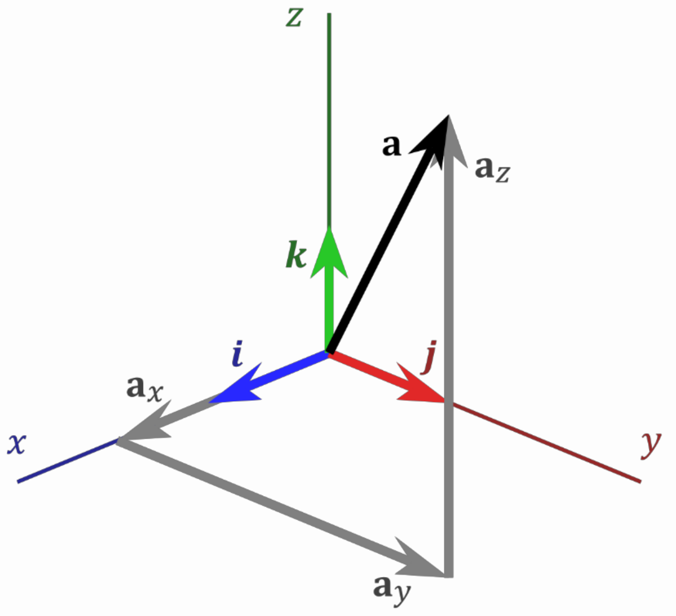
Vector Operations

Math for Machine Learning



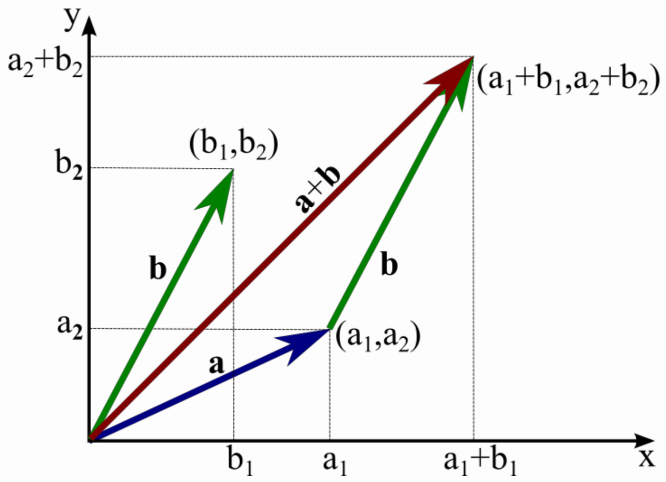
Vector Operations

1. Vector Addition

2. Vector Subtraction

3. Multiplying a vector by a Scalar

4. Angle between 2 Vectors



Vectors – Computer Science Approach

Scalar

24

Vector

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

row

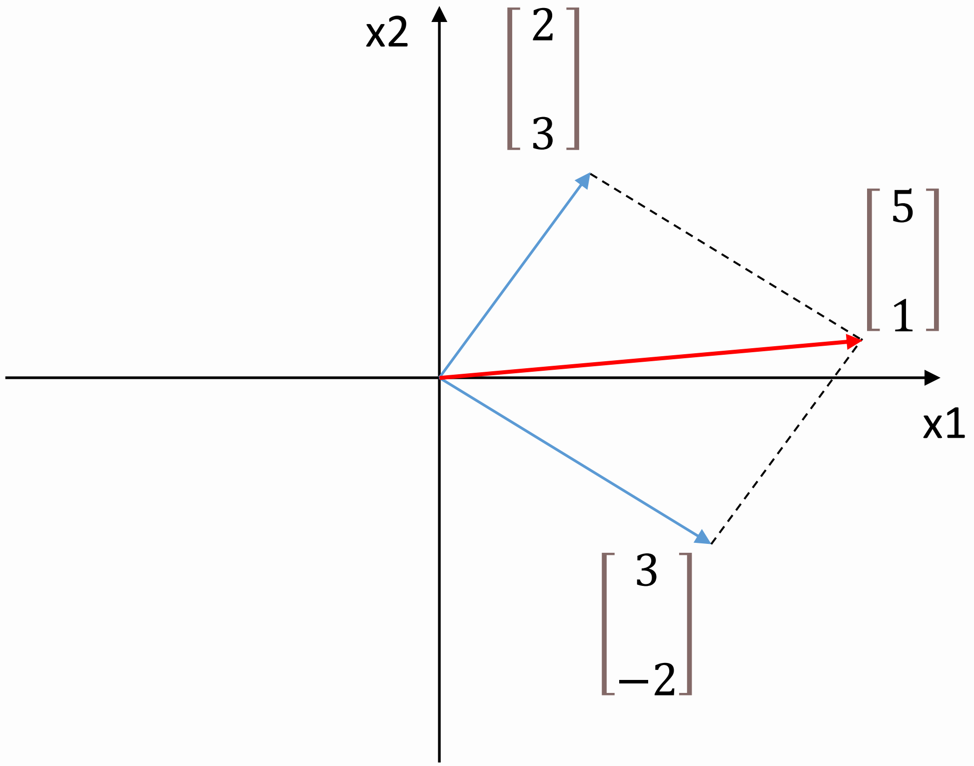
or

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

[89 66 23 94 28.1]

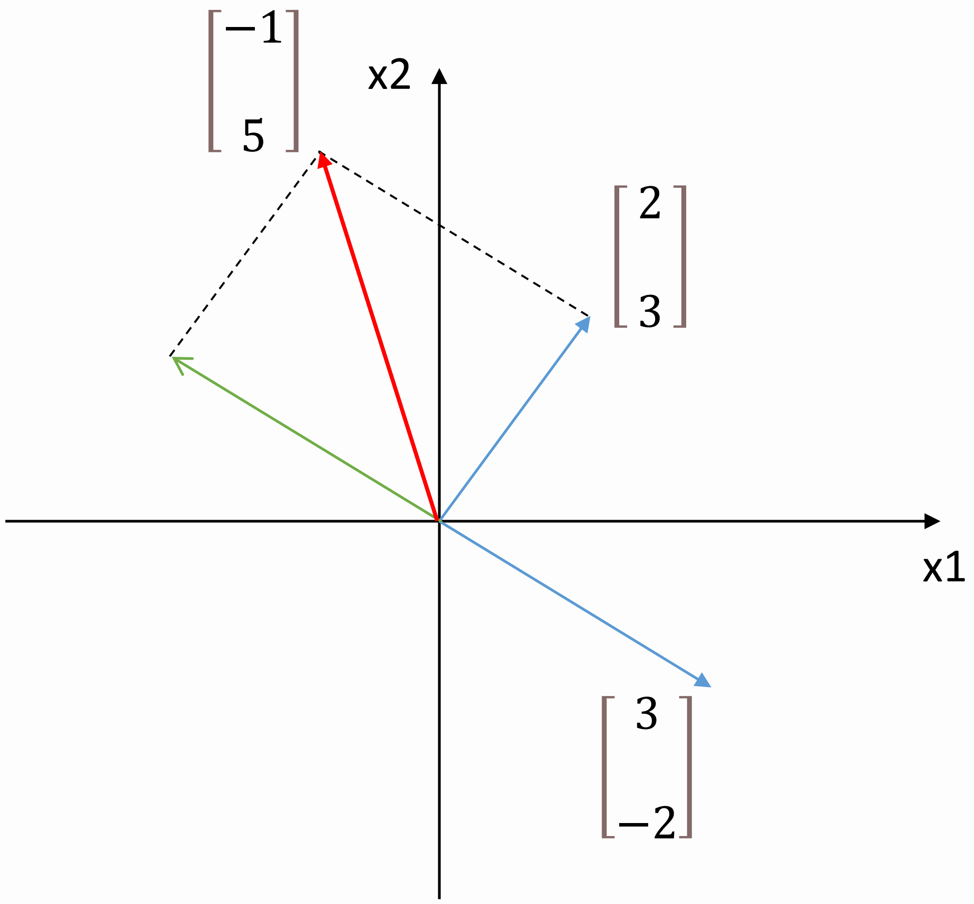
(Glucose, Blood Pressure, Skin Thickness, Insulin, BMI)

