Andy Liu
Math189R SU19
Homework 6
Tuesday, June 29, 2020

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

$$\boldsymbol{\mu}_k = \frac{\sum_i r_{ik}\mathbf{x}_i}{r_k}$$

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

(a) Note that the log likelihood given a $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ is equal to $-\frac{1}{2}\sum_k\sum_i r_{ik}(\log|\boldsymbol{\Sigma}_k| + (\mathbf{x}_i - \boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)))$. In order to find the optimal value of $\boldsymbol{\mu}_k$, which is what the M step is meant to compute, we can set the derivative of this (with respect to $\boldsymbol{\mu}_k$) and set it to zero. Here, the derivative of our log likelihood with respect to $\boldsymbol{\mu}_k$ is $\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)\boldsymbol{\Sigma}_k^{-1}$ (the log term is removed as it is a constant, and the derivative of $\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T$ is just $(\mathbf{x}_i - \boldsymbol{\mu}_k)$.

Note that this derivative is equal to 0 when $\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)\boldsymbol{\Sigma}_k^{-1} = 0$. Removing constants, note that this means that $\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k) = 0 \Rightarrow \sum_i r_{ik}\mathbf{x}_i = \sum_i r_{ik}\boldsymbol{\mu}_k$. Now, we can just solve for $\boldsymbol{\mu}_k$ to find that it is equal to $\frac{\sum_i r_{ik}\mathbf{x}_i}{\sum_i r_{ik}}$. Since $\sum_i r_{ik} = r_k$, this becomes $\frac{\sum_i r_{ik}\mathbf{x}_i}{r_k}$, as desired.

(b) Similarly to our strategy in part (a), we can take the derivative of the log likelihood with respect to $\boldsymbol{\Sigma}_k$ and set it equal to 0 to find our optimal value of $\boldsymbol{\Sigma}_k$, which is the goal of the M step. Note that the derivative of $-\frac{1}{2}\sum_k\sum_i r_{ik}(\log|\boldsymbol{\Sigma}_k| + (\mathbf{x}_i - \boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)))$ with respect to $\boldsymbol{\Sigma}_k$ is equal to $-\frac{1}{2}\sum_i r_{ik}(\boldsymbol{\Sigma}_k^{-1} - \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1})$.

Setting this derivative equal to zero, we find that $\sum_i r_{ik} \mathbf{\Sigma}_k^{-1} = \sum_i r_{ik} \mathbf{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \mathbf{\Sigma}_k^{-1}$. Multiplying by $\mathbf{\Sigma}_k^2$ (or multiplying by $\mathbf{\Sigma}_k$ from the left- and right-hand sides) on both sides, this equation becomes $\sum_i r_{ik} \mathbf{\Sigma}_k = \sum_i r_{ik} (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T$. Since $\sum_i r_{ik} = r_k$, we can solve for $\mathbf{\Sigma}_k$ to find that $\mathbf{\Sigma}_k = \frac{1}{r_k} \sum_i r_{ik} (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T$, as desired.

■