

CSC 242: Homework 1

Assigned on Jan. 22
Hand in before 2pm Feb. 10 (Tuesday)

Requirement

Please hand in your homework *before* our class (that is, before 2pm) on Feb. 10 (Tuesday). Include the **class ID number** assigned to you for this class, **your name**, and **how many days** late it is (if handed in late) in the headline. The ID number is used to simplify the grading procedure. If you do not know your ID number assigned to you, you can check the link.¹ If you are not on the list, please contact our TA Sean Esterkin (sesterki@u.rochester.edu). (*), (**), or (***) indicates the difficulty of each question.

Policy

We apply the late policy explained in syllabus to all homework. Any grading questions must be raised with the TA in two weeks after the homework is returned.

The homework must be completed individually. However, you are encouraged to discuss the general algorithms and ideas with classmates in order to help you answer the questions. You are also allowed to share examples that are not on the homework in order to demonstrate how to solve problems. If you work with one or more other people on the general discussion of the assignment questions, please record their names over every question they participated.

However, the following behaviors will receive heavy penalties (lose all points and apply the honest policy explained in syllabus)

- explicitly tell somebody else the answers;
- explicitly copy answers or code fragments from anyone or anywhere;
- allow your answers to be copied;
- get code from Web.

¹https://docs.google.com/spreadsheets/d/1hgxcxqFxBxZYb55_HHi2pWQt0rEyEppcStSZBKni8MU/edit?usp=sharing

1 (*) Linear Algebra: 2 points

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Calculate the value of Ab .

2 (**) Calculus: 2 points

Calculate the limit of the following equation

$$\lim_{x \rightarrow 0} \frac{\sin(x^2)}{(\sin x)^2}.$$

Hints: Use L'Hospital rule.

3 (**) Probability: 2 points

Throw two dices (each has six sides with values from 1 to 6) and compute the probability that the difference between two dices is 1.

4 (*) Ask Google for help: 2 points

List as least four applications of clustering algorithm.

5 (**) K-means: 8 points (programming homework)

Implement the K-means algorithm. Set the number of clusters as 3. The input data is a set of 2-D points shown in Figure 1 (You can download the coordinate file in "A.txt"). Please note that you may try a few times to obtain a reasonable clustering result.

You need to report

- the distortion function value

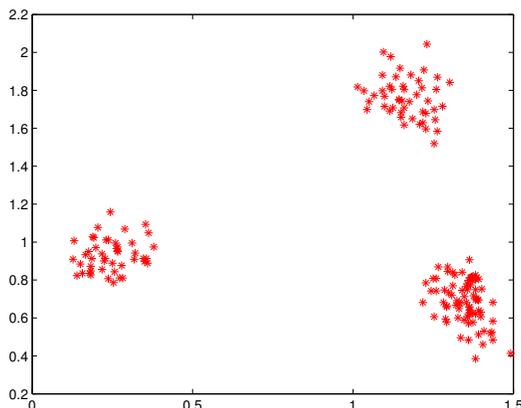
$$S = \sum_{i \in \text{group 1}} \|x_i - c_1\|^2 + \sum_{i \in \text{group 2}} \|x_i - c_2\|^2 + \sum_{i \in \text{group 3}} \|x_i - c_3\|^2$$

where c_1 , c_2 , and c_3 are the centers for three clusters you obtain;

- your clustering result shown in a figure (you should use different signs to mark 3 clusters; Photoshop or other similar softwares are forbidden in this problem).

You can use any programming language you feel comfortable. The code is not required to submit, but if TA had any question about your implementation or result, it is your obligation to provide runnable code which generates the result you submit. The fail to do that will cause losing points in this question. You should implement your own K-means algorithm. Directly calling existing functions for this algorithm from any programming languages is not permitted, but you are allowed to reuse any existing functions (even from your classmates) to print out your figure.

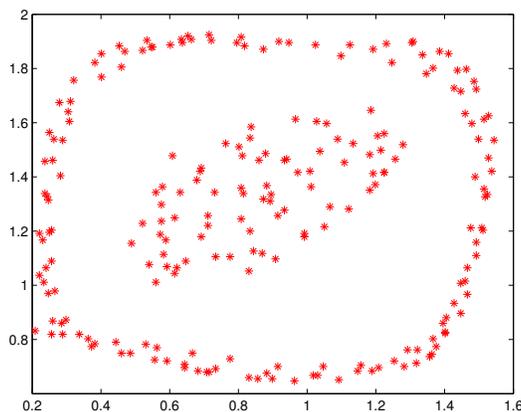
Figure 1: Input for K-means algorithm



6 (**) Hierarchical clustering: 8 points (programming homework)

Implement the hierarchical clustering algorithm. Choose the number of clusters as 2. The input data is a set of 2-D points shown in Figure 2 (You can download the coordinate file in “B.txt”). The reasonable clustering result is that the points in the center is a group while the points surrounding them is the other group. You should properly choose your cluster distance.

Figure 2: Input for Hierarchical clustering algorithm



You need to report

- your clustering result shown in a figure (you should use different signs to mark 2 clusters; Photoshop or other similar softwares are forbidden in this problem);
- your answer to the following question: can you get the same result using K-means and state your reason.

You can use any programming language you feel comfortable. The code is not required to submit, but if TA had any question about your implementation or result, it is your obligation to provide runnable code which generates the result you submit. The fail to do that will cause losing points in this question. You should implement your own Hierarchical clustering algorithm. Directly calling existing functions for this algorithm from any programming languages is not permitted, but you are allowed to reuse any existing functions (even from your classmates) to print out your figure.