Vector Addition



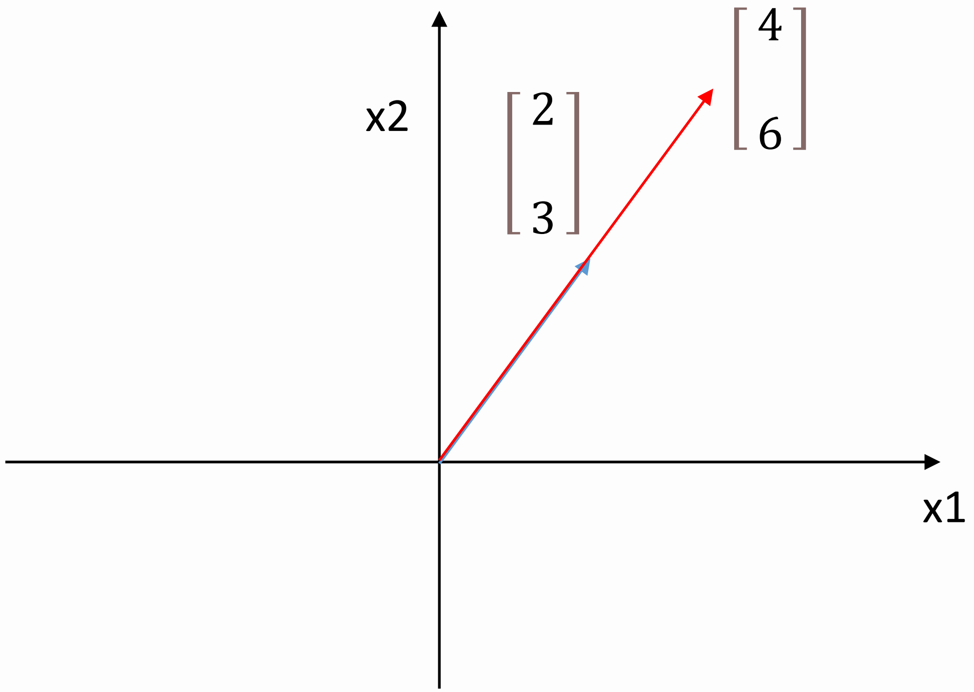
$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 2+3 \\ 3+(-2) \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

