CS570: Artificial Intelligence and Machine Learning

Programming Assignment 3

Due Date: Thursday 23:59, May 22

T.A: Bowon Nam (bwnam@ai.kaist.ac.kr)

Summary

In this assignment you will implement Expectation Maximization algorithm for fitting Gaussian Mixture Models for the re-scaled Old Faithful data set. Auxiliary function of EM for GMM is defined as follows:

$$Q(\theta, \theta^{(t-1)}) = \sum_{\mathrm{i}} \sum_{k} r_{ik} \log \pi_k + \sum_{i} \sum_{k} r_{ik} \log p(x_i|\theta_k)$$

where $\mathrm{r_{ik}} = p(z_i = k|x_i, \theta^{(t-1)})$ is responsibility that cluster k takes for data point.

The assignment involves developing one function in MATLAB

① Implement a MATLAB function named ***GMM_EM.m*** for doing EM for fitting Gaussian Mixture Model

② Plot a graph similar to **Figure 11.11 in textbook**

A. Consider 2 clusters, start with $\mu_1 = (-1, 1), \Sigma_1 = I, \mu_2 = (1, -1), \Sigma_2 = I$

B. Set the color : $\mathrm{color(i)} = \mathrm{r_{i1}} color1 + r_{i2} color2$

③ Consider the following options of covariance matrix and plot graphs of each options (see **Figure 2.13 in textbook**) :

A. Independent(not shared) isotropic(spherical) covariance matrix

B. Independent(not shared) diagonal covariance matrix

Template for these functions is given. **You should upload your MATLAB code for GMM_EM function along with summary document of your experiments (studentid_name.pdf) within zip file to KLMS.**

Ex) 20142014_BowonNam.zip

(GMM_EM.m and 201421014_BowonNam.pdf are in this zip file)

Delayed submission won't be allowed. Please submit before the due time.

(Sending to T.A. won't be also allowed.)

Old faithful Datasets

For your assignment you will be working with re-scaled Old faithful data set. The data is in the file *oldfaithful.mat*. It has two dimension dataset for waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA. A dataset has 272 instances 2 variables(columns, first column is eruptions, second column is waiting).

Your document must include followings :

① Details about your GMM_EM algorithm

② Images for graph similar to **Figure 11.11 in textbook**

③ Graph for using isotropic covariance matrix

④ Graph for using diagonal covariance matrix

- *You can change the function arguments if you want*