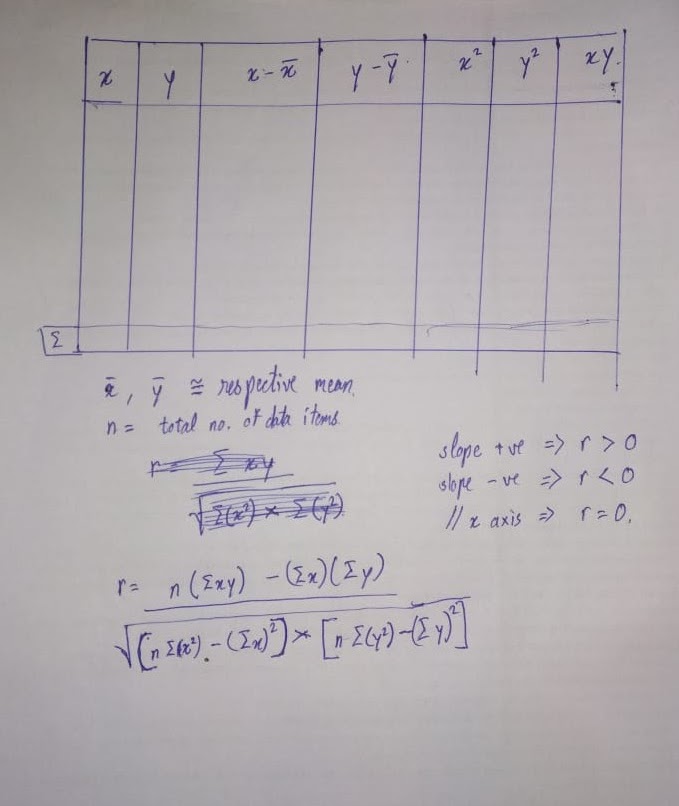
Linear regression model

Week 2

## Aniket Vaishnav

2017BTEIT00062

# Introduction

A Linear regression is a way of finding a linear relation between two variables. It may be extended to more than two variables in this it is known as multiple regression. It depends upon correlation coefficient.

Data Set chosen here is [Graduate Admission](https://www.kaggle.com/mohansacharya/graduate-admissions?select=Admission_Predict_Ver1.1.csv)

# Code

Written in python can be found at [GitHub](https://github.com/OBrutus/DataMining/blob/master/1.BasicInfo/main.py)

import math

import pandas

def selectattribute(csv\_file):

index = 0 # default value of index

for row in csv\_file.keys():

print(index, ': ', row)

index += 1

index = int(input('input : '))

return csv\_file.keys()[index]

if \_\_name\_\_ == '\_\_main\_\_':

target\_csv\_file\_name = 'dataset/Admission\_Predict.csv'

csv\_file = pandas.read\_csv(target\_csv\_file\_name)

print('select one of the Y attribute')

y\_attr = selectattribute(csv\_file)

print(' Y attribute set to : ', y\_attr)

print('select one of the X attribute')

x\_attr = selectattribute(csv\_file)

print(' X attribute set to : ', x\_attr)

# print('table creation')

y = list(csv\_file[y\_attr])

x = list(csv\_file[x\_attr])

y\_sum = csv\_file[y\_attr].sum()

x\_sum = csv\_file[x\_attr].sum()

y\_mean = csv\_file[y\_attr].mean()

x\_mean = csv\_file[x\_attr].mean()

y\_minus\_y\_mean = []

for ele in y:

y\_minus\_y\_mean.append(ele-y\_mean)

x\_minus\_x\_mean = []

for ele in x:

x\_minus\_x\_mean.append(ele-x\_mean)

y\_sq = []

for ele in y:

y\_sq.append(ele\*\*2)

x\_sq = []

for ele in x:

x\_sq.append(ele\*\*2)

xy = []

for p, q in zip(x, y):

xy.append(p\*q)

n = csv\_file[x\_attr].count()

r = ( n\*(sum(xy)) - sum(x)\*sum(y) ) / math.sqrt( (n\*sum(x\_sq)-sum(x)\*\*2) \* (n\*sum(y\_sq)-sum(y)\*\*2) )

print('r = ',r)

# Dependencies

Python 3+

Pandas

Run via python3 main.py

# References

<https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html>