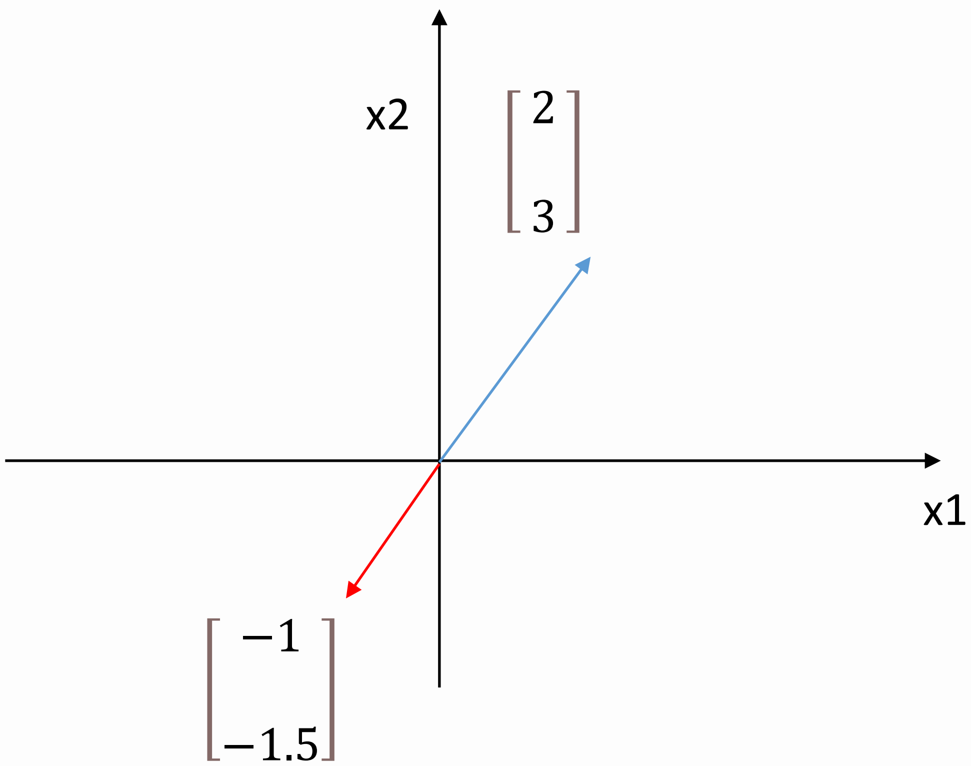
Vector Subtraction



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

Multiplying a vector by a Scalar

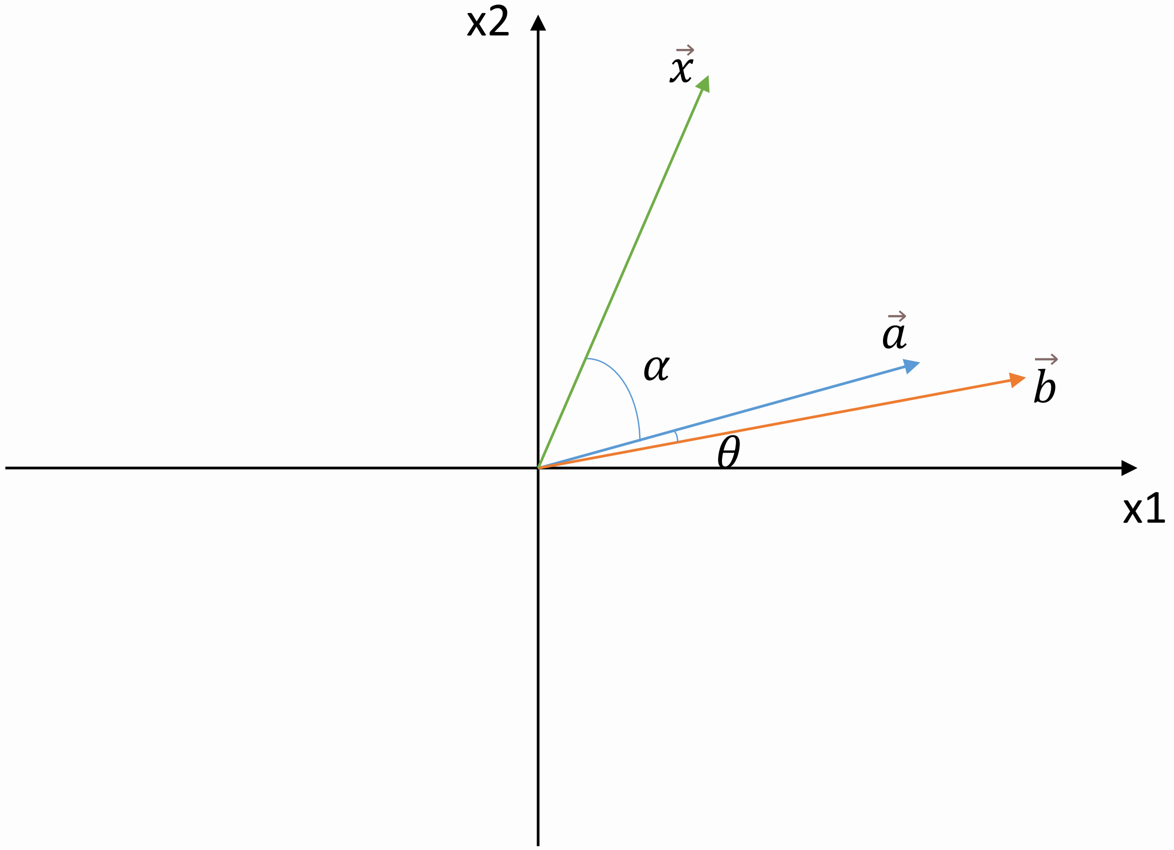




$$2 \times \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

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

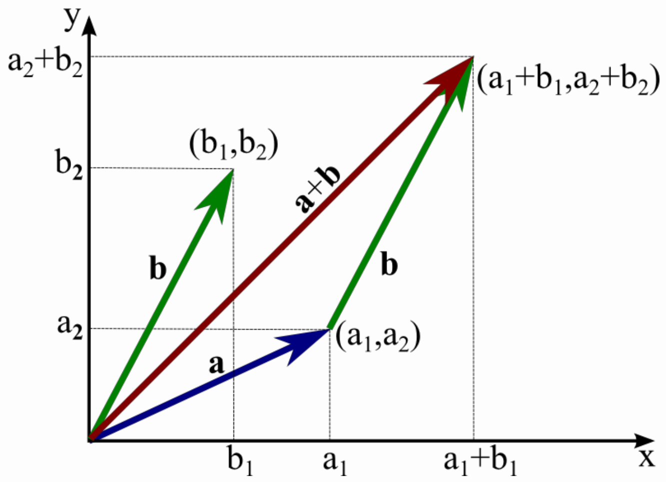
Angle Between 2 Vectors



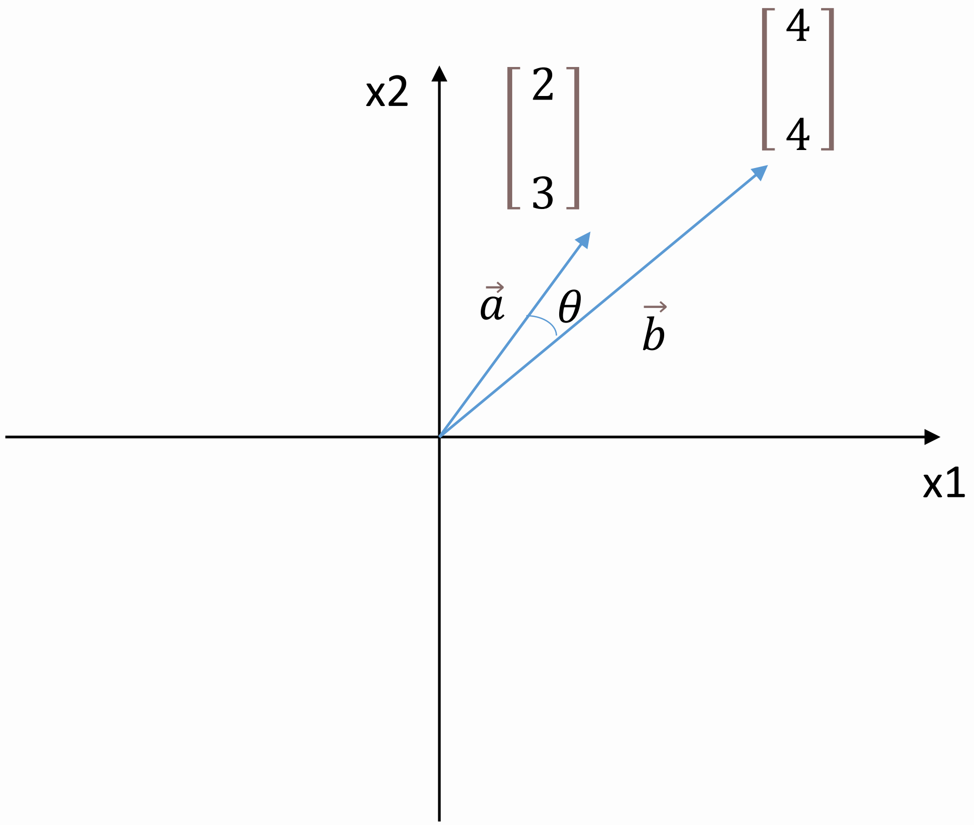
Inference:

- ✓ 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.

1. Dot Product of 2 Vectors
2. Cross Product of 2 Vectors
3. Projection of vector



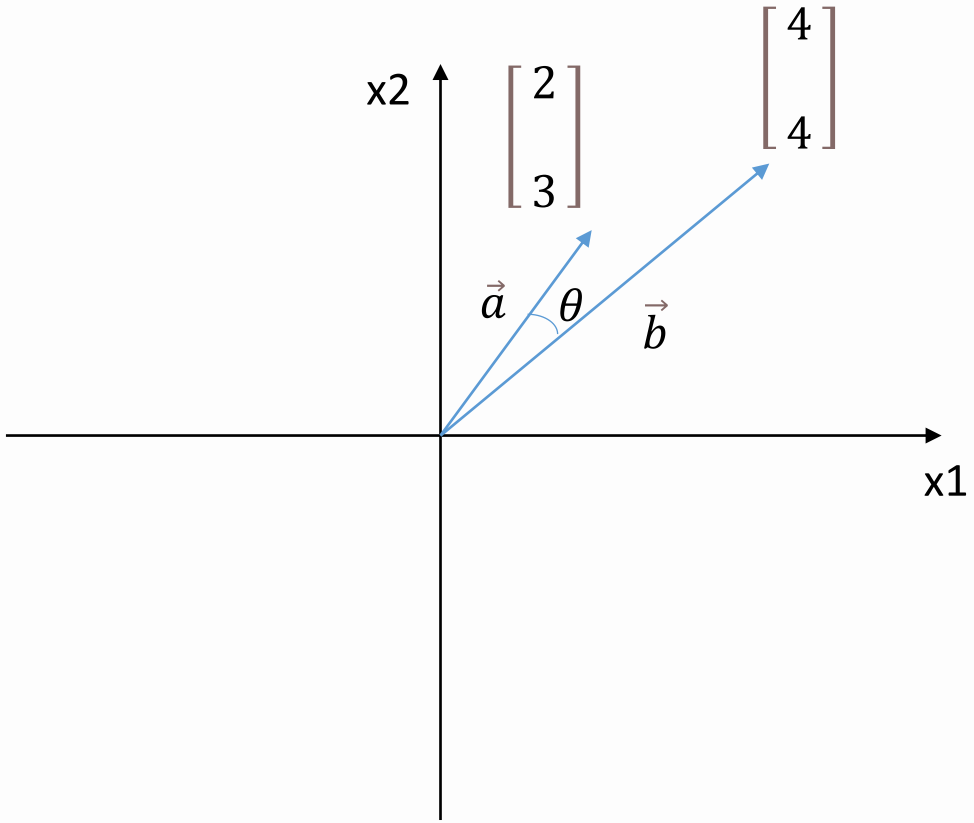
Dot Product of 2 Vectors



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

$$\overrightarrow{a} . \overrightarrow{b} = |\vec{a}|\left|\vec{b}\right| \cos\theta$$

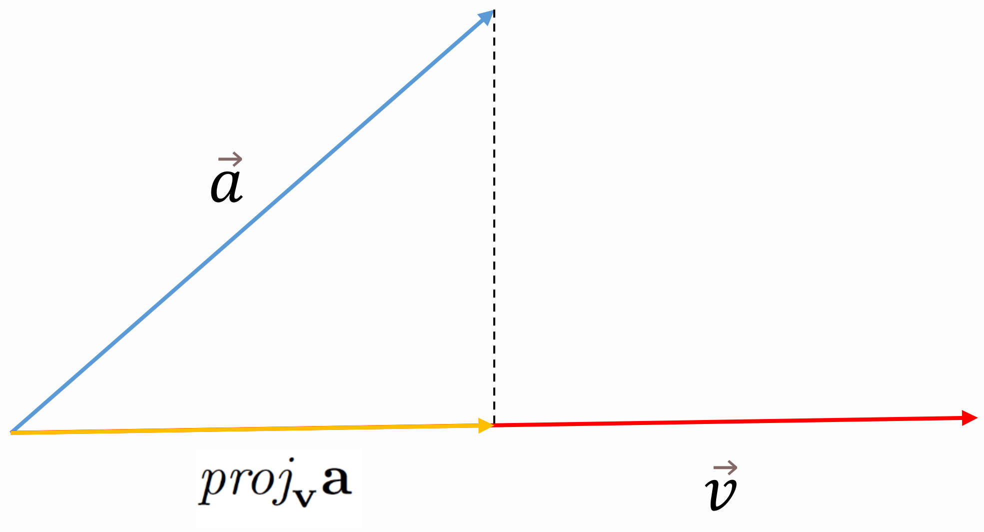
Cross Product of 2 Vectors



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

$$\overline{a} \times \overline{b} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 2 & 3 & 0 \\ 4 & 4 & 0 \end{vmatrix} = \boldsymbol{i}(3\cdot0 - 0\cdot4) - \boldsymbol{j}(2\cdot0 - 0\cdot4) + \boldsymbol{k}(2\cdot4 - 3\cdot4) = \boldsymbol{i}(0 - 0) - \boldsymbol{j}(0 - 0) + \boldsymbol{k}(8 - 12) = \{0; 0; -4\}$$

Projection of Vector



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

Matrix – Basics (Linear Algebra)

Math for Machine Learning

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

Matrix - Basics

1. Scalars; Vectors; Matrix
2. Shape of a Matrix
3. Different Types of Matrix
4. Transpose of a Matrix
5. Role of Matrix in Machine Learning

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

Scalars; Vectors; Matrix

Scalar

24

Vector

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

row

or column

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

Matrix

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

Shape of a Matrix

$$\begin{bmatrix} 2 & 5 \\ 4 & 8 \end{bmatrix}$$

2 x 2 Matrix

$$\begin{bmatrix} 8 & 6 & 1 \\ 2 & 9 & 2 \\ 3 & 4 & 3 \end{bmatrix}$$

3 x 3 Matrix

$$\begin{bmatrix} 2 & 3 \\ 6 & 4 \\ 7 & 8 \end{bmatrix}$$

3 x 2 Matrix

General Matrix Notation :

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

m x n Matrix

a_{ij} → Matrix element

i → Row number

j → Column number

Different Types of Matrices

Null Matrix or Zero Matrix :

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

2 x 2

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

3 x 3

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

4 x 4

Identity Matrix :

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

2 x 2

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

3 x 3

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

4 x 4

Transpose of a Matrix

Transpose of a matrix is formed by turning all the rows of a given matrix into columns and vice-versa

$$A = \begin{bmatrix} 2 & 5 \\ 4 & 8 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 2 & 4 \\ 5 & 8 \end{bmatrix}$$

$$B = \begin{bmatrix} 8 & 6 & 1 \\ 2 & 9 & 2 \\ 3 & 4 & 3 \end{bmatrix}$$

$$B^T = \begin{bmatrix} 8 & 2 & 3 \\ 6 & 9 & 4 \\ 1 & 2 & 3 \end{bmatrix}$$

Matrix Operations

1. Matrix Addition
2. Matrix Subtraction
3. Multiplying a Matrix by a Scalar
4. Multiplying 2 Matrices

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

Scalars; Vectors; Matrix

Scalar

24

Vector

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

row

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

Matrix

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

Matrix Addition

Rule : Two Matrices can be added only if the have the same shape, that is, both the matrix should have the same number of rows and columns

$$\begin{bmatrix} 2 & 3 \\ 10 & 5 \end{bmatrix} + \begin{bmatrix} 10 & 5 \\ 20 & 4 \end{bmatrix} = \begin{bmatrix} 12 & 8 \\ 30 & 9 \end{bmatrix}$$

2 x 2   2 x 2   2 x 2

$$\begin{bmatrix} 2 & 1 \\ 4 & 2 \\ 6 & 3 \end{bmatrix} + \begin{bmatrix} 5 & 2 \\ 3 & 6 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 7 & 3 \\ 7 & 8 \\ 8 & 8 \end{bmatrix}$$

