Math189R SP19
Homework 6
Monday, Apr 1, 2019

Feel free to work with other students, but make sure you write up the homework and code on your own (no copying homework *or* code; no pair programming). Feel free to ask students or instructors for help debugging code or whatever else, though.

The starter files for problem 2 can be found under the Resource tab on course website. Please print out all the graphs generated by your own code and submit them together with the written part, and make sure you upload the code to your Github repository.

1 (Murphy 11.2 - EM for Mixtures of Gaussians) Show that the M step for ML estimation of a mixture of Gaussians is given by

$$\boldsymbol{\mu}_k = \frac{\sum_i r_{ik}\mathbf{x}_i}{r_k}$$
$$\boldsymbol{\Sigma}_k = \frac{1}{r_k}\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top = \frac{1}{r_k}\sum_i r_{ik}\mathbf{x}_i\mathbf{x}_i^\top - r_k\boldsymbol{\mu}_k\boldsymbol{\mu}_k^\top.$$

article amsmath

We have the complete data log likelihood

$$\begin{aligned}\mathcal{L}(\theta_k, \Sigma_k) &= \sum_k\sum_i r_{ik}\log p(x_i|\theta_k)\\ &= -\frac{1}{2}\sum_i r_{ik}\left(\log|\Sigma_k| + (x_i - \theta_k)^\top\Sigma_k^{-1}(x_i - \theta_k)\right) \quad \text{(to a constant)}\end{aligned}$$

Then differentiating with respect to θ_k we have

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\theta_k} &= \sum_i r_{ik}\Sigma_k^{-1}(x_i - \theta_k)\\ &= \Sigma_k^{-1}\sum_i r_{ik}(x_i - \theta_k) = \mathbf{0} \quad \text{(since } \Sigma^{-1} \text{ is linear)}\end{aligned}$$

so at optimality we have

$$\sum_i r_{ik}x_i = \theta_k\sum_i r_{ik},$$

which gives the desired result. Differentiating with respect to Σ_k we have

$$\frac{\partial\mathcal{L}}{\partial\Sigma_k} = -\frac{1}{2}\sum_i r_{ik}\left(\Sigma_k^{-1} - \Sigma_k^{-1}(x_i - \theta_k)(x_i - \theta_k)^\top\Sigma_k^{-1}\right) = \mathbf{0}.$$

This gives us the optimality condition that

$$\sum_i r_{ik} I = \sum_i r_{ik} (x_i - \theta_k)(x_i - \theta_k)^\top \Sigma_k^{-1}.$$

Multiplying by Σ_k on the right and dividing by $r_k = \sum_i r_{ik}$ gives the desired result.