

LA Assignment 2 Report

Achint Chaudhary

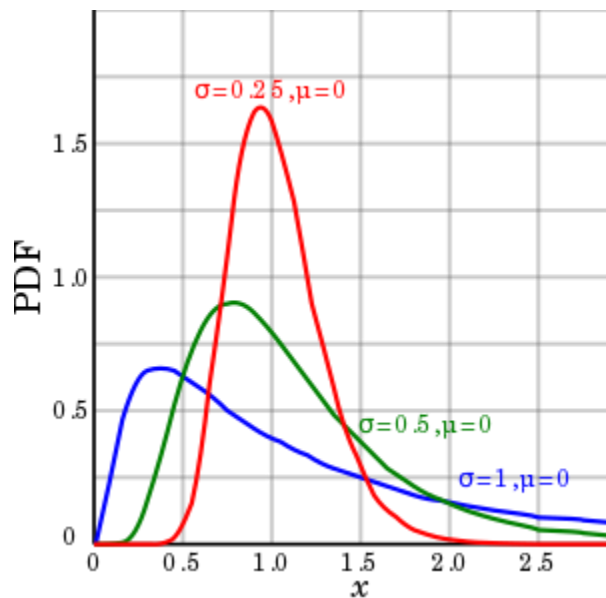
15879, M.Tech, CSA

Problem 1

Task 1

In Community graph, like one given in example, we have high number of other nodes which are connected to less nodes. While, some nodes have very high degrees & that form group within the main graph

The Distribution of Degree is Similar to Discrete case of Log-normal distribution, which is skewed on either side unlike symmetric Gaussian distribution. Below is an example of how continuous log-normal distribution looks like



Task 5

Difference between obtained & ideal smallest eigen value & it's corresponding eigen vector is written on to the output file. Almost, no difference from Ideal results were observed

Task 6

Eigen vector corresponding to second smallest eigen value tries to divide the graph into two parts, each of which have high information within it & less mutual information between these two divided clusters

Bonus 2

The technique used to divide the nodes into two group can be extended to divide them into any number, say K cluster. What we'll follow is a Hierarchical division algorithm an idea of which is presented below.

Like what we have observed that Eigen vector corresponding to second smallest Eigen value divides the nodes into two cluster. If we want more clusters, we can pick the Eigen vector corresponding to next smallest Eigen value, & chose a cluster out of present clusters that have maximum variance, this cluster can be divided into two parts by similar technique, this Eigen vector of 3rd smallest Eigen value can divided the initial data into at max of 4 clusters at this step.

If desired number of Clusters are not obtained, we can chose Eigen vector corresponding to next smallest Eigen value & continue the process, until we have K clusters at our disposal.

Bonus 3

If Eigen vector corresponding to second largest eigen value is taken, we'll not be able to divide clusters into two group as before with eigen vector corresponding to second smallest eigen value.

The classification made by eigen vector corresponding to second largest eigen value tries to separate out two nodes connected by an edge into different classes

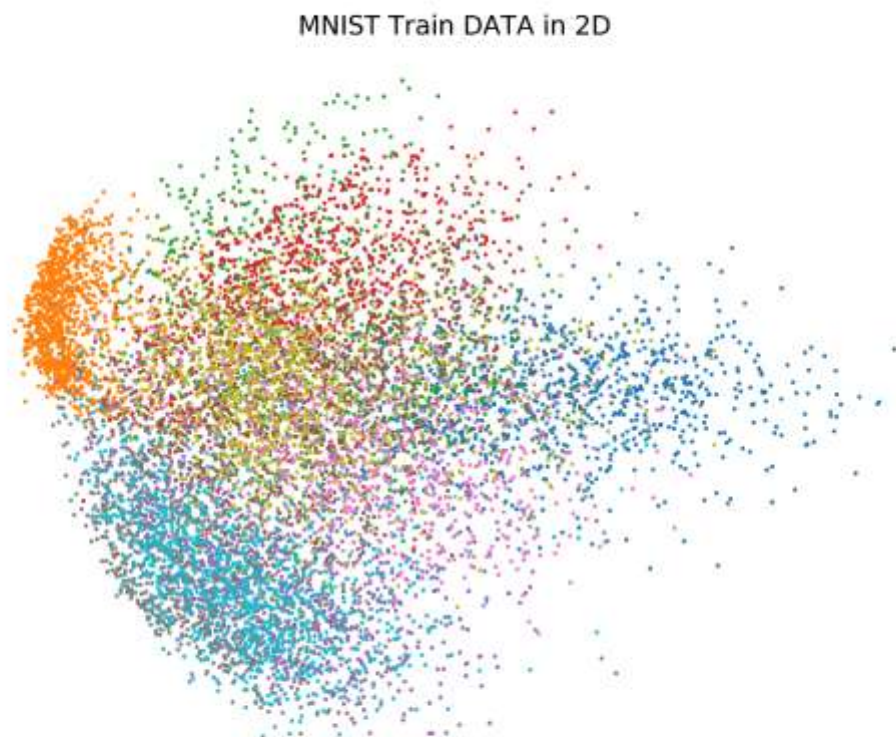
Bonus 4

Most central edge in the graph is the one that has two connecting vertices that will belong to different clustering, that were obtained from clustering process

Problem 2

Bonus 1

Image obtained after projection of given Dataset into 2D space is provided into output_plots folder. What we can infer that, in 2D dimension, these 10 classes of hand written digits do overlap each other, so it is hard for any ML algorithm to predict output labels with high accuracy if given as 2D data.



But also there are observable pattern which shows, that in 2D also all classes don't cover 2D space with uniformity. There are groups of data of same class/digit (mentioned with different colors) in figure in output_plots directory