3 x 2   3 x 2   3 x 2

Matrix Subtraction

Rule : Two Matrices can be subtracted only if they have the same shape, that is, both the matrix should have the same number of rows and columns

$$\begin{bmatrix} 2 & 3 \\ 10 & 5 \end{bmatrix} - \begin{bmatrix} 10 & 5 \\ 20 & 4 \end{bmatrix} = \begin{bmatrix} -8 & -2 \\ -10 & 1 \end{bmatrix}$$

2 x 2 2 x 2 2 x 2

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

3 x 2 3 x 2 3 x 2

Multiplying a Matrix by a Scalar

$$5 \times \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 5 \times 2 \\ 5 \times 4 \\ 5 \times 6 \end{bmatrix} = \begin{bmatrix} 10 \\ 20 \\ 30 \end{bmatrix}$$

3 x 1 3 x 1

$$5 \times \begin{bmatrix} 2 & 1 \\ 4 & 2 \\ 6 & 3 \end{bmatrix} = \begin{bmatrix} 10 & 5 \\ 20 & 10 \\ 30 & 15 \end{bmatrix}$$

3 x 2 3 x 2

Note : Vectors are a type of Matrix with either one row or one column

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

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

2 x 2   2 x 2

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

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

3 x 2   3 x 2

Cannot be multiplied.

Multiplying 2 Matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}$$

$$\begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix} \text{ x } \begin{bmatrix} 5 & 6 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 2x5 + 4x3 & 2x6 + 4x4 \\ 3x5 + 6x3 & 3x6 + 6x4 \end{bmatrix} = \begin{bmatrix} 22 & 28 \\ 33 & 42 \end{bmatrix}$$

2 x 2   2 x 2   2 x 2

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

$$\begin{bmatrix} 10 & 5 \\ 20 & 10 \\ 30 & 15 \end{bmatrix} + \begin{bmatrix} 10 & 5 \\ 20 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 \\ 4 & 2 \\ 6 & 3 \end{bmatrix} + \begin{bmatrix} 5 & 2 \\ 3 & 6 \\ 2 & 5 \end{bmatrix} = ? \quad \begin{bmatrix} 7 & 3 \\ 8 & 7 \\ 5 & 11 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 5 & 6 \\ 3 & 2 & 4 \\ 0 & 1 & 2 \\ 5 & 4 & 6 \end{bmatrix} \times \begin{bmatrix} 5 & 2 & 4 \\ 1 & 0 & 3 \\ 2 & 7 & 1 \end{bmatrix} = ?$$

$$\begin{bmatrix} 22 & 44 & 20 \\ 22 & 14 & 22 \\ 15 & 18 & 25 \\ 41 & 40 & 16 \end{bmatrix}$$

Statistics for Machine Learning

Math for Machine Learning

What is Statistics ?

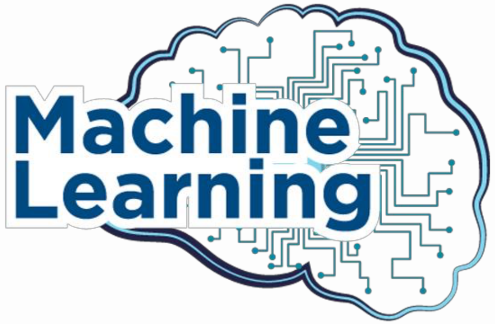
Statistics is the science concerned with developing and studying methods for collecting, analysing, interpreting and presenting data.

Correlation

Statistical Measures:

1. Range
2. Mean
3. Standard Deviation

Role of Statistics in ML



Data

Machine Learning Models

Topics covered in this module:

1. Basics of Statistics
2. Types of Statistics
3. Population & Sample
4. Central Tendencies
5. Percentiles & Dispersion
6. Statistics Implementation with Python -1
7. Range, Sample Variance & Standard Deviation
8. Correlation & Causation
9. Hypothesis Testing
10. Statistics Implementation with Python -2

Basics of Statistics
& Types of Data

Math for Machine Learning

Basics of Statistics

1. Why we need Statistics?
2. Applications of Statistics
3. Types of Data

Why we need Statistics ?

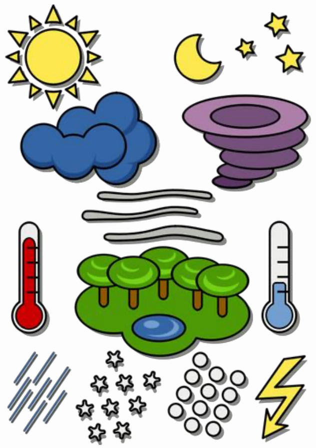
Statistics is a tool that helps us to extract information & Knowledge from data

Who is the best Batsman in the world in the time period 2010 to 2020?

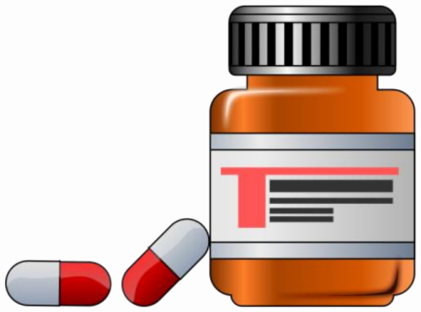
Few Applications of Statistics

6σ SIX SIGMA

Business



Weather Forecast



Clinical Trial of Medicines

Types of Data

Data

Categorical

Nominal

Nominal data *is a classification of categorical variables, that do not provide any quantitative value.*

Ordinal

Ordinal data *are the type of data in which the values follow a natural order.*

Numerical

Discrete

Discrete Data *are the type of data that can only take certain values.*

Continuous

Continuous data *can have almost any numeric value. Unlike Discrete data, they can have decimal values.*

Types of Statistics:

- Descriptive Statistics
- Inferential Statistics

Math for Machine Learning

Types of Statistical Studies

Math for Machine Learning

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

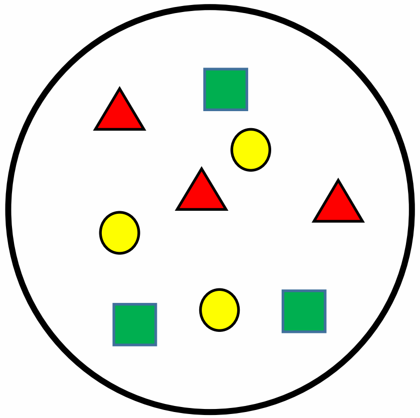
*An **observational study** is a study where we simply collect and analyze data. We won't inject any changes. We just observe the correlation in the data.*

Experimental Study

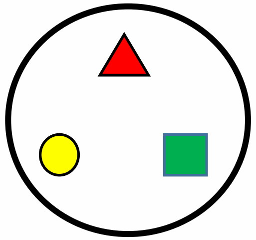
*An **experimental study** is a **study** in which conditions are controlled and manipulated by the experimenter.*

1. Sample Study

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



Population

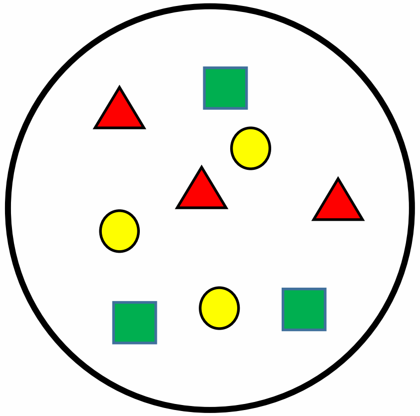


Sample

Average Blood Sugar Level = ?

2. Observational Study

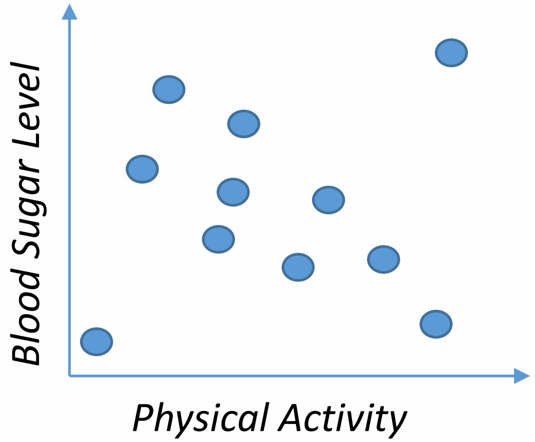
*An **observational study** is a study where we simply collect and analyze data. We won't inject any changes. We just observe the correlation in the data.*



Population

Relation between:

1. Blood Sugar Level
2. Physical Activity



Inference: Blood Sugar Level & Physical Activity are Negatively Correlated

3. Experimental Study

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

People following healthy food habits & doing regular exercises

Group 1

Group 2

People with unhealthy food habits & doing very minimal exercises

Control Groups

Conclusive Result

(People in Group 1 have optimum blood sugar level)

