfitting

X	1	2	3	4	5	6	7	8	9	10
Y	1.38	101.41	23.34	39.89	55.23	129.91	119.33	221.09	207.43	287.80



Underfitting

Underfitting happens when the model **does not learn enough** from the data. Underfitting occurs when a machine learning model cannot capture the underlying trend of the data

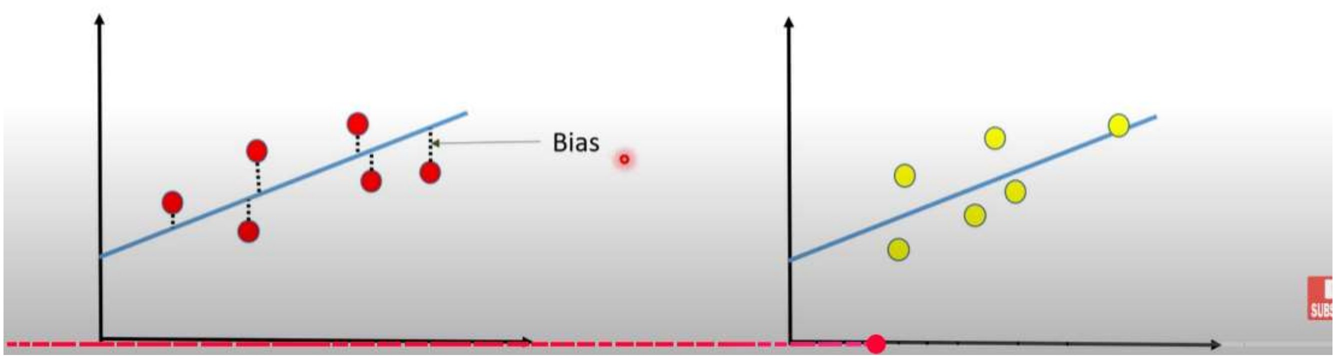
Sign: Low training data accuracy

To solve the problem of underfitting and overfitting we use Bias - Variance Tradeoff

Bias – Variance Tradeoff

Bias :

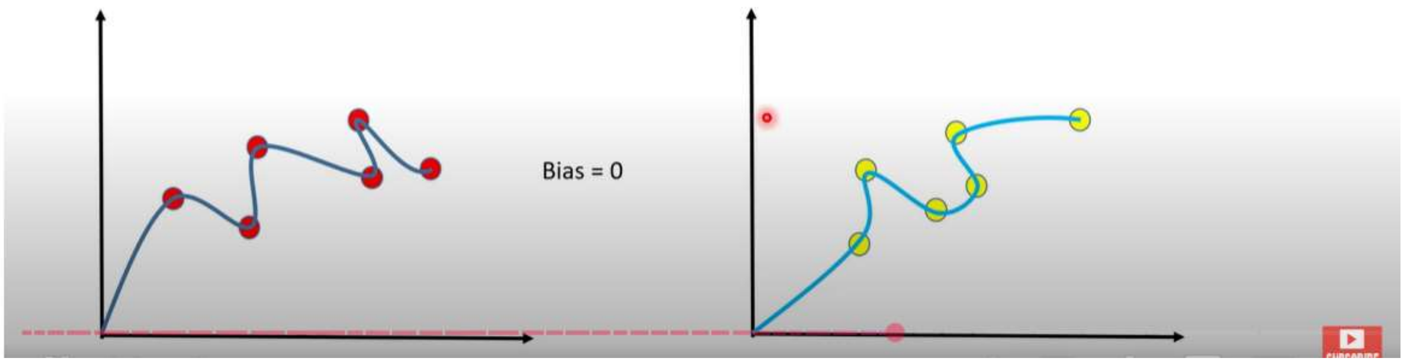
Bias is the difference between the average prediction of our model and the correct value which we are trying to predict.



Bias – Variance Tradeoff

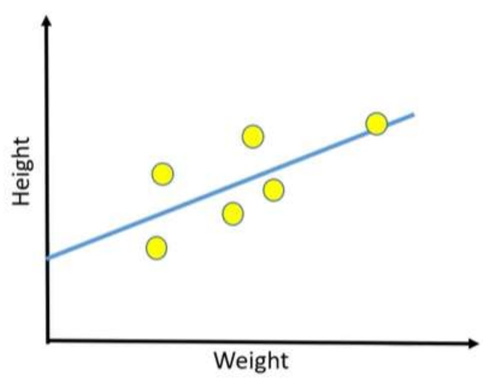
Variance :

Variance is the amount that the estimate of the target function will change if different training data was used.



