COMS 4030A

Adaptive Computation and Machine Learning

Assignment 2

This assignment counts 10% towards your final mark.
The submission is to be done on moodle and will be graded by the autograder which also uses a plagiarism checker. You may **not** use any Python machine learning libraries.
Submissions will be accepted until **Monday 20 May** at **23h00**.

You are required to create a Python program that implements the k-MEANS ALGORITHM.
Your Python **submission** to moodle will be a .py file in which:

(1) The number of clusters is set to 3.

(2) The dataset below on page 3 is hard-coded into the algorithm.

Your algorithm must then:

(3) Read in from standard input a list of 6 numbers, such as
0.45
0.55
0.70
0.71
0.11
0.67

The first 2 values are the initial values for cluster centre 1, so $\boldsymbol{\mu}^1 = (0.45, 0.55)$.
The next 2 values are the initial values for cluster centre 2, so $\boldsymbol{\mu}^2 = (0.70, 0.71)$.
The last 2 values are the initial values for cluster centre 3, so $\boldsymbol{\mu}^3 = (0.11, 0.67)$.

(4) Run k-MEANS ALGORITHM using the hard-coded dataset and starting with cluster centres from step 3.

(5) Halt the algorithm when the centres have converged – that is, there are no changes to the cluster centres from one iteration to the next.

(6) Compute the sum-of-squares error on the dataset with respect to the final cluster centres, using the formula

$$\text{sum-of-squares error} = \sum_{j=1}^{3} \sum_{\boldsymbol{x} \in \text{cluster } j} d(\boldsymbol{x}, \boldsymbol{\mu}^j)^2.$$

(7) **The following value must be output using standard output:**
the sum-of-squares error with respect to the final cluster centres.
Round off to **4 decimal places**.

Samples:

(i) For the input values given above, the output is:

1.1053

(ii) For the following input values, the output is given below:

0.85
0.14
0.32
0.76
0.21
0.36

Output:

0.4379

Use the following dataset consisting of datapoints in $\mathbb{R}^2$:

0.22, 0.33
0.45, 0.76
0.73, 0.39
0.25, 0.35
0.51, 0.69
0.69, 0.42
0.41, 0.49
0.15, 0.29
0.81, 0.32
0.50, 0.88
0.23, 0.31
0.77, 0.30
0.56, 0.75
0.11, 0.38
0.81, 0.33
0.59, 0.77
0.10, 0.89
0.55, 0.09
0.75, 0.35
0.44, 0.55