Programming Assignment: Implementing Clustering Validation Measures

✅ Passed and verified · 10/10 points

Deadline   The assignment was due on December 10, 11:59 PM PST
You can still pass this assignment before the course ends.

Instructions   My submission   Discussions

In this assignment, you will be implementing two clustering validation measures: Normalized Mutual Information (NMI) and Jaccard similarity.

You will be given one ground-truth clustering (partition) results and five clustering test cases. You need to evaluate the clustering test cases with regard to the ground-truth by NMI and Jaccard measures and submit your measures. You will be graded based on whether your measures are correct.

The ground-truth clustering (partition) results are stored in file "partitions.txt"; the five clustering result test cases are stored in file "clustering_1.txt", ..., "clustering_5.txt".

All files including partitions.txt, clustering_1.txt, ..., can be downloaded from the data.zip file attached below.

> data.zip

Each clustering result (both ground-truth and test cases) is represented by a file. Each line in a file consists of two integers, separated by a space. The first integer represents the id of a data item, and the second integer represents the id of the cluster which this item belongs to.

You need to submit a file titled "scores.txt" consisting of 5 lines. Each line contains two float numbers separated by a space. The first number of the i-th line represents the NMI measure you calculated for the i-th test case i (i.e. "clustering_i.txt") with regard to the ground-truth given in "partitions.txt", and the second number of the i-th line represents the Jaccard measure you calculated for the i-th test case.

As an example, a valid submission may look like:

```
1   0.1000000 0.2000000
2   0.3000000 0.4000000
3   0.5000000 0.6000000
4   0.7000000 0.8000000
5   0.9000000 1.0000000
```

You will be graded based on whether your file format is correct and on how many measures you submitted are correct.