

Coding Assignment 4

Due: 2016/01/19 23:59:59

January 8, 2016

- Given a dataset with 30475 instances with 2688 dimensions, your assignment is to implement a cluster method, and **cluster them in 10 groups**. We'll test your cluster method on the same machine, and grade your program based on its performances (including executing time, intercluster separation and intracluster cohesion) and your report.

- Clustering metric :

$$\frac{\text{intercluster separation}}{\text{intracluster cohesion}} \times \frac{1}{\text{executing time}} = \frac{\sum_{i,j} (m_i - m_j)^2}{\sum_i \frac{1}{|G_i|} \sum_{x \in G_i} (x - m_i)^2} \times \frac{1}{(\text{sec})}$$

- Specification

- Follow the OOP (Object-Oriented Programming) paradigm in Matlab.
- **Cluster data into 10 groups**.
- You are free to use any data preprocessing techniques.
- You can't use the library of clustering, ex. kmeans method in matlab.

- Output format :
$$\begin{bmatrix} \text{clustering index of 1st instance} \\ \text{clustering index of 2ed instance} \\ \vdots \\ \text{clustering index of 30475th instance} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_{30475} \end{bmatrix},$$
 where $C_i \in \{1, 2, \dots, 10\} \forall i = 1 \sim 30475$.

- Submission Requirements:

- A report (REPORT.*) describing:
 - * How you implement the clustering.
 - * How you accelerate your clustering method.
 - * Anything worth mentioning.
- The code:
 - * Cluster.m
 - * You have to follow the specs (as described above) strictly.
- Put your work under package model.cluster (i.e. under directory +model/+cluster).
- Follow the submission guide (see README.md for details).

- High Performance Rewards:

- You'll get extra credits if the performance of your classifier is top 20% of the class (this will contribute to 10% of your final grade).