REPORT

CSE 512: Assignment 3

Chumki Acharya (112683478)

Performance:

Algorithm	Dataset	K-value	Loss test	Loss train
GMM	Optdigits	1	-0.52	-0.32
GMM	Optdigits	3	-0.50	-0.31
GMM	Optdigits	5	-0.43	-0.28

GMM:

The Gaussian Mixture model has been implemented with the initial parameters being initialized by k-means and then using expectation maximization with the implementation of loss function. It has been observed that the loss function minimizes the loss with the increasing components. The dataset given is multidimensional, thus GMM is implemented for multidimensional data. The probability of each digit is calculated, and the total probability is given by average of all the digits. It has been observed that the loss is generally in case of test data than train data.

ReadMe:

GMM Algorithm

- File Name: gmm_learner.py
- Datasets: Optdigits
- components: number of components
- train: file path to train data folder
- test: file path to test data folder
- Run following on command line mentioning the appropriate dataset file and mode:

  python gmm_learner.py --components 3
  --train /path/to/training/data/
  --test /path/to/ test/data/

References:

- https://towardsdatascience.com/gaussian-mixture-modelling-gmm-833c88587c7f
- https://www.kaggle.com/dfoly1/gaussian-mixture-model/data?select=emails.csv
- https://www.youtube.com/watch?v=JNlEIEwe-Cg
- https://www.youtube.com/watch?v=0NMC2NfJGqo