

**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

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

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

For the maximization step we find the  $\mu_k$  and  $\Sigma_k$  that maximizes the log likelihood. The book gives us the log likelihood.

$$\mu_k = \arg \max_{\mu_k} \left[ -\frac{1}{2} \sum_i r_{ik} (\log |\Sigma_k| + (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)) \right]$$

$$= \arg \min_{\mu_k} \sum_i r_{ik} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)$$

Now to find compute the minimum I take the derivative  $\mu_k$  and set it to 0. Then solve. The derivative with respect to  $\mu_k$  is,

$$\begin{aligned} \frac{\partial}{\partial \mu_k} \left( \sum_i r_{ik} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right) \\ &= \sum_i r_{ik} \left( \Sigma_k^{-1} (x_i - \mu_k) + \Sigma_k^{-T} (x_i - \mu_k) \right) \\ &= 2 \Sigma_k^{-1} \sum_i r_{ik} (x_i - \mu_k) \end{aligned} \quad \text{Since } \Sigma^{-1} \text{ is symmetric}$$

Now I set the derivative equal to 0 and solve for  $\mu_k$ ,

$$\begin{aligned} 2 \Sigma_k^{-1} \sum_i r_{ik} (x_i - \mu_k) &= 0 \\ \implies \sum_i r_{ik} (x_i - \mu_k) &= 0 \\ \implies \sum_i r_{ik} x_i &= \sum_i r_{ik} \mu_k \\ \implies \mu_k &= \frac{\sum_i r_{ik} x_i}{\sum_i r_{ik}} \\ &= \boxed{\frac{\sum_i r_{ik} x_i}{r_k}} \end{aligned}$$

I used the solution for this part.  
Now I'll repeat to find  $\Sigma_k^{-1}$ ,

$$\begin{aligned}\Sigma_k &= \arg \max_{\Sigma_k} \left[ -\frac{1}{2} \sum_i r_{ik} (\log |\Sigma_k| + (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)) \right] \\ &= \arg \min_{\Sigma_k} \sum_i r_{ik} (\log |\Sigma_k| + (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k))\end{aligned}$$

The derivative with respect to  $\Sigma_k^{-1}$  is,

$$\begin{aligned}& \frac{\partial}{\partial \Sigma_k^{-1}} \sum_i r_{ik} (\log |\Sigma_k| + (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)) \\ &= \sum_i r_{ik} \left( \frac{1}{|\Sigma_k|} |\Sigma_k| \Sigma_k^{-1} - \Sigma_k^{-1} (x_i - \mu_k) (x_i - \mu_k)^T \Sigma_k^{-1} \right) \\ &= \Sigma_k^{-1} \sum_i r_{ik} (I - (x_i - \mu_k) (x_i - \mu_k)^T \Sigma_k^{-1})\end{aligned}$$

Set the derivative equal to 0 and solve for  $\Sigma_k$ ,

$$\begin{aligned}\Sigma_k^{-1} \sum_i r_{ik} (I - (x_i - \mu_k) (x_i - \mu_k)^T \Sigma_k^{-1}) &= 0 \\ \implies \sum_i r_{ik} (I - (x_i - \mu_k) (x_i - \mu_k)^T \Sigma_k^{-1}) &= 0 \\ \implies I \sum_i r_{ik} &= \left( \sum_i r_{ik} (x_i - \mu_k) (x_i - \mu_k)^T \right) \Sigma_k^{-1} \\ \implies \Sigma_k \sum_i r_{ik} &= \sum_i r_{ik} (x_i - \mu_k) (x_i - \mu_k)^T \\ \implies \Sigma_k &= \frac{\sum_i r_{ik} (x_i - \mu_k) (x_i - \mu_k)^T}{\sum_i r_{ik}}\end{aligned}$$

■

**2 (SVD Image Compression)** In this problem, we will use the image of a scary clown online to perform image compression. In the starter code, we have already load the image into a matrix/array for you. However, you might need internet connection to access the image and therefore successfully run the starter code. The code requires Python library Pillow in order to run.

Plot the progression of the 100 largest singular values for the original image and a randomly shuffled version of the same image (all on the same plot). In a single figure plot a grid of four images: the original image, and a rank  $k$  truncated SVD approximation of the original image for  $k \in \{2, 10, 20\}$ .

■