Topics in Deep Learning

Lab Assignment 01

By

Tilak Nanavati
201811001

Problem Statement :

You are given a Dataset. Split the Dataset into two parts, one part will be used for the training purpose of our linear regression model and the other part will be used for the testing purpose.

Solution:

Initially we are taking the normalization of the given training set and then build the model. For that I am using gradient descent method to find the thetas.

Code :

```python
import pandas as pd
import numpy as np
import statistics as s

data = pd.read_csv("Training_Set.csv")
colname = []
l = []
for i in data:
    colname.append(i)
#print(colname)
colname = colname[:-1]
for i in colname:
    l.append(data[i])
#print(l)

mn = []
#print(data.x1)
total = len(data.x1)

print("Mean Values :")
print("*************")
for i in l:
    #print(i)
    print(sum(i)/total)
    mn.append(sum(i)/total)

std = []
print()
print("Std Values :")
print("***********")

for i in l:
    #print(i)
    print(s.stdev(i))
    std.append(s.stdev(i))

count = 0
for i in l:
    for j in range(total):
        #print(i.iloc[j])
        i.iloc[j] = (i.iloc[j] - mn[count])/(std[count])
        #print(i.iloc[j])
    count+=1
```

```python
#data

alpha = 0.01

theta = [0,0,0,0,0,0]

for it in range(0,1000):
    cm = [0,0,0,0,0,0]
    for i in range(total):

        xi = [1]

        for j in range(0,5):
            xi.append(data.iloc[i,j])

        #print("XI :",xi)
        #print("Theta :",theta)
        c = 0
        diff = 0

        for j in range(len(xi)):
            c = c + xi[j]*theta[j]

        #print("c:",c)

        diff = (c - data.iloc[i,5])

        for j in range(len(xi)):
            cm[j] += diff*xi[j]

        #print("CM:",cm)

    for i in range(0,6):

        theta[i] = theta[i] - ((alpha*cm[i])/total)

#new_theta = theta - ((alpha*cm)/total)
print()
print("Trained Model : ")
print("****************")
print("Theta Values :")
print(theta)
```

Results :

```
Mean Values :
*************
15.19139784946237
5.526881720430108
201.3978494623656
109.34408602150538
2962.505376344086

Std Values :
************
3.2246892642338256
1.7544440333885434
110.37339193618556
45.354520081449515
805.9667921691789
```

```
Trained Model :
*****************
Theta Values :
[23.724782178952555, -0.6189386025654944, -2.488613838813864, -0.13462728894325218, -1.7504571477345958, -3.3334374718515862]
```

Above thetas I am getting for the trained model. And following are the predicted values of the testing set that is given to us.

Predicted Values
24.78532681
13.7176476
14.72221891
13.38412084
14.62316609
18.67611474

And now taking the model as polynomial linear regression model.

I am obtaining the following theta values :

```
Trained Model :
*****************
Theta Values :
[19.821415549782397, -0.8728688420050662, -2.151932468610275, -0.6442565982906642, -5.8202318273177385, -1.6555264775063554, 0.1
703770921360513, 1.0492752760667947, 1.2960670710359252, -0.9008087084620638, -0.7227097121710119, 1.9637936227051045, 0.622878
577448875, 0.18866384517813325, 0.3241171839897036, 1.0443930428533243, -0.754734617361033, 0.2812655146154923, 1.19572871223388
65, 0.1957889146063764, -0.784928105085089]
```

Following are the predicted values for the theta we have obtained above.

Predicted Values
24.77518652
14.09037412
14.18819034
11.27316283
10.81452723

24.32589116