Types of Statistics

Statistics

Descriptive Statistics

- Describe the data
- Data Analysis

Inferential Statistics

- Get inferences about the data
- Data Science

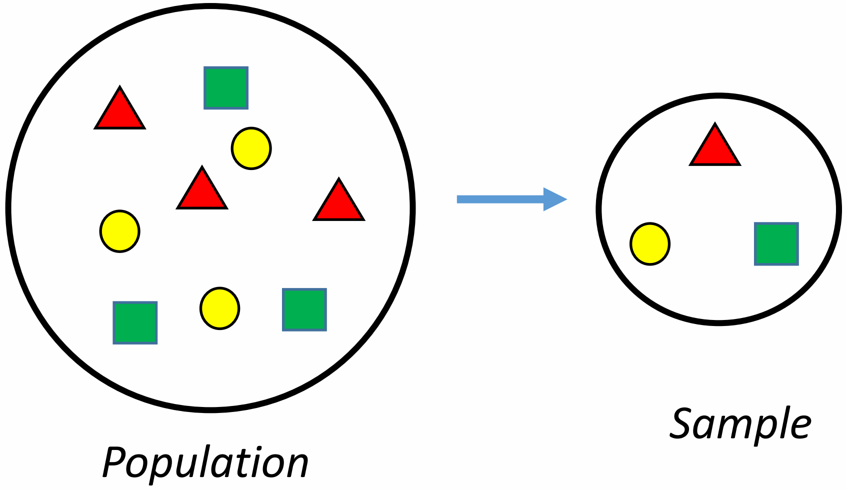
Types of Statistics

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)

Descriptive Statistics of House Price Dataset

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	price
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000
mean	3.613524	11.363636	11.136779	0.069170	0.554695	6.284634	68.574901	3.795043	9.549407	408.237154	18.455534	356.674032	12.653063	22.532806
std	8.601545	23.322453	6.860353	0.253994	0.115878	0.702617	28.148861	2.105710	8.707259	168.537116	2.164946	91.294864	7.141062	9.197104
min	0.006320	0.000000	0.460000	0.000000	0.385000	3.561000	2.900000	1.129600	1.000000	187.000000	12.600000	0.320000	1.730000	5.000000
25%	0.082045	0.000000	5.190000	0.000000	0.449000	5.885500	45.025000	2.100175	4.000000	279.000000	17.400000	375.377500	6.950000	17.025000
50%	0.256510	0.000000	9.690000	0.000000	0.538000	6.208500	77.500000	3.207450	5.000000	330.000000	19.050000	391.440000	11.360000	21.200000
75%	3.677083	12.500000	18.100000	0.000000	0.624000	6.623500	94.075000	5.188425	24.000000	666.000000	20.200000	396.225000	16.955000	25.000000
max	88.976200	100.000000	27.740000	1.000000	0.871000	8.780000	100.000000	12.126500	24.000000	711.000000	22.000000	396.900000	37.970000	50.000000

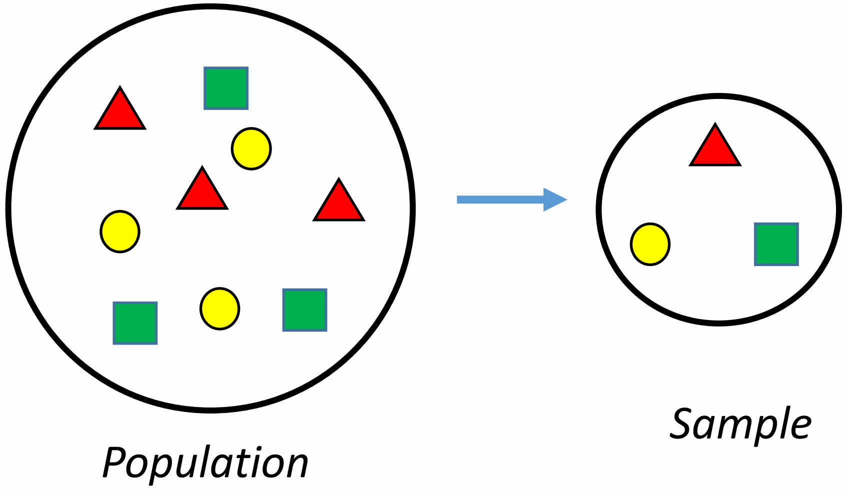
Boston House Price Dataset

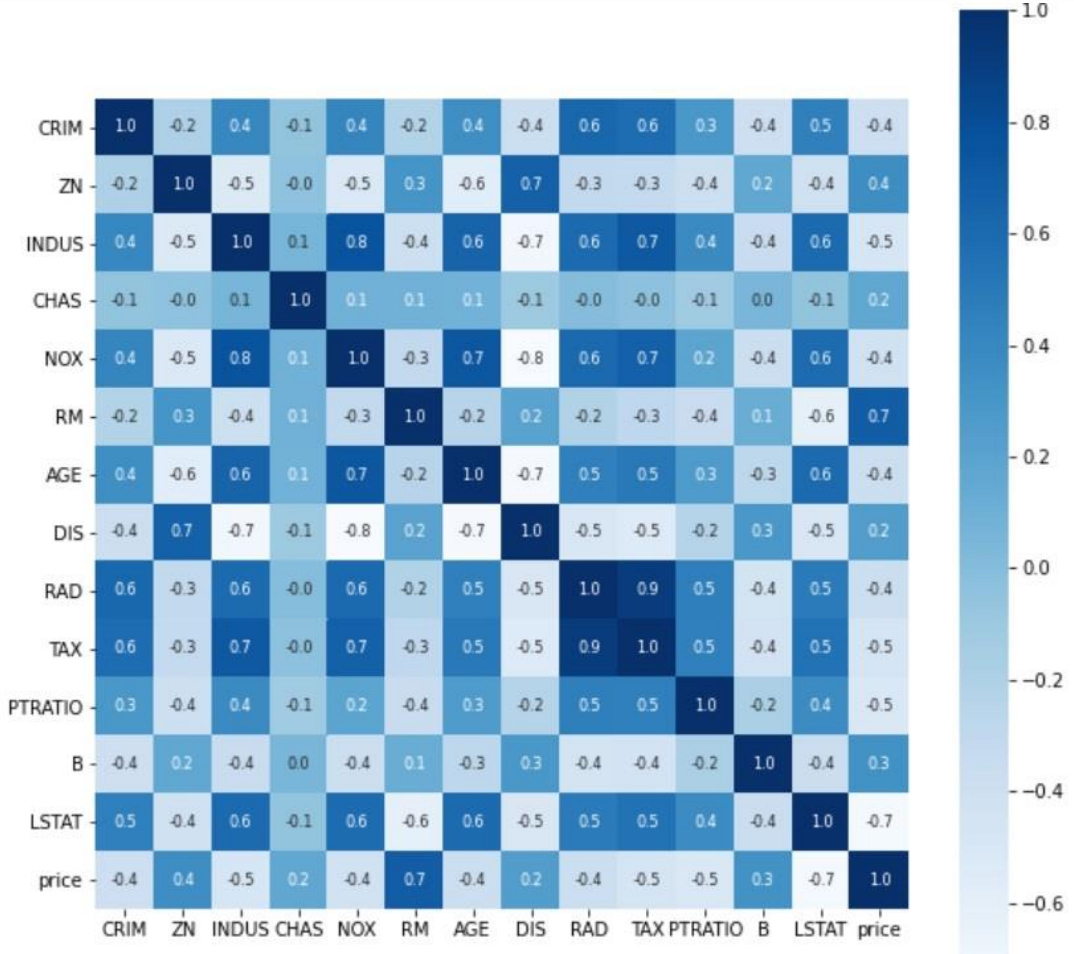
The dataset used in this project comes from the UCI Machine Learning Repository. This data was collected in 1978 and each of the 506 entries represents aggregate information about 14 features of homes from various suburbs located in Boston.

crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	b	lstat	price
0.00632	18	2.31	0	0.538	6.575	65.2	4.09	1	296	15.3	396.9	4.98	24
0.02731	0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.9	9.14	21.6
0.02729	0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
0.03237	0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4

Inferential Statistics

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

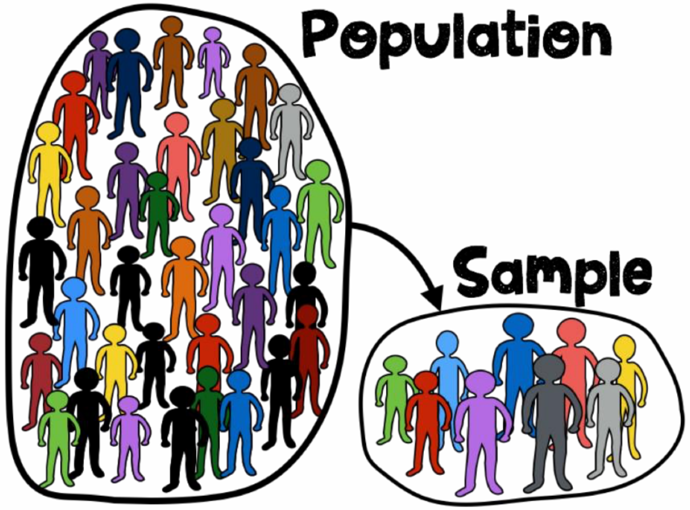




Correlation of House Price Data

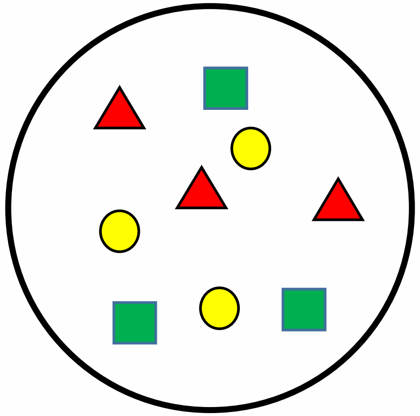
Population & Sample - Sampling Techniques

