CSCI 567, Fall 2021
Haipeng Luo

Written Assignment #4

Due: Nov 09, 2021, 11:59 pm, PT

Instructions

Submission: Assignment submission will be via `courses.uscden.net`. By the submission date, there will be a folder set up in which you can submit your files. Please be sure to follow all directions outlined here.

You can submit multiple times, but only *the last submission* counts. That means if you finish some problems and want to submit something first and update later when you finish, that's fine. In fact you are encouraged to do this: that way, if you forget to finish the homework on time or something happens, you still get credit for whatever you have turned in.

Problem sets must be typewritten or neatly handwritten when submitted. In both cases, your submission must be a single PDF. Please also follow the rules below:

- The file should be named as `firstname_lastname_USCID.pdf` (e.g., Joe_Doe_1234567890.pdf).
- Do not have any spaces in your file name when uploading it.
- Please include your name and USCID in the header of the report as well.

Total points: 40 points

Notes on notation:

- Unless stated otherwise, scalars are denoted by small letter in normal font, vectors are denoted by small letters in bold font and matrices are denoted by capital letters in bold font.
- $\|.\|$ means L2-norm unless specified otherwise i.e. $\|.\| = \|.\|_2$

Problem 1 Optimization over the simplex (14 points)

In this exercise you will prove two optimization results over the simplex that we used multiple times in the lectures. These results will also help you solve other problems in this homework.

The $K-1$ dimensional simplex is simply the set of all distributions over K elements, denoted by $\Delta = \{\mathbf{q} \in \mathbb{R}^K \mid q_k \geq 0, \ \forall k \text{ and } \sum_{k=1}^K q_k = 1\}$.

1.1 Let $a_1, \ldots, a_K$ be K positive numbers. Prove that the solution of the following optimization problem

$$\operatorname*{arg\,max}_{\mathbf{q}\in\Delta} \sum_{k=1}^{K} a_k \ln q_k$$

is $\mathbf{q}^*$ such that $q_k^* = \frac{a_k}{\sum_{k'} a_{k'}}$ (that is, $q_k^* \propto a_k$). Hint: the Lagrangian of this problem is

$$L(\mathbf{q}, \lambda, \lambda_1, \ldots, \lambda_K) = \sum_{k=1}^{K} a_k \ln q_k + \lambda \left(\sum_{k=1}^{K} q_k - 1 \right) + \sum_{k=1}^{K} \lambda_k q_k$$

for Lagrangian multipliers $\lambda \neq 0$ and $\lambda_1, \ldots, \lambda_k \geq 0$. Now apply KKT conditions to find $\mathbf{q}^*$. **(5 points)**

1.2 Let $b_1, \ldots, b_K$ be K real numbers and H be the entropy function. Prove that the solution of the following optimization problem

$$\operatorname*{arg\,max}_{\mathbf{q}\in\Delta} \mathbf{b}^T\mathbf{q} + H(\mathbf{q}) = \operatorname*{arg\,max}_{\mathbf{q}\in\Delta} \sum_{k=1}^{K} \left(q_k b_k - q_k \ln q_k \right)$$

is $\mathbf{q}^*$ such that $q_k^* \propto e^{b_k}$. Hint: follow the exact same steps as in the previous problem, that is, write down the Lagrangian and then apply KKT conditions. **(7 points)**

1.3 In the lecture we derived EM through a lower bound of the log-likelihood function. Specifically, on Slide 42 of Lecture 8, we find the tightest lower bound by solving

$$\operatorname*{arg\,max}_{\mathbf{q}_n\in\Delta} \mathbb{E}_{z_n \sim \mathbf{q}_n} \left[\ln p(\mathbf{x}_n, z_n \,; \theta^{(t)}) \right] + H(\mathbf{q}_n).$$

Use the result from Problem 1.2 to find the solution (you already know what it is from Slide 42). **(2 points)**

Problem 2 Gaussian Mixture Model (8 points)

