HW4, CSE 569 Fall 2019

Due 11:59pm Dec 1, 2019

Maximum score: 70

Q.1) (**70 pts**) Clustering and Mixture Models

The file `Binomial_20_flips.txt` contains the outcomes of coin tosses with 3 different coins. A coin is selected based on an unknown prior probability from a set of 3 coins. The selected coin is tossed 20 times and the number of heads is noted. This procedure is repeated 1000 times to give rise to the 1000 entries in the `Binomial_20_flips.txt` file. The goal is to determine the parameters (probability of heads) for the 3 coins $[\theta_1, \theta_2, \theta_3]$ and the prior probabilities of picking a coin $[\pi_1, \pi_2, \pi_3]$.

1) (20 pts) Apply K-Means algorithm with 3 clusters to determine the parameters $[\theta_1, \theta_2, \theta_3]$ and the prior probabilities $[\pi_1, \pi_2, \pi_3]$.
2) (20 pts) The entries in the `Binomial_20_flips.txt` file are the outcomes of a binomial distribution,

$$\text{Bin}[\theta, n, k] = \binom{n}{k} \theta^k (1-\theta)^{n-k}, \tag{1}$$

where, $n = 20$ and k is the number of heads when the coin with parameter θ is tossed n times. Assume the entries in the `Binomial_20_flips.txt` are the outcomes of a mixture distribution,

$$p(k) = \sum_{i=1}^{3} \pi_i \text{Bin}[\theta_i, n = 20, k]. \tag{2}$$

Apply the EM algorithm to determine the parameters $[\theta_1, \theta_2, \theta_3]$ and the prior probabilities $[\pi_1, \pi_2, \pi_3]$. Plot the curves for the negative log-likelihood (or the likelihood) vs. the iterations.

3) (20 pts) Repeat experiment 2 (EM algorithm) but this time initialize the parameters $[\theta_1, \theta_2, \theta_3]$ and the prior probabilities $[\pi_1, \pi_2, \pi_3]$ with the outcomes of the K-Means algorithm.
4) (10 pts) In about 200 - 300 words discuss the results of your experiments.

Your submission must be a .pdf file with your results followed by the code. Typeset your code and **do not** paste an image of your code in the report.