COMP4433 Quiz 3

Submission deadline: 3:30pm 31 Oct 2023

Hi, Fu Lai Korris. When you submit this form, the owner will see your name and email address.

1. Referring to slide 8 of "Clustering I" lecture notes, a dataset with n records and p attributes will form a nxp data matrix. how many dissimilarity values (i.e. d(i,j)) are needed to compute for the dissimilarity matrix?

   - ( ) n * p
   - ( ) n * n / 2
   - (•) (n * n - n)/2
   - ( ) n * n - n
   - ( ) None of the above

Dissimilarity matrix has nxn elements. The self dissimilarity d(i,i) values do not need to compute and there are n values. Dissimilarity values are symmetric, i.e. d(i,j)=d(j,i). So, the answer is (n * n - n)/2

2. This question is about distance measure for binary variables/attributes (i.e. slides 8-9 of "Clustering I" lecture notes. Which of the followings is true? Pick the best answer.

   - ( ) For a dataset with p asymmetric binary attributes, the range of Jaccard coefficient d(i,j) is from 0 to 1.
   - ( ) For a dataset with p symmetric binary attributes, the range of simple matching coefficient d(i,j) is from 0 to 1.
   - ( ) For a dataset with p (asymmetric or symmetric) binary attributes, the sum of a, b, c, and d of the contingency table is equal to p, i.e., a+b+c+d=p.
   - (•) All of the above (i.e., the first, second and third statements above are correct).
   - ( ) Only the second and the third statements are correct.

Very obvious for the first two statements, i.e., (b+c)/(a+b+c+d) or (b+c)/(a+b+c) and the third statement a+b+c+d=p which is the definition.

3. For a dataset with p categorical attributes, how many binary attributes will be created when one-hot encoding is used? Pick the best answer.

   - (•) It depends on the number of distinct attributes in each categorical attribute.
   - ( ) It depends on whether the attribute values are floating point or integer type.
   - ( ) It is equal to p.
   - ( ) None of the above

The final answer should be $\sum_i n_i$ where n_i denotes the number of distinct attributes in categorical attribute i. So, only the first answer is valid.

Some statistics for your reference:

COMP4433 Quiz 3

121	27:18	Active
Responses	Average time to complete	Status

View results

Open in Excel

1. Referring to slide 8 of "Clustering I" lecture notes, a dataset with n records and p attributes will form a nxp data matrix. how many dissimilarity values (i.e. d(i,j)) are needed to compute for the dissimilarity matrix?

More Details Insights

n * p	1
n * n / 2	5
(n * n - n)/2	106
n * n - n	3
None of the above	6

2. This question is about distance measure for binary variables/attributes (i.e. slides 8-9 of "Clustering I" lecture notes. Which of the followings is true? Pick the best answer.

More Details Insights

For a dataset with p asymmetric...	2
For a dataset with p symmetric ...	2
For a dataset with p (asymmetri...	7
All of the above (i.e., the first, se...	96
Only the second and the third st...	14

3. For a dataset with p categorical attributes, how many binary attributes will be created when one-hot encoding is used? Pick the best answer.

More Details

It depends on the number of dis...	108
It depends on whether the attri...	4
It is equal to p.	6
None of the above	3