School of Computer Science Engineering and Technology

Course- BTech

Type- Core

Course Code-

Course Name- Statistical Machine learning

Year- 2023-2024

Semester- odd

Date – 03-08- 2023

Batch- ALL

Question 1:

Normalize the given data (0-1) with min_max scaling

Gender	Age Range	Head Size(cm^3)	Brain Weight(grams)
1	1	4512	1530
1	1	3738	1297
1	1	4261	1335
1	1	3777	1282
1	1	4177	1590
1	1	3585	1300
1	1	3785	1400
1	1	3559	1255
1	2	3613	1355
2	2	3982	1375
2	2	3443	1340
2	2	3993	1380
2	2	3640	1355
2	2	4208	1522
2	2	3832	1208

Question 2:

https://www.kaggle.com/datasets/camnugent/california-housing-prices

Download the dataset from the above link.

a) Read the data with pandas and describe the data
b) Find data type and shape of each column
c) Find the null values (if yes fill the null values with ‘0’ or mean of that column)

Question 3:

https://www.kaggle.com/datasets/camnugent/california-housing-prices

a) Read the data with pandas and find features and target variables
b) Normalize the data with min-max scaling
c) Split the data into train and test.

Answers:

1.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

```
Import numpy as np

M1=Np.min(Head_size)

M2=Np.max(Head_size)

L=[]

For I in range(0,len(Head_size):

        X=(x-m1)/m2-m1

        L.append(X)

Print(L)
```

2.

```
Import pandas as pd

Import numpy as np
```

a) D=pd.read_csv('https://www.kaggle.com/datasets/camnugent/california-housing-prices ')
   Print(d.describe())
b) Print(d.type())
   Print(d.shape)
c) Print(d.isnull())

```
X=np.mean(Total_rooms)

        D1=d.fillna(X)
```

3:

a) X=d.drop(columns=['Price'])

Y=d[[“price"]]

b)

M1=Np.min(Head_size)

M2=Np.max(Head_size)

L=[]

For I in range(0,len(Head_size):

    X=(x-m1)/m2-m1

    L.append(X)

Print(L)

e)

X=D

print("Enter the splitting factor (i.e) ratio between train and test")

s_f = float(input())

n_train = math.floor(s_f * X.shape[0])

n_test = math.ceil((1-s_f) * X.shape[0])

X_train = X[:n_train]

y_train = y[:n_train]

X_test = X[n_train:]

y_test = y[n_train:]

print("Total Number of rows in train:",X_train.shape[0])

print("Total Number of rows in test:",X_test.shape[0])