NYU Center for Data Science: DS-GA 1003

Machine Learning and Computational Statistics (Spring 2018)

Brett Bernstein

May 12, 2018

Instructions: Following most lab and lecture sections, we will be providing concept checks for review. Each concept check will:

- List the lab/lecture learning objectives. You will be responsible for mastering these objectives, and demonstrating mastery through homework assignments, exams (midterm and final), and on the final course project.
- Include concept check questions. These questions are intended to reinforce the lab/lectures, and help you master the learning objectives.

You are strongly encourage to complete all concept check questions, and to discuss these (and related) problems on Piazza and at office hours. However, problems marked with a ($\star$) are considered optional.

EM: Concept Check

EM/Mixture Model Objectives

- Write down the probability model corresponding to the GMM problem (multinomial distribution on mixture component z, multivariate Gaussian conditionals $x|z$).
- Write down the joint density $p(x, z)$ for the GMM model.
- Give an expression for the marginal log-likelihood for the observed data x for the GMM model, and explain why it doesn't simplify as nicely as the log-likelihood of a multivariate Gaussian model.
- Give pseudocode for the EM Algorithm for GMM (as in slide 29).
- Give an expression for the probability model for a generic latent variable model, where x is observed, z is latent (i.e. unobserved), and the parameters are represented by θ.
- Give EM algorithm pseudocode (as in slide 27).

EM Question

Poisson Mixture Model Setup: Consider the poisson mixture model, where each data instance is generated as follows:

1. Draw an [unobserved] cluster assignment z from a multinomial distribution $\pi = (\pi_1, \cdots, \pi_k)$ on k clusters.

2. Draw a count from a Poisson distribution with PMF:

$$p(x \mid \lambda_z) = \frac{\lambda_z^x e^{-\lambda_z}}{x!},$$

   where $\lambda = (\lambda_1, \ldots, \lambda_k) \in (0, \infty)^k$.

To keep things concise, we'll write $\theta = (\pi, \lambda)$ to represent all of the unknown parameters.

Problems:

1. To start, let x, z be the count and cluster assignment for a single instance. Give an expression for $p(x, z \mid \theta)$ in terms of $p(z \mid \theta)$ and $p(x|z, \theta)$.

   Solution.

$$p(x, z \mid \theta) = p(z \mid \theta)p(x|z, \theta) = \pi_z \frac{\lambda_z^x e^{-\lambda_z}}{x!}$$

2. Give an expression for $p(x \mid \theta)$, the marginal distribution for a single observed x, in terms of π, λ, and x.

   Solution.

$$p(x \mid \theta) = \sum_z p(x, z \mid \theta) = \sum_z \pi_z \frac{\lambda_z^x e^{-\lambda_z}}{x!}$$

3. Give an expression for the conditional distribution $p(z \mid x, \theta)$ in terms of $p(x, z \mid \theta)$ and $p(x \mid \theta)$. (Basic probability review)

   Solution.

$$p(z|x, \theta) = \frac{p(x, z \mid \theta)}{p(x \mid \theta)}$$

4. Now assume we have some training set of size n. We observe $x = (x_1, \cdots, x_n)$, but don't observe $z = (z_1, \cdots, z_n)$. We'll work through the EM algorithm for this problem. First, let's tackle the "E step", in which we evaluate the responsibilities $\gamma_i^j = p(z_i = j|x_i)$ for each $j \in \{1, \cdots, k\}$. Give an expression for this responsibility for cluster j and instance i.

   Solution.

$$\gamma_i^j = p(z_i = j|x_i, \theta) = \frac{p(x_i|\lambda_j)}{\sum_{z=1}^k \pi_z p(x_i|\lambda_z)} = \frac{\pi_z \frac{\lambda_z^{x_i} e^{-\lambda_z}}{x_i!}}{\sum_{z=1}^k \pi_z \frac{\lambda_z^{x_i} e^{-\lambda_z}}{x_i!}}$$

5. Before we move on to the "M step", let's apply this "E step" result a toy problem. Imagine $k = 3$, and we have $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 3$. Find $p(z = 2|x = 1)$ in terms of π_i for i in $\{1, 2, 3\}$. Hint: Note $p(x)$ is constant for all k, so its straightforward to give proportional expressions for each of $p(z = k|x = 1)$ then normalize.

   Solution.

$$p(z = 1|x = 1) \propto p(x = 1|z = 1)p(z = 1) = \pi_1 e^{-1}$$
$$P(z = 2|x = 1) \propto p(x = 1|z = 2)p(z = 2) = \pi_2 2e^{-2}$$
$$P(z = 3|x = 1) \propto p(x = 1|z = 3)p(z = 3) = \pi_3 3e^{-3}$$

$$P(z = 2|X = 1) = \frac{\pi_2 2e^{-2}}{\pi_1 e^{-1} + \pi_2 2e^{-2} + \pi_3 3e^{-3}}$$

6. Now we will tackle the "M step" of the EM algorithm, during which we will update π_z and λ_z. To start, find our objective (the expectation of the complete log likelihood).

   Solution. The complete log-likelihood is

$$\begin{aligned} p(x, z|\lambda) &= \sum_{i=1}^{n} \log \left[\pi_z \frac{\lambda_z^{x_i} e^{-\lambda_z}}{x_i!} \right] \\ &= \sum_{i=1}^{n} [\log \pi_z + x_i \log(\lambda_z) - \lambda_z - \log x_i!] \end{aligned}$$

   Taking the expected complete log likelihood with respect to q^* yields

$$\begin{aligned} J(\lambda, \pi) &= \sum_{z=1}^{k} q^*(z) \log p(x, z|\lambda) \\ &= \sum_{i=1}^{n} \sum_{i=1}^{k} \gamma_i^j [\log \pi_z + x_i \log(\lambda_z) - \lambda_z - \log x_i!] \end{aligned}$$

7. Finally give the expression for λ_z^{new} (that maximizes the objective you found above).

   Solution.

$$\frac{\partial J(\lambda, \pi)}{\partial \lambda_z} = \sum_{i=1}^{n} \gamma_i^z \left[-1 + \frac{x_i}{\lambda_z} \right]$$

   Setting this equal to 0 and solving yields

$$\sum_{i=1}^{n} \gamma_i^z \left[-1 + \frac{x_i}{\lambda_z}\right] = 0$$

$$\sum_{i=1}^{n} \gamma_i^z \frac{x_i}{\lambda_z} = \sum_{i=1}^{n} \gamma_i^z$$

$$\frac{\sum_{i=1}^{n} \gamma_i^z x_i}{\sum_{i=1}^{n} \gamma_i^z} = \lambda_z^{new}$$