1. Show M step for ML estimation of Gaussian mixture.

Oscar Scholin
189 HW 6

Gaussian mixture is data greated from several Gaussians with unknown params. EM alg. finds max likelihood estimates of the parameters.
E step finds expected val of log likelihood wrt conditional distr of latent variables given observed data & current estimate of params
M step maximizes expected vals.

The log likelihood

$$\ell(\vec{\mu}_k, \Sigma_k) = \sum_k \sum_i r_{ik} \log \mathbb{P}(\vec{x}_i \mid \vec{\theta}_k)$$

Note $\mathbb{P}(\vec{x}_i \mid \vec{\theta}_k)$ assuming multivariate Gaussian prior (i.e. what distribution the params should belong to)

$$\mathbb{P}(\vec{x}_i \mid \mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_k|}} \exp\left(-\frac{1}{2}(\vec{x}-\mu_k)^T \Sigma^{-1} (\vec{x}-\mu_k)\right)$$

where Σ_k is the covariance matrix, $|\Sigma_k| = \det \Sigma_k$.

$$\log \mathbb{P}(\vec{x} \mid \mu_k, \Sigma_k) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_k| - \frac{1}{2}(\vec{x}-\vec{\mu}_k)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_k)$$

$\therefore$ up to a constant,

$$\ell(\vec{\mu}_k, \Sigma_k) = -\frac{1}{2} \sum_i r_{ik} \left(\log |\Sigma_k| + (\vec{x}_i - \vec{\mu}_k)^T \Sigma_k^{-1} (\vec{x}_i - \vec{\mu}_k) \right)$$

Take derivative wrt $\vec{\mu}_k$:

$$\frac{\partial \ell}{\partial \vec{\mu}_k} = -\frac{1}{2} \sum_i r_{ik}\, 2 \Sigma_k^{-1} (\vec{x}_i - \vec{\mu}_k)$$

$$= -\sum_i r_{ik} \Sigma_k^{-1} (\vec{x}_i - \vec{\mu}_k)$$

$$= -\Sigma_k^{-1} \sum_i r_{ik} (\vec{x}_i - \vec{\mu}_k) \overset{!}{=} 0$$

$$\therefore \sum_i r_{ik} x_i = \vec{\mu}_k \sum_i r_{ik}$$

$$\therefore \vec{\mu}_k = \frac{\sum_i r_{ik} x_i}{r_k}$$

① https://statisticaloddsandends.wordpress.com/2018/05/24/derivative-of-log-det-x/

② https://www2.imm.dtu.dk/pubdb/edoc/imm3274.pdf
↳ equation 61, page 8

$$\left\{ \begin{aligned} \frac{\partial \log |X|}{\partial X} &= (X^{-1})^T \quad ① \\ \frac{\partial a^T X^{-1} b}{\partial X} &= -X^{-T} a b^T X^{-T} \quad ② \end{aligned} \right.$$

Now take derivative wrt Σ_k:

$$\frac{\partial \ell}{\partial \Sigma_k} = -\frac{1}{2} \sum_i r_{ik} \left(\Sigma_k^{-T} + \Sigma_k^{-T} (\vec{x}_i - \vec{\mu}_k)(\vec{x}_i - \vec{\mu}_k)^T \Sigma_k^{-T} \right)$$

$$= -\frac{1}{2}\sum_i r_{ik}\left(\vec{\Sigma}_k^{-1} + \vec{\Sigma}_k^{-1}(\vec{x}_i - \vec{\mu}_k)(\vec{x}_i - \vec{\mu}_k)^T \vec{\Sigma}_k^{-1}\right) \quad \text{since } \vec{\Sigma}_k \text{ is symmetric.}$$

$$= -\frac{1}{2}\vec{\Sigma}_k^{-1}\sum_i r_{ik} - \frac{1}{2}\vec{\Sigma}_k^{-1}\sum_i r_{ik}(\vec{x}_i - \vec{\mu}_k)(\vec{x}_i - \vec{\mu}_k)^T \vec{\Sigma}_k^{-1} \stackrel{!}{=} 0$$

$$\therefore \frac{1}{2}\vec{\Sigma}_k^{-1}\sum_i r_{ik} = \vec{\Sigma}_k^{-1}\sum_i r_{ik}(\vec{x}_i - \vec{\mu}_k)(\vec{x}_i - \vec{\mu}_k)^T \vec{\Sigma}_k^{-1}$$

$$\therefore \sum_i r_{ik} I = \sum_i r_{ik}(\vec{x}_i - \vec{\mu}_k)(\vec{x}_i - \vec{\mu}_k)^T \vec{\Sigma}_k^{-1}$$

$$\therefore \vec{\Sigma}_k = \frac{\sum_i r_{ik}(\vec{x}_i - \vec{\mu}_k)(\vec{x}_i - \vec{\mu}_k)^T}{r_k}$$

b) SVD image compression. Plot 1^{st} 100 singular values for the original image and randomly shuffled.