Eli weissler

Math189R SU17
Homework 6
Wednesday, June 7, 2017

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

$$\mu_k = \frac{\sum_i r_{ik}\mathbf{x}_i}{r_k}$$

$$\Sigma_k = \frac{1}{r_k}\sum_i r_{ik}(\mathbf{x}_i - \mu_k)(\mathbf{x}_i - \mu_k)^\top = \frac{1}{r_k}\sum_i r_{ik}\mathbf{x}_i\mathbf{x}_i^\top - r_k\mu_k\mu_k^\top.$$

First recall that $r_{ik} \triangleq (z_i = k \mid x_i, \theta^{(t-1)})$ is the responsibility that cluster k takes for data point i

In the M step, we optimize Q with respect to π and θ, where

$\pi_k = p(z_i = k)$ (i.e. the coefficients on the gaussian)

$\theta_k = (\mu_k, \Sigma_k)$ are the params for gaussian K

Recall

$$Q(\theta^t, \theta^{(t-1)}) = \sum_i \sum_K r_{ik} \log \pi_K + \sum_i \sum_K r_{ik} \log p(x_i \mid \theta_K)$$

■

For this problem we only care about the second term

$$\rightarrow \ell(\mu_K, \Sigma_K) = \sum_K \sum_i r_{iK} \log P(x_i | \theta_K)$$

$$= -\frac{1}{2} \sum_i r_{iK} \left[\log|\Sigma_K| + (x_i - \mu_K)^T \Sigma_K^{-1} (x_i - \mu_K) \right]$$

First we will find the best μ_K holding everything else constant

$$\frac{\partial \ell}{\partial \mu_K} = \frac{\partial}{\partial \mu_K} -\frac{1}{2} \sum_i r_{iK} \left(\log|\Sigma_K| - \Sigma_K^{-1} x_i - x_i^T \Sigma_K^{-1} + \Sigma_K^{-1} \mu_K + \mu_K^T \Sigma_K^{-1} \right)$$

$$= \sum_i r_{iK} \Sigma_K^{-1} (x_i - \mu_i) = \Sigma_K^{-1} \sum_i r_{iK} (x_i - \mu_i) = 0$$

Setting this equal to zero yields

$$\sum_i r_{iK} \mu_K = \sum_i r_{iK} \mu_K = \mu_K \sum_i r_{iK}$$

$$\rightarrow \mu_K = \frac{\sum_i r_{iK} x_i}{\sum_i r_{iK}} = \boxed{\frac{\sum_i r_{iK} x_i}{r_K}}$$

Next we find the best Σ_K, holding everything else constant

Using the fact that

$$\frac{\partial a^T X^{-1} b}{\partial X} = -X^{-1} a b X^{-1} \quad \& \quad \frac{\partial \det(X)}{\partial X} = \det(X) X^{-1}$$

we take

$$\frac{\partial}{\partial \Sigma_K} -\frac{1}{2} \sum_i r_{ik} \left(\log|\Sigma_K| + x_i^T \Sigma_K^{-1} x_i - \mu_K^T \Sigma_K^{-1} x_i - x_i^T \Sigma^{-1} \mu_K + \mu_K^T \Sigma_K^{-1} \mu_K \right)$$

$$= -\frac{1}{2} \sum_i r_{ik} \left(\frac{1}{|\Sigma_K|} \cdot |\Sigma_K| \Sigma^{-1} + \Sigma^{-1} (x_i - \mu_K)(x_i - \mu_K) \Sigma^{-1} \right)$$

$$= -\frac{1}{2} \sum_i r_{ik} \left(\Sigma^{-1} + \Sigma^{-1} (x_i - \mu_K)(x_i - \mu_K) \Sigma^{-1} \right) = 0$$

Setting equal to zero yields

$$\Sigma^{-1} \sum_i r_{ik} = \sum_i \Sigma^{-1} (x_i - \mu_K)(x_i - \mu_K) \Sigma^{-1}$$

$$\rightarrow \Sigma^{-1} = \frac{1}{r_K} \sum_i \Sigma^{-1} (x_i - \mu_K)(x_i - \mu_K) \Sigma^{-1}$$

Multiplying left and right by Σ yields

$$\boxed{\Sigma = \frac{1}{r_K} \sum_i (x_i - \mu_K)(x_i - \mu_K)}$$