Linear regression model

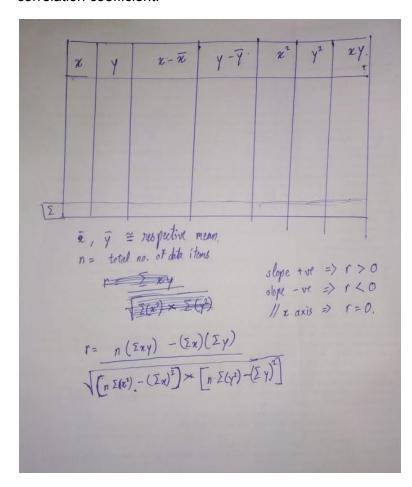
Week 2

Aniket Vaishnay

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 <u>Graduate</u> Admission

Code

Written in python can be found at GitHub

```
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