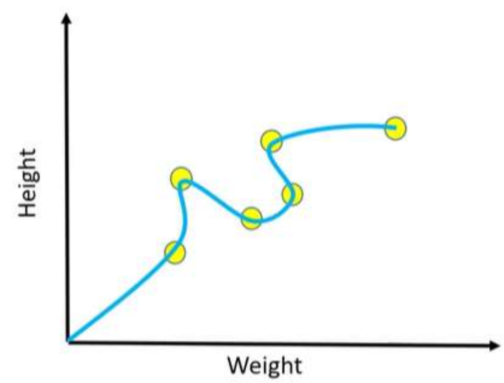
Bias – Variance Tradeoff

(Testing with different data)



(i) Underfitting

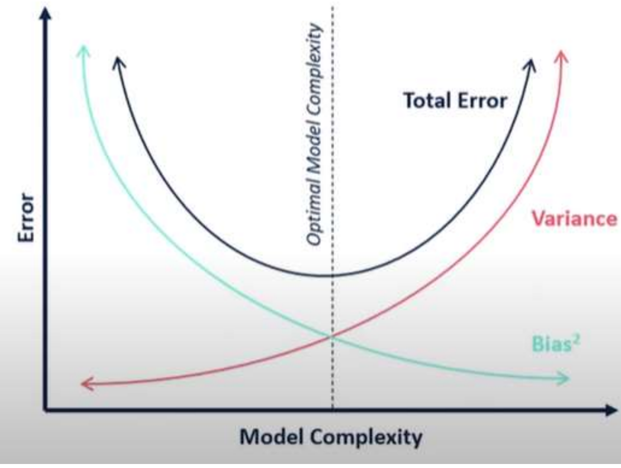
Inference: *a. High Bias*
b. Low Variance



(ii) Overfitting

Inference: *a. Low Bias*
b. High Variance

Bias – Variance Tradeoff

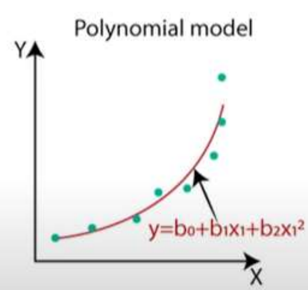


Left side: Underfitting
Right side: Overfitting

Bias – Variance Tradeoff

Techniques to have better Bias – Variance Tradeoff :

1. *Good Model Selection*
2. *Regularization*
3. *Dimensionality Reduction*
4. *Ensemble methods*



Loss Function

Loss function measures how far an estimated value is from its true value.

It is helpful to determine which model performs better & which parameters are better.

$$Loss = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Types of Loss Function:

- Cross Entropy Loss
- Squared Error Loss
- KL Divergence

Low loss function value = High Accuracy

Model Evaluation

Supervised Learning

Classification

Classification is about predicting a class or discrete values
Eg: Male or Female; True or False

Evaluation metric for Classification: **Accuracy score**

Regression

Regression is about predicting a quantity or continuous values
Eg: Salary; age; Price.

Evaluation metric for Regression: **Mean Absolute Error**

Accuracy Score

In Classification, **Accuracy Score** is the ratio of **number of correct predictions** to the **total number of input data points**.

$$\text{Accuracy Score} = \frac{\text{Number of correct predictions}}{\text{Total Number of data points}} \times 100\ \%$$

Mean Squared Error for calculating the Regression Accuracy

Mean Squared Error

Mean Squared Error measures the average of the squares of the errors, that is, the average squared difference between the estimated values and the actual value.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y_i})^2$$

Actual Value (Y_i = 140 mg/dL) Predicted Value (Y_i = 160 mg/dL)

Model Parameters

Types of Parameters

Parameters

Model Parameters

These are the parameters of the model that can be determined by training with training data. These can be considered as internal Parameters.

- **Weights**
- **Bias**

Y = w*X + b

Hyperparameters

Hyperparameters are parameters whose values control the learning process. These are adjustable parameters used to obtain an optimal model. External Parameters.

- **Learning rate**
- **Number of Epochs**

Model Parameters

Weights: Weight decides how much influence the input will have on the output.