Math for Machine Learning

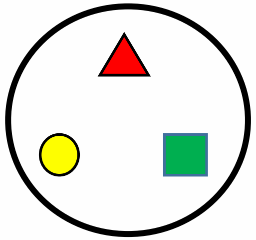


1. Sample Study

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



Population



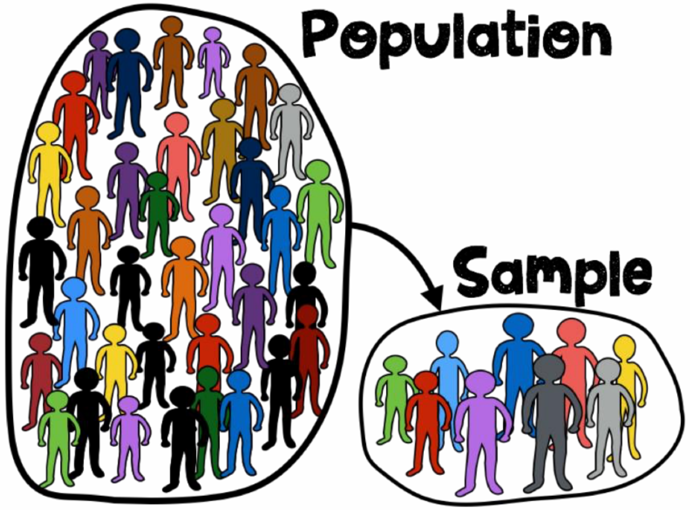
Sample

Average Blood Sugar Level = ?

Types of Sampling Techniques

Sampling Techniques:

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



(Probability Sampling Techniques)

(Non-Probability Sampling Techniques)

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

Measure of Central Tendencies: Mean, Median & Mode

Math for Machine Learning

Central Tendency

Central Tendency:

*A measure of **central tendency** is a value that represents the center point or typical value of a dataset. It is a value that summarizes the data.*

Central Tendency

- ***Mean***
- ***Median***
- ***Mode***

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

160 165 168 172 174 176

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

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

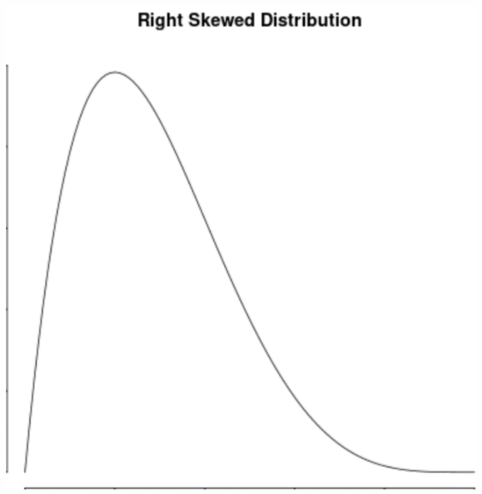
160
172
160
168
174

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: Range, Variance & Standard Deviation

Math for Machine Learning

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.*

$$Range = Max\ value\ - Min\ Value$$

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.*

$$\sigma^2 = \frac{\sum(\chi - \mu)^2}{N}$$

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.

