

Components of math that contribute to Data Science:

- Linear Algebra
- Calculus

Linear Algebra:

Used in image recognition, text analysis and dimensionality reduction.

When classifying images in Machine Learning, the computer can recognize the images in the form of matrices. Matrices are an important component of linear Algebra.

Types of Matrices

- Vectors: single dimension matrices
- 2Dimensional Matrix : most common and takes the format of n-rows and n-columns
- Identity matrix: It is a given square matrix of any order which contains on its main diagonal elements with value of one, while the rest of the matrix elements are equal to zero.

One of the support libraries of Linear Algebra is NumPy

Calculus

Used in optimization techniques

One can carry out mathematical modelling of artificial neural networks and also increase their accuracy and performance.

Classification of calculus:

- Differential calculus

Studies the rate at which quantities change.

Derivates are mostly used for finding the minima and maxima of functions Derivatives are used in optimization techniques where we have to find the minima in order to maximize the error function.

- Integral Calculus

Study of the accumulation of quantities and used in finding the area under the curve.

Bayesian Inference is an important field in machine learning that makes use of integral calculus

Statistics for Data Science

Statistics deals with raw data and helps industries to take careful data-driven decisions.

1.Descriptive Statistics

Deals with quantitative summarization of data

Key concepts:

- Normal Distribution:

Represents large amount of Data in a plot.

- Central Tendency

Identifies the central point of data

- Skewness and Kurtosis:

Skewness is a measure of symmetry while Kurtosis is a measure of tailedness meaning it measures extreme values in either tails of the graph.

- Variability

Measures the distance of the data point from the central mean of the distribution. Measures of variability include range, variance, standard deviation and inter quartile ranges.

## 2. Inferential statistics

Process of inferring or concluding from the data.

Techniques include:

- Central Limit Theorem

In a central Limit Theorem, the mean of the smaller sample is the same as that of the mean of the larger population.

- Hypothesis Testing

Measure of assumption

- ANOVA (Analysis of Variance)

Test hypothesis for multiple groups, performs testing with minimal error rate

- Qualitative Data Analysis

Consists of correlation and regression

1. Correlation: measure of finding relationships between random variables and bi-variate data.
2. Regression: Estimation of the relationship between variables

## Exercises answers

### Exercise 1

Write a Python program to get the difference between a given number and 17, if the number is greater than 17 return double the absolute difference.

Python code:

```
def difference(n):  
    if n <= 17:  
        return 17 - n  
    else:  
        return (n - 17) * 2  
print (difference (22))  
print (difference (14))
```

## Exercise 2

write a python program to differentiate  $f(x) = 2x^2 + 3$

```
>>> from sympy import *  
>>> #create a "symbol" called x  
>>> x = Symbol('x')  
>>> f = 2*x**2+3  
>>> f_prime = f.diff(x)  
>>> f = lambdify(x, f)  
>>> f_prime = lambdify(x, f_prime)  
>>> # Let's test it out  
>>> f(3)21  
>>> f_prime(3)12
```