Solution Part I: Exponential Mixture Models

Rebecca C. Steorts

Please read Ghojogh et. al (2020): https://arxiv.org/pdf/1901.06708, which provides a tutorial regarding mixture models and the EM algorithm. They authors provide nice derivations for the generalized case that we have talked about in class, the Gaussian mixture model, and the Poisson mixture model.

Specifically, if you had a question in class on Wednesday, November 6th, the answer likely lies in the tutorial. For example, Section 3 has a very nice treatment of the likelihood, log-likelihood, and how the latent component is important here. Please read through this in detail and let me know if you have follow up questions.

In this document, I will provide the derivations for the exponential mixture model, relying on some proofs in the tutorial.

Model Formulation and Notation

Assume that we have a mixture of K exponential distributions.

Following the notation in the article, assume that

$$g_k(x, \lambda_k) = \lambda_k e^{-\lambda_k x}, x > 0$$

From equation (1) in the article, it follows that

$$f(x, \lambda_1, \dots, \lambda_K) = \sum_{k=1}^{K} \pi_k \lambda_k e^{-\lambda_k x},$$

which is a K-component exponential distribution with rate parameter λ_k and mixing proportions (or weights) π_k.

E-Step

From equation (38) in the article, we can update γ_{ik} as follows, where I am omitting hat notation:

$$\gamma_{ik} = \frac{\pi_k \lambda_k e^{-\lambda_k x_i}}{\sum_{j=1}^{K} \pi_j \lambda_j e^{-\lambda_j x_i}}$$

M-Step

From section 3, it follows that (using a latent variable trick to make Q easier to work with)

$$\begin{aligned} Q(\lambda_1, \ldots, \lambda_K) &= \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{ik} \log[\pi_k g_k(x_i, \lambda_k)] && (1) \\ &= \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{ik} \log[\pi_k \lambda_k e^{-\lambda_k x_i}] && (2) \\ &= \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{ik} [\log \pi_k + \log \lambda_k - \lambda_k x_i] && (3) \\ & && (4) \end{aligned}$$

Recall that we have a constraint in our problem, which is $\sum_{k=1}^{K} \pi_k = 1$ and this must be taken into account.

The Lagrangian becomes

$$\mathcal{L}(\lambda, \pi, \alpha) = \sum_{i=1}^{n} \sum_{k=1}^{K} [\gamma_{ik} \log \pi_k + \gamma_{ik} \log \lambda_k - \gamma_{ik} \lambda_k x_i] - \alpha(\sum_{k=1}^{K} \pi_k - 1). \quad (5)$$

This implies that

$$\frac{\partial \mathcal{L}}{\partial \lambda_k} = \sum_{i=1}^{n} \left[\frac{\gamma_{ik}}{\lambda_k} - \gamma_{ik} x_i \right] = 0 \implies \tag{6}$$

$$\frac{\sum_{i=1}^{n} \gamma_{ik}}{\lambda_k} = \sum_{i=1}^{n} \gamma_{ik} x_i \implies \tag{7}$$

$$\lambda_k = \frac{\sum_{i=1}^{n} \gamma_{ik}}{\sum_{i=1}^{n} \gamma_{ik} x_i} \tag{8}$$

We can use a similar approach to find π_k or we can utilize the fact that this can be proven generally and is always updated the same way as illustrated on page 7 and equation (41) of the tutorial.

Specifically, using the tutorial, it follows directly that

$$\pi_k = \frac{\sum_{i=1}^{n} \gamma_{ik}}{\sum_{i=1}^{n} \sum_{j=1}^{K} \gamma_{ij}} = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ik}.$$

You can also derive this directly from the Lagrangian as follows:

$$\frac{\partial \mathcal{L}}{\partial \pi_k} = \frac{\sum_{i=1}^{n} \gamma_{ik}}{\pi_k} - \alpha = 0 \implies \tag{9}$$

$$\pi_k = \frac{1}{\alpha} \sum_{i=1}^{n} \gamma_{ik} \implies \tag{10}$$

$$\sum_{k=1}^{K} \pi_k = \frac{1}{\alpha} \sum_{k=1}^{K} \sum_{i=1}^{n} \gamma_{ik} \implies \tag{11}$$

$$1 = \frac{1}{\alpha} \sum_{k=1}^{K} \sum_{i=1}^{n} \gamma_{ik} \implies \tag{12}$$

$$\alpha = \frac{1}{\sum_{k=1}^{K} \sum_{i=1}^{n} \gamma_{ik}} \tag{13}$$

Putting this together, we find that

$$\pi_k = \frac{\sum_{i=1}^{n} \gamma_{ik}}{\sum_{i=1}^{n} \sum_{j=1}^{K} \gamma_{ij}} = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ik}.$$

Summary

Hopefully, you have now learned a few things:

1. Estimating the unknown parameters for the exponential mixture model using the EM algorithm
2. Recall that this does a type of soft clustering as discussed in class. The clusters can overlap. This contrasts the k-means algorithm, where the clusters cannot overlap.
3. What is an application we could apply this to that we have looked at previously in class?
4. Why did we need to derive these estimates if we want to work with the exponential mixture model? (Recall there is no code in R for this particular model).
5. Why is it good practice to know how to do derivations and write our own code?
6. You will explore coding up the mixture model in part II of the exercise for a simulation study.

You now have parameter estimates using the EM algorithm and should be able to utilize this into Part II to code these up.