$$SD = \sqrt[2]{\sigma}$$

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 & Quantiles

Math for Machine Learning

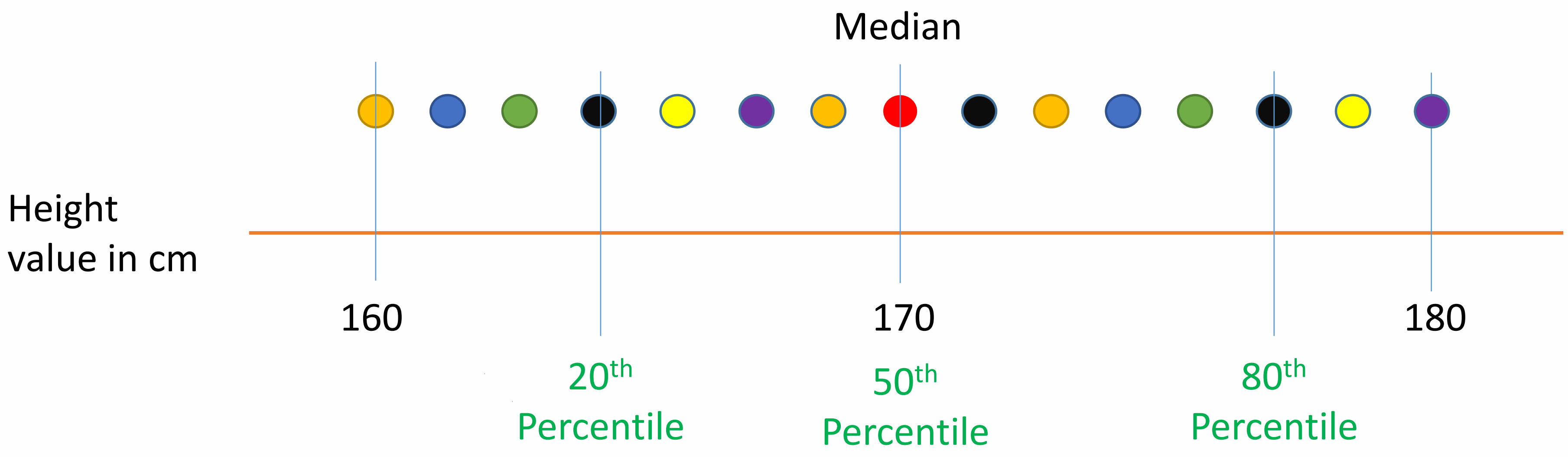
Purpose of these Measurements

Distribution of data points in a dataset

Percentiles

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

Dataset with Height of 15 people



Correlation & Causation

Math for Machine Learning

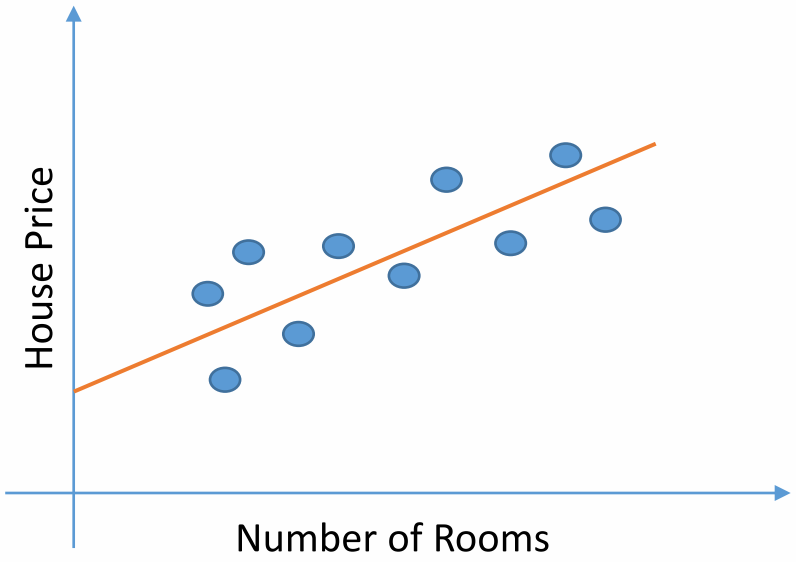
Correlation & Causation

Relationship between the features in a Dataset

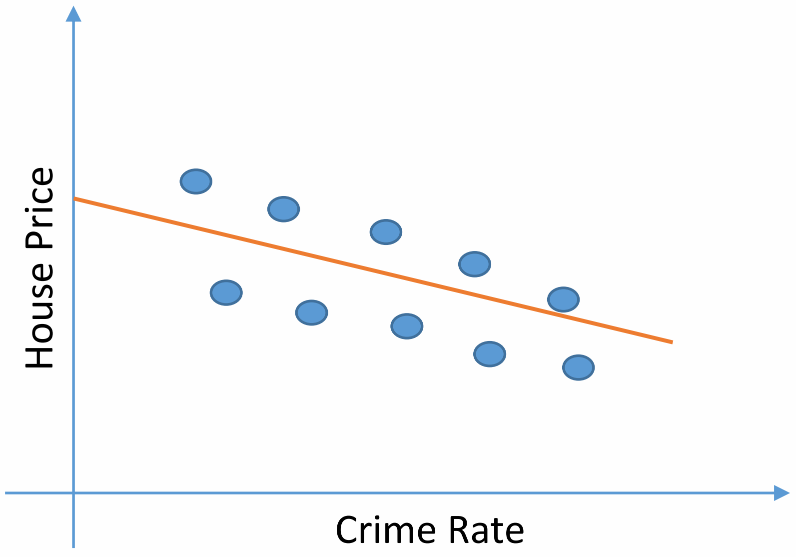
Correlation

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

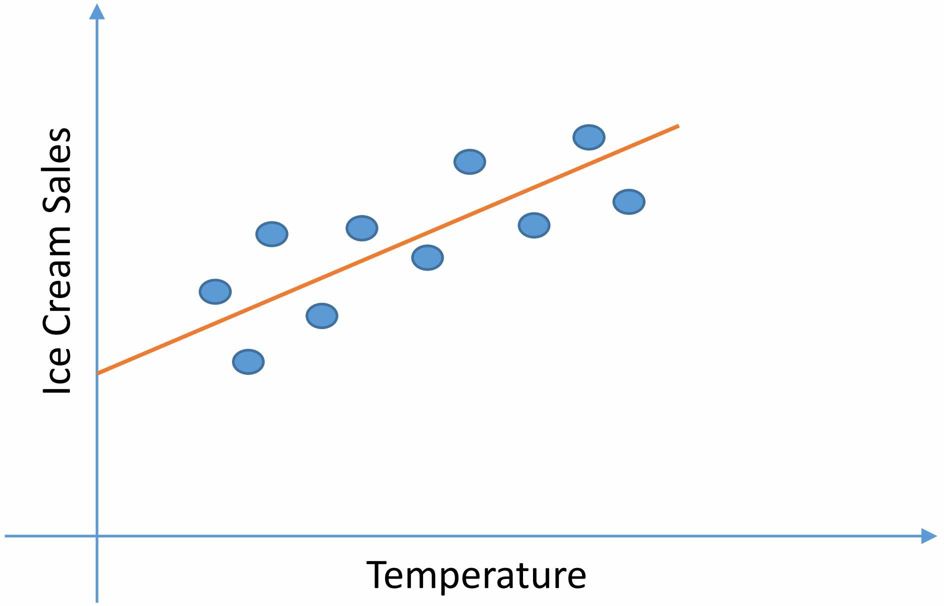


Negative Correlation



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.



Hypothesis Testing:
- Null Hypothesis & Alternative Hypothesis

Math for Machine Learning

Hypothesis

Hypothesis is an assumption that is made based on the observations of an experiment.

Hypothesis

Null Hypothesis

Null Hypothesis (H_0) is the commonly accepted fact.

Example: Ptolemy proposed that sun, stars and other planets revolve around the earth.

Alternative Hypothesis

Alternative Hypothesis (H_a) is opposite to null hypothesis and it challenges the null hypothesis.

Example: Aryabhata proposed that earth and other planets revolve around the sun.

Hypothesis Testing

Hypothesis is an assumption that is made based on the observations of an experiment. Hypothesis Testing is a method carried out to tests the assumptions made in the experiment.

Pharmaceutical Company

Drug A

Drug B

Headache

Hypothesis Testing

GROUP 1

Drug A

[12,   8,   13,   10,   7]

(Time taken for recovery
in minutes)

Average Time taken = 10 minutes

GROUP 2

Drug B

[15,   12,   18,   16,   14]

(Time taken for recovery
in minutes)

Average Time taken = 15 minutes

NULL HYPOTHESIS: Drug A takes 10 minutes on an average to cure headache; Drug B takes 15 minutes on an average to cure headache. Hence, Drug A is more quicker.

Hypothesis Testing

NULL HYPOTHESIS (H_0): Drug A is more quicker than Drug B.

ALTERNATIVE HYPOTHESIS (H_a): Drug B is more quicker than Drug A.

Possible Outcomes of Hypothesis Testing:

- Reject the Null Hypothesis
- Fail to reject the Null Hypothesis

Drug B

Drug A

Probability for Machine Learning

Math for Machine Learning

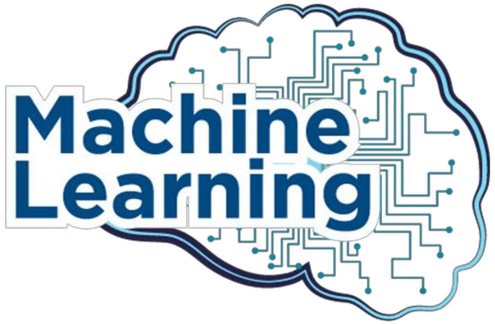
What is Probability ?

Probability is a branch of Mathematics that deals with calculating the likelihood of a given event to occur.

Simple Examples:

1. Roll a Dice
2. Toss a coin
3. Bag containing different coloured balls

Importance of Probability in ML



Machine Learning Model

$$\frac{1}{n}\sum(y_i - \hat{y}_i)^2$$

Loss Function

Topics covered in this module:

1. Basics of Probability
2. Random Variables
3. Probability Distributions
4. Maximum Likelihood
5. Bayes Theorem
6. Information Theory
7. Cross Entropy
8. Information Gain

Basics of Probability

Probability is a branch of Mathematics that deals with calculating the likelihood of a given event to occur.
The Probability value lies between 0 and 1.

Consider Rolling a Die

- What is the probability of getting a number greater than 10 when we roll a die?

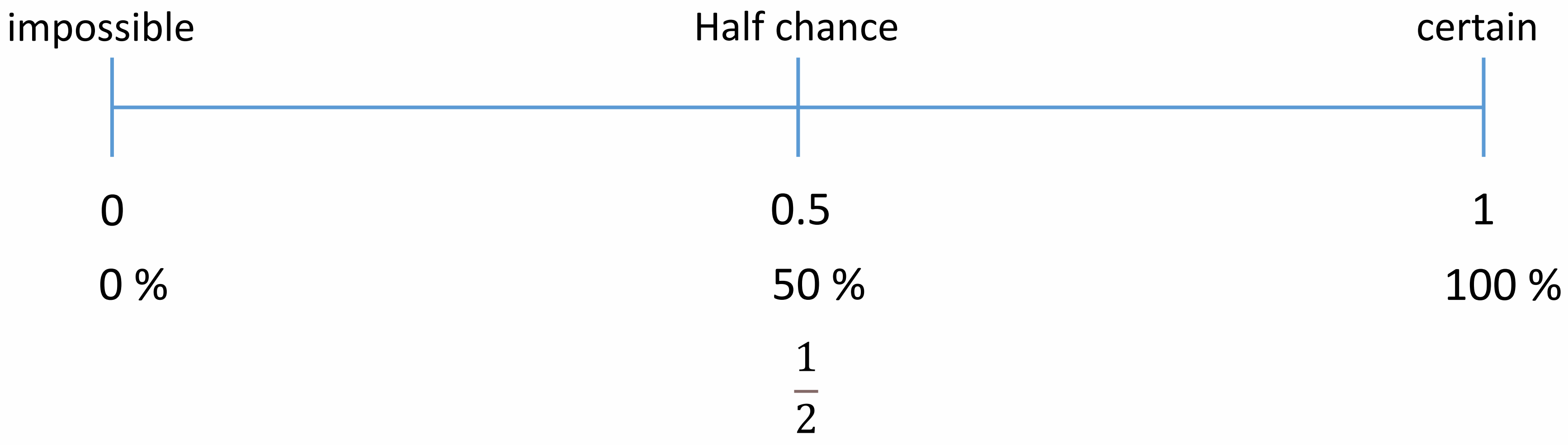
  Answer: 0

- What is the probability of getting a number less than 10 when we roll a die?

  Answer: 1

Probability Value

The Probability value lies between 0 and 1.



Basics of Probability

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

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;
Types of Random Variables

Math for Machine Learning

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 | *Possible Values* | *Random Events*

$$X = \begin{cases} 0 \leftarrow \text{Head} \\ 1 \leftarrow \text{Tail} \end{cases}$$

Head

Tail

Random Variables

Few Examples of Random Variables:

$$X = \begin{cases} 0, & \text{if Heads} \\ 1, & \text{if Tail} \end{cases}$$

Y = Weight of a random person in a class

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

$P (Y < 60)$

Applications:

- Turnover of a company in a given time period.
- Price change of an asset over a given time period

Types of Data

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.

Probability Distribution for Random Variable

Math for Machine Learning

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 — *Possible Values* — *Random Events*

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

Head

Tail

Probability Distributions

The **probability distribution** for a random variable describes how the probabilities are distributed over the values of the random variable.

