

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

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

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

The data log likelihood is as follows:

$$l(\mu_k, \Sigma_k) = \sum_k \sum_i r_{ik} \log \mathbb{P}(\mathbf{x}_i | \theta_k)$$

$$= -0.5 \sum_i r_{ik} (\log |\Sigma_k| + (\mathbf{x}_i - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x}_i - \mu_k))$$

Next, take the partial with respect to μ_k :

$$\frac{\partial l}{\partial \mu_k} = \Sigma_k^{-1} \sum_i r_{ik} (\mathbf{x}_i - \mu_k) = 0$$

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

Had to look at solutions for this step:

Now, take the derivative with respect to Σ_k , we will get:

$$\frac{\partial l}{\partial \Sigma_k} = -0.5 \sum_i r_{ik} (\Sigma_k^{-1} - \Sigma_k^{-1} (\mathbf{x}_i - \mu_k)^\top \Sigma_k^{-1}) = 0$$

Then we can multiply by Σ_k and divide by $r_k = \sum_i r_{ik}$ to get the desired result ■

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

I've included my graph and pictures as part of the submission to github

■