Applicant's Details

Name	Degree	College	C	C++	Python	Height	Weight	No. of Backlogs
A	B.E	ABC college	✓	✗	✓	165	72	1
B	M.E	XYZ College	✓	✓	✗	168	80	0
C	M.C.A	State College	✓	✗	✗	175	67	0
D	B.E	ZYX College	✓	✓	✓	168	70	2
✗	✓	✓	✓	✓	✓	✗	✗	✓

Name, Height and weight will have 0 weight value
But the other fields have a value where an important field will have a high weight value
As other fields can help us determine whether the person should be hired for SWE role

Model Parameters

Weights:

Weight decides how much influence the input will have on the output.

$$Y = w*X + b$$

$$Y = w_1*X_1 + w_2*X_2 + w_3*X_3 + b$$

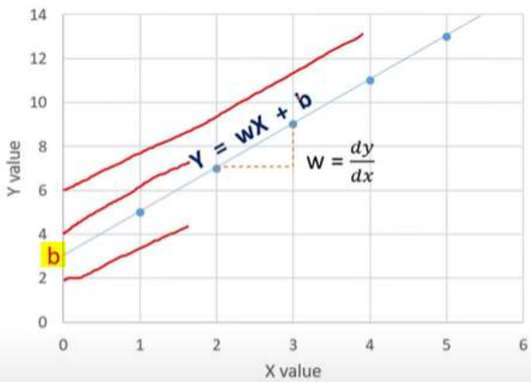
X – feature or input variable
Y – Target or output variable
w – weight
b – bias

Bias:

Bias is the offset value given to the model. Bias is used to shift the model in a particular direction. It is similar to a Y-intercept. 'b' is equal to 'Y' when all the feature values are zero.

Bias is the Y-intercept

Linear Regression



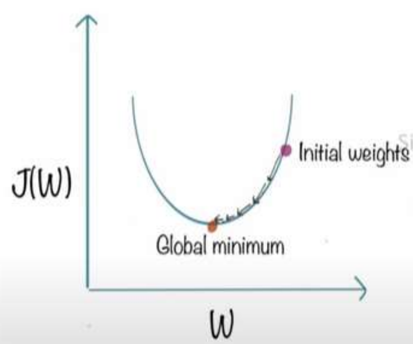
Y = wX + b

X --> X value
Y --> Y value
w --> weight
b --> bias

Hyperparameters

Learning Rate:

The **Learning Rate** is a tuning parameter in an optimization algorithm that determines the step size at each iteration while moving toward a minimum of a loss function.



Gradient Descent

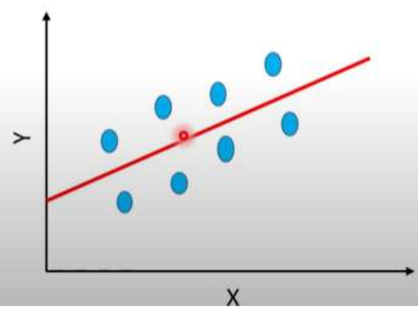
Number of Epochs:

Number of Epochs represents the number of times the model iterates over the entire dataset.

Learning rate is basically the step size to reduce or increase weight to reduce the loss function
Number of Epochs is no. of times the model is trained on the entire dataset

Model Optimization

Optimization refers to determining best parameters for a model, such that the loss function of the model decreases, as a result of which the model can predict more accurately.

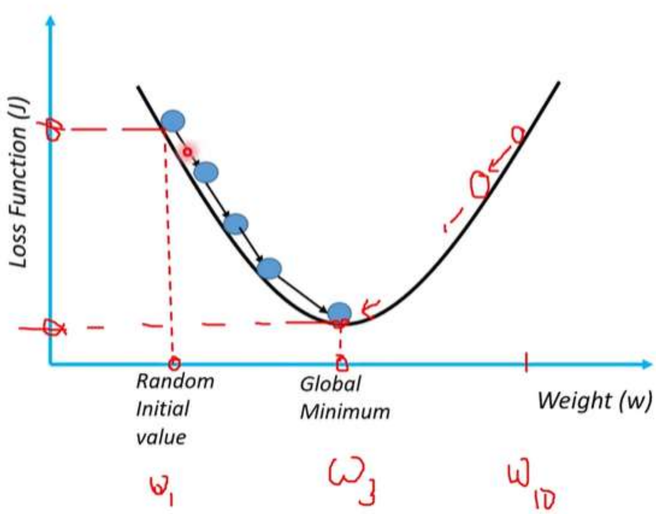


$Y = w_3X + b_3$

Hence, w_3 & b_3 are the best parameters

Gradient descent is a model optimization which is used to reduce the loss function

Gradient Descent



Gradient Descent

Gradient Descent is an optimization algorithm used for minimizing the loss function in various machine learning algorithms. It is used for updating the parameters of the learning model.

w = w - L*dw

b = b - L*db

w --> weight
b --> bias
L --> Learning Rate

dw --> Partial Derivative of loss function with respect to ~~m~~ w

db --> Partial Derivative of loss function with respect to ~~c~~ b