HW8

CISC648010 - Spring 2021

Due Date: April 25th at 11 PM

1 EM algorithm 20 pts

Consider the exponential mixture model defined as follows,

$$p(x; \theta) = \sum_{k=1}^{K} \phi_k \cdot \lambda_k \exp\{-\lambda_k x\}, x \text{ is non-negative real number,} \quad (1)$$

where $\theta = [\phi_1, \ldots, \phi_K, \lambda_1, \ldots, \lambda_K]$ and $\sum_{k=1}^{K} \phi_k = 1$.

Assume that $K = 2$, and the observational data is $D = [x_1 = 1]$ (we have only one data point). We initialize $\lambda_1 = 1, \lambda_2 = 2$ and $\phi_1 = 0.5$ and $\phi_2 = 0.5$ before running the EM algorithm. Find the objective function of the optimization problem that is solved during the M step in the first iteration (10 pts). Solve the optimization problem and find the values for $[\phi_1, \phi_2, \lambda_1, \lambda_2]$ after the first iteration (10 pts).

2 EM algorithm (Programming) 20 pts

This problem is a programming assignment. Download dataset EM.csv from canvas. Run the following lines of code to load the dataset:

```
import csv
import numpy as np
import matplotlib.pyplot as plt
with open('EM.csv') as csv_file:
    csv_reader = csv.reader(csv_file, delimiter=',', quoting=csv.QUOTE_NONNUMERIC)
    data1 = [ ]
    for row in csv_reader:
        data1.append(row)
data1 = np.array(data1)
```

The dataset includes 1000 feature vectors (the dimension of each feature vector is 2) generated by a Gaussian Mixture Model with $K = 3$. Use the EM algorithm with 50 iterations to estimate the parameters of the Gaussian Mixture Model. You should report $\{\phi_j, \mu_j, \Sigma_j\}_{j=1,2,3}$. Upload your code on canvas to receive the full credit.