Tossing 3 Coins

X = Sum of number of Heads when 3 coins are tossed

HHH = 3
TTT = 0
HHT = 2
HTH = 2

THH = 2
TTH = 1
HTT = 1
THT = 1

Probability Distributions

HHH = 3

TTT = 0

HHT = 2

HTH = 2

THH = 2

TTH = 1

HTT = 1

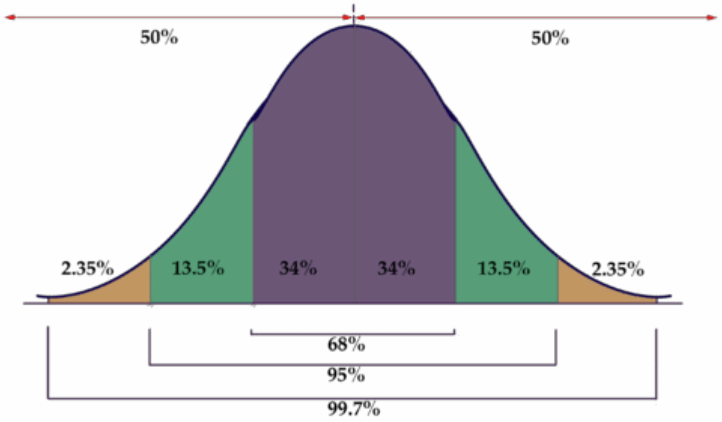
THT = 1

X (No. of Heads)	P(X = x)	P(X = x)
0	1/8	0.125
1	3/8	0.375
2	3/8	0.375
3	1/8	0.125

Discrete Probability Distributions

Normal Distribution & Skewness

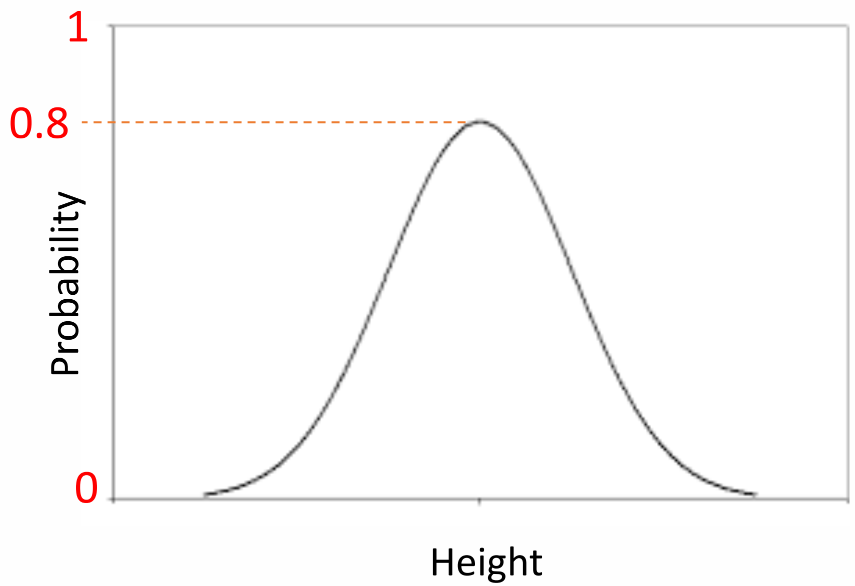
Math for Machine Learning



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.

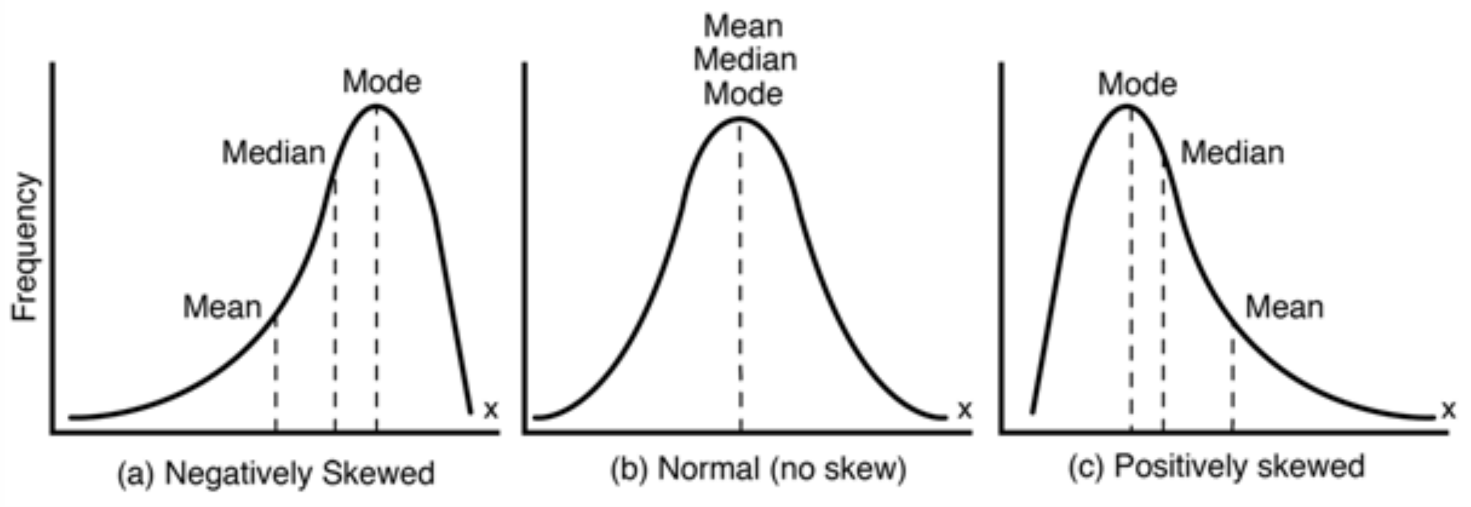
Normal Distribution is also known as **Gaussian Distribution**.



Bell Shaped Curve

Skewness

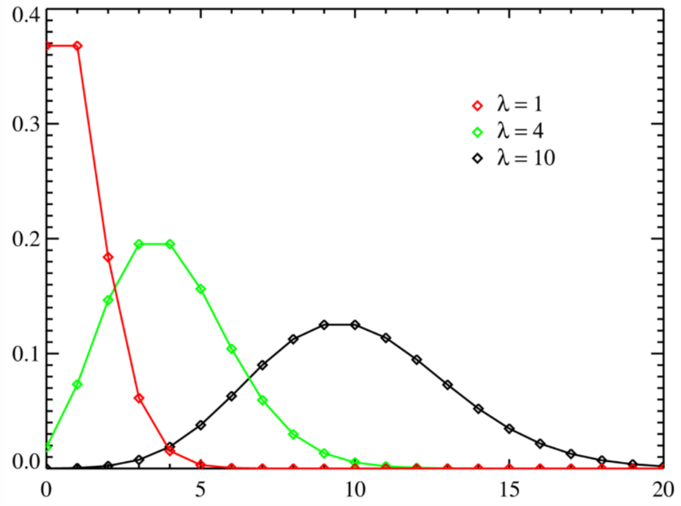
A data is considered **skewed** when the distribution curve appears distorted or skewed either to the left or to the right, in a statistical distribution.



Example: Average income of people in different cities

Poisson Distribution

Math for Machine Learning



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

Poisson Distribution

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

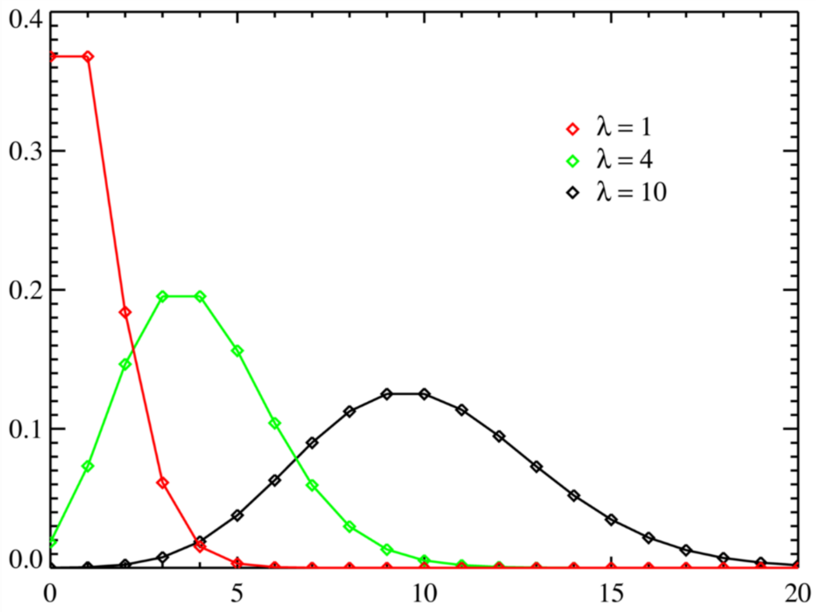
x --> Number of times the event occurs

p(x) --> Probability

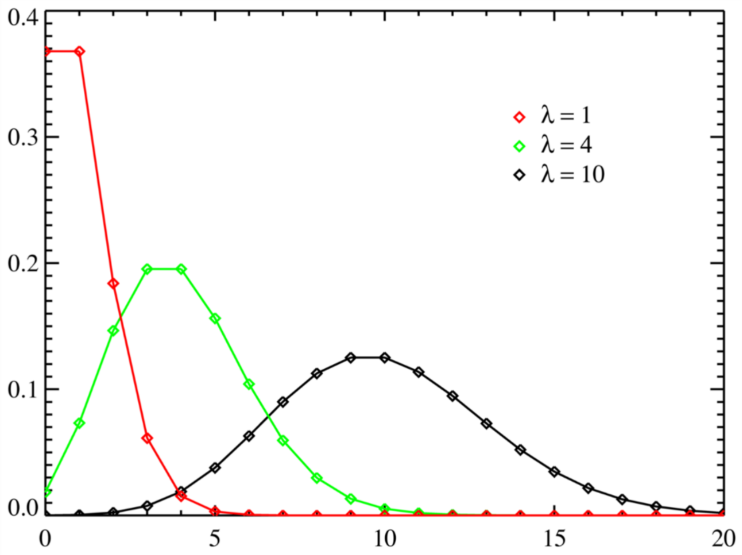
λ --> Mean number of events

x! --> Factorial of x

e --> Euler's Number (2.71828)



Poisson Distribution



Data
Science