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

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

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

First expand

$$\boldsymbol{\mu}_k\boldsymbol{\mu}_k^\top = \frac{\sum_i r_{ik}\mathbf{x}_i}{r_k}\frac{\sum_j r_{jk}\mathbf{x}_j^\top}{r_k}$$

$$= \frac{1}{r_k^2}(\sum_i r_{ik}\mathbf{x}_i)(\sum_j r_{jk}\mathbf{x}_j^\top)$$

$$=$$

$$\boldsymbol{\Sigma}_k = \frac{1}{r_k}\sum_i r_{ik}(\mathbf{x}_i - \frac{\sum_j r_{jk}\mathbf{x}_j}{r_k})(\mathbf{x}_i - \frac{\sum_j r_{lk}\mathbf{x}_l}{r_k})^\top$$

$$= \frac{1}{r_k}\sum_{ijl} r_{ik}(\mathbf{x}_i - \frac{r_{jk}\mathbf{x}_j}{r_k})(\mathbf{x}_i - \frac{r_{lk}\mathbf{x}_l}{r_k})^\top$$

$$= \frac{1}{r_k}\sum_{ijl} r_{ik}(\mathbf{x}_i\mathbf{x}_i^\top - (r_{lk}\mathbf{x}_i\mathbf{x}_l^\top + r_{jk}\mathbf{x}_j\mathbf{x}_i^\top)\frac{1}{r_k} + \frac{r_{jk}r_{lk}\mathbf{x}_j\mathbf{x}_l^\top}{r_k^2})$$

$$= \frac{1}{r_k}\sum_i r_{ik}\mathbf{x}_i\mathbf{x}_i^\top - \sum_{ijl}\frac{r_{ik}}{r_k^2}(r_{lk}\mathbf{x}_i\mathbf{x}_l^\top + r_{jk}\mathbf{x}_j\mathbf{x}_i^\top) - \frac{r_{ik}r_{jk}r_{lk}\mathbf{x}_j\mathbf{x}_l^\top}{r_k^3}$$

$$= \frac{1}{r_k}\sum_i r_{ik}\mathbf{x}_i\mathbf{x}_i^\top \ldots$$

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

Checked solution; this was not the correct approach.

■

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 hw6pr2 in repo