USAMA SARWAR
FA17-BCS-090-B

Question # 01

Data Normalization in Python

Python provides the preprocessing library, which contains the normalize function to normalize the data. Here is a sample code.

Code

```
from sklearn import preprocessing
import numpy as np
a = np.random((1,4))
a = a*20
print ('Data = ', a)
normalized = preprocessing.normalize(a)
print('Normalized Data = ', normalized
```

Output

```
Data = [[4.00366677 8.09852736 4.34348565
  15.6718225]]

Normalized Data = [[ 0.21521128 0.43532457
  0.23347775  0.84241603]]
```

Question # 02.

Given.

Transaction & List of items

Support Threshhold = 50%, Confidence 60%.

Steps.

Step 1:

Support threshold = 50%.

$\Rightarrow 0.5 * 6 = 3$

$\Rightarrow min\_sup = 3$

Count of Each Item

$I1 = 4$, $I2 = 5$, $I3 = 4$, $I4 = 4$

$I5 = 2$

Step 2:

Prune Step $I5$ Deleted because

$I5 = 2 < min\_sup = 3$

Step 3:

Join Step: From 2-itemset. From the table find out the occurrence of 2-itemset

Item	Count
I1, I2	4
I1, I3	3
I1, I4	2
I2, I3	4
I2, I4	3
I3, I4	2

Step 4

Prune Step Table shows that item set {I1, I4} {{I3, I4} does not meet min_sup, thus deleted

Now

I1, I2	4
I1, I3	3
I2, I3	4
I2, I4	3

{I1, I2, I3} subsets, {I1, I2}, {I1, I3}, {I2, I3 are occuring in table thus {I2, I1, I3} is frequent. where {I2, I1, I4} ~~or~~ is not frequent so deleted

Now Item

I1, I2, I3

I1, I2, I4

I1, I3, I4

I2, I3, I4

Only {I1, I2, I3} is frequent

Step 6.

Generate Association Rule

{I1, I2} => {I3}

Confidence = $\frac{3}{4} \times 100 = 75\%$

{I1, I3} => {I2}

$C = 3/3 \times 100 = 100\%$

{I2, I3} => {I1}

$C = 3/4 \times 100 = 75\%$

{I1} => {I2, I3} $C = 3/4 \times 100 = 75\%$

{I2} => {I1, I3} $C = 3/5 \times 100 = 60\%$

{I3} => {I1, I2} $C = 3/4 \times 100 = 75\%$

This shows that all the above association rules are strong if min confidence threshold is 60%.

Question # 04

Pseudocode of Delta Learning Algorithm

Initialization

Define w_i, $i = 0, 1, 2, \ldots N$ } set w_i to small random values, e.g in the range $[-1, 1]$

Set $x_0(k) = 1$, for all k = 1, 2, ... K

Set training rate tr to a value in $[0, 1]$

```
STOP_EPOCH = 100
ERROR = STOP_ERROR
epoch = 0

do {
        epoch++
        ERROR = 0
          for i=0 to N {
              Delta[i] = 0
              }
        for k=1 to k=K {
            o(k) = 0
              for i=0 to i=N {
                  o(K) += wi * xi(k)
              }
              ERROR += (o(k) - d(k))^2
              for i=0 to N {
                  wi = wi + tr * (d(k) - o(k)) * xi(k)
              }
              for i=0 to N {
                  Delta[i] += (d(k) - o(k)) * xi(k)
              }
        } // end of k loop
```

```
for i=0 to N {
    wi = wi + tr * Delta[i]
}
} // end of do loop
while (ERROR < STOP_ERROR)
or
while (epoch < STOP_EPOCH)
```

Question #05

Difference between types of regression

The difference between simple leanear regression and multiple linear regression is that, multiple linear regression has (>1) independent variables, whereas simple linear regression has only 1 independent variable

Question # 05.

Consider the following data set (Training Set)

Income	No. of Siblings (x_2)	High School Grade x_1	Scholarship Amount (y)
1 M	3	2.3	No
0.5 M	4	3	Yes
0.2 M	2	3.5	Yes
0.9 M	3	2.9	No

Testing Set

x_1	x_2	x_3	y
0.7 M	2	3	?

Using Euclidean Formula

let $K=2$

1st row $(x_1 - y_2)^2 + (x_1 - y_2)^2$

$(1-0.7)^2 + (2.3-3)^2 = 0.98$

Similarly

2nd row $(0.5-0.7)^2 + (3-3)^2 = 0.66$

3rd row $(0.2-0.7)^2 + (3.5-3)^2 = 0.2$

4th row $(0.9-0.7)^2 + (2.9-3)^2 = 2.33$

So,

x_1	x_2	E.D	Rank	Y	Included in 2 nearst neigh
1	2.3	0.98	3	No	No
0.5	3	0.66	2	Yes	Yes
0.2	3.5	0.2	1	Yes	Yes
0.9	2.9	2.33	4	No	No

The only 2 neighbours are included they both are awarded scholarship, so our test data will also result in a scholarship awarded = Yes

∴ Y = Yes | Awarded