

HOMEWORK 6

Problem 1.

Visualize the observations in `faithful.csv` using a scatter plot. It contains two features: one of the features corresponds to the waiting time between eruptions and the second feature corresponds to the duration of each eruption at the Old Faithful geyser in Yellowstone National Park in Wyoming. It contains 272 observations on 2 variables.

Implement the k -means clustering algorithm from scratch and use it to cluster the data. Check for various values of k and visually identify the appropriate cluster size.

Problem 2. Go over the Jupyter notebook (`clustering.ipynb`) provided to you. It contains explanations and applications of some clustering methods. It also contains an application of the k -means clustering algorithm to an image compression problem (Section 2, Another Application of K -means clustering). In the Jupyter notebook provided to you, a black-box (mini-batch) k -means algorithm is used for compressing image. Using your k -means from scratch implementation from Problem 1, re-perform the same clustering approach to obtain a compressed image. If your implementation fails for this problem, describe why.

Bonus.

In the Jupyter notebook (`clustering.ipynb`) provided to you, go over Section 3, Mixture Models. It provides an application of the Gaussian mixture model to a dataset `clustering_gmm.csv` using a black-box implementation. Implement an implementation of the Gaussian mixture model from scratch and apply it on the dataset.