Probabilistic Graphical Models

Home Work 3

Faculty of New Science & Technologies
University Of Tehran
Fall 2018-2019

Part 1

General Homework Policies

1. Due date of this homework is on *Tuesday 21 Day 97 (11 Jan. 2019)*, so you need to submit it before the due date[midnight 21 Day] otherwise you won't get any score!

2. Try to budget your time because due dates are hardly changeable, and we will not accept late homework for any reason.

3. You are welcome to collaborate, cooperate, and consult with your classmates provided that you write-up the solutions independently.

4. Don't plagiarize! Write everything in your own words, and properly cite every outside source you use. Taking credit for work as well as ideas that are not your own is plagiarism. Students who plagiarize will not get any score and they will be introduced in the class.

5. Please create reference for all sources(books, papers, websites) which you use.

6. Please create a cover letter for your report which is simply is the Homework#, title of the course, your name, surname, and student number.

7. You may post questions asking for clarifications and alternate perspectives on concepts on piazza or in the class.

8. Email your final file of assignment to *masoudbimar@gmail.com* and cc to *soheila.molaei2006@gmail.com* with subject [PGM hw# Surname] which # indicates number of the home work.

Questions

1 Expectation Maximization [60 points]

In this problem, you will implement the EM algorithm to learn the parameters of a two-class Gaussian mixture model. Recall that a mixture model is a density created by drawing each instance X from one of two possible distributions, $P(X|Y = 0)$ or $P(X|Y = 1)$. Y is a hidden variable over classes that simply indicates the distribution each instance is drawn from. We will assume that $P(Y)$ is a Bernoulli distribution and each $P(X|Y)$ is a 1-dimensional Gaussian with unit variance. The joint density is therefore:

$$P(X = x) = \sum_{y \in \{0,1\}} P(X = x|Y = y) \times P(Y = y)$$

$$P(X = x; \mu, \theta) = \sum_{y \in \{0,1\}} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(x - \mu_y)^2}{2}\} \times \theta_y$$

The parameters of this model are $\mu = [\mu_0, \mu_1]$ and $\theta = [\theta_0, \theta_1]$, where μ_y is the mean of the Gaussian for class y, and $\theta_y = P(Y = y)$ is the probability that an instance is drawn from class y. (Note that $\theta_0 + \theta_1 = 1$.) We will use EM to estimate these parameters from a data set $\{x^i\}_{i=1}^n$, where $x^i \in \mathbf{R}$.

1. [**5 points**] Let p_{iy} denote the probability that the ith instance is drawn from class y (i.e., $p_{iy} = P(Y = y|X = x^i)$). During the iteration t, the E-step computes p_{iy} for all i, y using the parameters from the previous iteration, $\mu^{(t-1)}$ and $\theta^{(t-1)}$. Write down an expression for p_{iy} in terms of these parameters.

2. [**5 points**] The M-step treats the p_{iy} variables as fractional counts for the unobserved y values and updates μ, θ as if the point (x^i, y) were observed p_{iy} times. Write down an update equation for $\mu^{(t)}$ and $\theta^{(t)}$ in terms of p_{iy}.

3. [**30 points**] Implement EM using the equations you derived in parts 1 and 2. Print out your code and submit it with your solution.

4. [**10 points**] Download the data set from the course channel. Each row of this file is a training instance x^i. Run your EM implementation on this data, using $\mu = [1, 2]$ and

$\theta = [.33, .67]$ as your initial parameters. What are the final values of μ and θ? Plot a histogram of the data and your estimated mixture density $P(X)$. Is the mixture density an accurate model for the data?

To plot the density in Matlab, you can use:

```
density = @(x) (<class 1 prior> * normpdf(x, <class 1 mean>, 1)) + ...
(<class 2 prior> * normpdf(x, <class 2 mean>, 1));
fplot(density, [-5, 6]);
```

Recall from class that EM attempts to maximize the marginal data loglikelihood $\ell(\mu, \theta) = \sum_{i=1}^{n} \log P(X = x^i; \mu, \theta)$, but that EM can get stuck in local optima. In this part, we will explore the shape of the loglikelihood function and determine if local optima are a problem. For the remainder of the problem, we will assume that both classes are equally likely, i.e., $\theta_y = \frac{1}{2}$ for $y = 0, 1$. In this case, the data loglikelihood ℓ only depends on the mean parameters μ.

1. [**10 points**] Create a contour plot of the loglikelihood ℓ as a function of the two mean parameters, μ. Vary the range of each μ_k from -1 to 4, evaluating the loglikelihood at intervals of .25. You can create a contour plot in Matlab using the `contourf` function. Print out your plot and include in with your solution.

   Does the loglikelihood have multiple local optima? Is it possible for EM to find a non-globally optimal solution? Why or why not?