

HW

Student Performance Dataset (33 attributes)

ASSIGNMENT. NO: 3

Aim: Dissimilarity Matrix Generation for binary, nominal and numerical attributes.

```
import pandas as pd
import numpy as np
import seaborn as sbs
import matplotlib.pyplot as plt
```

```
df=pd.read_csv("student-mat.csv",sep=',')
df.head()
```

```
dfs=df[['schoolsup','famsup','paid','activities','nursery','higher','internet','romantic']]
```

```
dfs.head()
dfs=dfs.replace('no', 0)
dfs=dfs.replace('yes', 1)
```

****BINARY ATTRIBUTES****

```
n=np.array(dfs[['schoolsup','famsup']])
n=n.reshape(-1,2)
n.shape
```

→ # -1 means, Numpy will calculate this number for you.

```
m=np.array(dfs[['romantic','internet']])
m=m.reshape(-1,2)
m.shape
```

```
from scipy.spatial import distance
dist_matrix = distance.cdist(n,m)
dist_matrix.shape
```

```
print(dist_matrix)
sbs.heatmap(dist_matrix)
plt.show()
```

****NUMERIC ATTRIBUTES****

```
numeric=df[['age','Medu','Fedu','traveltime','studytime','failures']]
numeric.head()
```

```
num1=np.array(numeric[['age','failures']])
num1.reshape(-1,2)
num1.shape
```



```
num2=np.array(numeric[['Fedu','Medu']])
num2.reshape(-1,2)
num2.shape
```

```
from scipy.spatial import distance
dist_matrix = distance.cdist(num1,num2)
```

print(dist_matrix)

```
dist_matrix.shape
```

```
sbs.heatmap(dist_matrix)
```

****NOMINAL ATTRIBUTES****

```
nominal=df[['Mjob','Fjob','reason','guardian']]
nominal=nominal.replace('at_home','home')
nominal=(nominal.astype('category'))
```

```
from sklearn.preprocessing import LabelEncoder
lb=LabelEncoder()
```

```
nominal['guardian']=lb.fit_transform(nominal['guardian'])
```

```
nominal['Mjob']=lb.fit_transform(nominal['Mjob'])
```

```
nominal['Fjob']=lb.fit_transform(nominal['Fjob'])
nominal['reason']=lb.fit_transform(nominal['reason'])
```

```
nominal.head()
```

fit label encoder and return encoded labels.

```
nom1=np.array(nominal)
nom1.reshape(-1,2)
```

```
nom2=np.array(nominal)
nom2.reshape(-1,2)
```

```
from scipy.spatial import distance
dist_matrix = distance.cdist(nom1,nom2)
```

```
sbs.heatmap(dist_matrix)
```

```
dist_matrix.shape
```


Spicy, spatial, distance, `cdist(xA, xB, metric = 'euclidean', *args, **kwargs)`.

→ compute distance between each pair of the two collections of inputs.

→ returns an 'ndarray'.

`cdist(x, y, 'euclidean')`

→ computes the distance between m points using euclidean distance (2-norm) as the distance metric between the points. The points are arranged as m, n -dimensional row vectors in the matrix X .

Heat map: plot rectangular data as a color-encoded matrix.

→ A heat map is a graphical representation of data where the individual values contained in a matrix are represented as colors.

→ It is a bit like looking at a data table.

→ It is really useful to display a general view of numerical data.

class sklearn.preprocessing.LabelEncoder :-

→ Encode target labels with value between 0 & $n_{classes}-1$.

→ This transformer should be used to encode target values.

→ LabelEncoder is a utility class to help normalize labels such that they contain only values between 0 and $n_{classes}-1$.