2. K-means

Project due Apr 26, 2023 08:59 -03 Completed

K-means

1.0/1.0 point (graded)

For this part of the project you will compare clustering obtained via K-means to the (so
EM. In order to do so, our K-means algorithm will differ a bit from the one you learned. I
estimated exactly as before but the algorithm returns additional information. More spec
resulting clusters of points to estimate a Gaussian model for each cluster. Thus, our K-m
returns a mixture model where the means of the component Gaussians are the K cent
K-means algorithm. This is to make it such that we can now directly plot and compare s
two algorithms as if they were both estimating mixtures.

Read a 2D toy dataset using `X = np.loadtxt('toy_data.txt')`. Your task is to run th
this data using the implementation we have provided in `kmeans.py` Initialize K-means
`common.init(X, K, seed)`, where K is the number of clusters and `seed` is the rand
randomly initialize the parameters.

Note that `init(X,K)` returns a K-component mixture model with means, variances an
The K-means algorithm will only care about the means, however, and returns a mixture
on the K-means solution.

Try $K = [1, 2, 3, 4]$ on this data, plotting each solution using our `common.plot` functi
initialization is random, please use seeds $0, 1, 2, 3, 4$ to and select the one that minimiz
the associated plots (best solution for each K). The code for this task can be written in

Report the lowest cost for each K:

$\text{Cost}|_{K=1} =$ 5462.297452340002 ✔

$\text{Cost}|_{K=2} =$ 1684.9079502962372 ✔

$\text{Cost}|_{K=3} =$ 1329.59486715443 ✔

$\text{Cost}|_{K=4} =$ 1035.499826539466 ✔

Submit You have used 2 of 50 attempts

Previous Next