In the lecture we applied EM to learn Gaussian Mixture Models (GMMs) and showed the M-Step without a proof on Slide 48 of Lecture 8 . In this problem you will prove this for the simpler one-dimensional case. Specifically consider a one-dimensional GMM that has the following density function for x:

$$p(x) = \sum_{k=1}^{K} \omega_k \mathcal{N}(x \mid \mu_k, \sigma_k) = \sum_{k=1}^{K} \frac{\omega_k}{\sqrt{2\pi}\sigma_k} \exp\left(\frac{-(x-\mu_k)^2}{2\sigma_k^2}\right)$$

where:

- K is the number of Gaussians components
- μ_k and σ_k^2 are the mean and the variance of the k-th component
- ω_k is the mixture weight for component k and satisfies:

$$\forall k, \omega_k > 0 \text{ and } \sum_k \omega_k = 1.$$

Prove that the maximizer of the expected complete log-likelihood (with γ_{nk} being the posterior of latent variables computed from the previous E-Step)

$$\sum_n \sum_k \gamma_{nk} \ln \omega_k + \sum_n \sum_k \gamma_{nk} \ln \mathcal{N}(x_n \mid \mu_k, \sigma_k)$$

is the following

$$\omega_k = \frac{\sum_n \gamma_{nk}}{N}, \qquad \mu_k = \frac{1}{\sum_n \gamma_{nk}} \sum_n \gamma_{nk} x_n, \qquad \sigma_k^2 = \frac{1}{\sum_n \gamma_{nk}} \sum_n \gamma_{nk} (x_n - \mu_k)^2.$$

Hint: you can make use of the result from Problem 1.1.

Problem 3 EM (18 points)

Consider the following probabilistic model to generate a non-negative integer x. First, draw a hidden binary variable z from a Bernoulli distribution with mean $\pi \in [0, 1]$, that is, $p(z = 1; \pi) = \pi$ and $p(z = 0; \pi) = 1 - \pi$. If $z = 0$, set $x = 0$; otherwise, draw x from a Poisson distribution with parameter λ so that

$$p(x|z = 1; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}.$$

Given N samples $x_1, \ldots, x_N$ generated independently in this way, follow the steps below to derive the EM algorithm for this model. (This is also a good example to see why finding the exact MLE is difficulty; you might try it yourself.)

3.1 Fixing the model parameters π and λ, for each sample n, find the posterior distribution of the corresponding hidden variable z_n: $p(z_n|x_n; \pi, \lambda)$. You will find it useful to consider the cases $x_n > 0$ and $x_n = 0$ separately. Given that z_n is binary, this means that you have to find the value of the following four quantities:

- $\gamma'_0 = p(z_n = 0|x_n > 0; \pi, \lambda)$,
- $\gamma'_1 = p(z_n = 1|x_n > 0; \pi, \lambda)$,
- $\gamma_0 = p(z_n = 0|x_n = 0; \pi, \lambda)$,
- $\gamma_1 = p(z_n = 1|x_n = 0; \pi, \lambda)$. **(6 points)**

3.2 Suppose that we have computed $\gamma_0, \gamma_1, \gamma'_0, \gamma'_1$ from the previous value of π and λ. Now, write down the expected complete likelihood function $Q(\pi, \lambda)$ in terms of the data $x_1, \ldots, x_n$, the posteriors $\gamma_0, \gamma_1, \gamma'_0, \gamma'_1$, and the parameters π and λ (show intermediate steps). This completes the E-step. **(3 points)**

3.3 Find the maximizer π^* and λ^* for the function $Q(\pi, \lambda)$ from the previous question (show your derivation). You might find it convenient to use the notation $N_0 = |\{n : x_n = 0\}|$ (that is, the number of examples with value 0) in your solution. This completes the M-step. **(6 points)**

3.4 Combining all the results, write down the EM update formulas for π^{new} and λ^{new}, using only the data $x_1, \ldots, x_n$ and the previous parameter values π^{old} and λ^{old} (do not use $\gamma_0, \gamma_1, \gamma'_0, \gamma'_1$). **(3 points)**