

Clustering Project

You are asked to implement several clustering methods:

1. **lloyd.py** : Lloyd's k-means.
2. **kmeanspp.py** : k-means++.
3. **cmeans.py** : fuzzy c-means.

Notice that the fuzzy c-means algorithm should be implemented to produce hard clustering. This should be done by first computing soft clustering and then converting the soft clustering to hard clustering.

Input1: One data file. The data is a comma separated matrix of size $n \times m$. Here the data points are the rows, not the columns.

Input2: k , the number of desired clusters.

Input3: r , the number of random iterations.

Output: A comma separated file containing n integer values. Each value is in the range $0 \dots k-1$.

What you need to submit

- Source code and documentation for the python scripts.

You **must** be available to demonstrate your program to the TAs. Time slots and additional instructions will be announced later.

Deadline:

TBA.