

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

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

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

From page 351 of the textbook, we see that

$$\ell(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \sum_k \sum_i r_{ik} \log p(\mathbf{x}_i | \theta_k) \quad (1)$$

$$= -\frac{1}{2} \sum_i r_{ik} [\log |\boldsymbol{\Sigma}_k| + (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k)]. \quad (2)$$

We shall now take the respective gradients and set the equal to zero to obtain our optimality condition. Starting with $\boldsymbol{\mu}_k$ (using product rule):

$$\nabla_{\boldsymbol{\mu}_k} \ell = \sum_i r_{ik} \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k) \quad (3)$$

$$= \boldsymbol{\Sigma}_k^{-1} \sum_i r_{ik} (\mathbf{x}_i - \boldsymbol{\mu}_k) \quad (4)$$

$$= 0 \quad (5)$$

$$\implies \sum_i r_{ik} \mathbf{x}_i = \boldsymbol{\mu}_k \sum_i r_{ik} \implies \boldsymbol{\mu}_k = \frac{\sum_i r_{ik} \mathbf{x}_i}{r_k}. \quad (6)$$

Then for $\boldsymbol{\Sigma}_k$:

$$\nabla_{\boldsymbol{\Sigma}_k} \ell = -\frac{1}{2} \sum_i r_{ik} [\boldsymbol{\Sigma}_k^{-1} - \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}] = 0 \quad (7)$$

$$(8)$$

Cancelling a Σ_k^{-1} and the $-\frac{1}{2}$ and rearranging we have

$$\sum_i r_{ik} [I - (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \Sigma_k^{-1}] = 0 \quad (9)$$

$$\implies \sum_i r_{ik} = \sum_i r_{ik} (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} \quad (10)$$

then right multiplying by Σ_k and dividing by r_k :

$$\Sigma_k^{-1} = \frac{1}{r_k} \sum_i r_{ik} \mathbf{x}_i \mathbf{x}_i^\top - r_k \boldsymbol{\mu}_k \boldsymbol{\mu}_k^\top, \quad (11)$$

as desired. ■

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 $k \in \{2, 10, 20\}$.

See images in starter file folder. ■