EE5601: Representation Learning, Fall 2018 (34)
Indian Institute of Technology Hyderabad

HW 0, Assigned: Monday 10.09.2018.
Due: Sunday 16.09.2018 at 11:59 pm.

Note: The programming exercises must be solved in Python. *Please do not use built-in functions for K-means, PCA and MLE.*

1. Implement the $K-$means algorithm. Your program must accept as input an $d \times N$ data matrix $\mathbf{X}$, the number of clusters K, and the stopping threshold ϵ. Your program must output the following:

    (a) the error at the end of each iteration,

    (b) the K centroids at convergence,

    (c) the K clusters at convergence.

    You are encouraged to draw inputs $\mathbf{X}$ from realistic scenarios. For example, you could use a color image and cluster the pixels based on their color. For such an example, you can display the clusters by assigning unique colors to each of them. (10)

2. Implement principal component analysis (PCA) to decorrelate data. How will you check if your algorithm is working as intended? As with the previous exercise, draw inputs $\mathbf{X}$ from realistic scenarios. For example, try to remove redundancies in local image patches using PCA. Find examples where PCA fails. (10)

3. Find the maximum-likelihood estimates (MLE) for the following distributions:

    (a) Binomial,

    (b) Poisson,

    (c) Exponential,

    (d) Gaussian,

    (e) Laplacian.

    Now, write a program to perform MLE for the same. Generate samples from your estimated parameters and compare with your data $\mathbf{X}$. (20)