Colin Adams
Math189R SU20
Homework 6
June 2020

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 Murphy, the part of Q that is dependent on $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ is contained in

$$l(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \sum_{k,i} r_{ik} \log p(\mathbf{x}_i|\boldsymbol{\theta}_k) = -\frac{1}{2}\sum_i r_{ik}\big[\log(\det(\boldsymbol{\Sigma}_k)) + (\mathbf{x}_i - \boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\big]$$

where we ignore any constants not related $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$. To solve for $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ we will need to take a gradient of l. Let's start with $\boldsymbol{\mu}_k$:

$$\frac{\mathrm{d}l}{\mathrm{d}\boldsymbol{\mu}_k} = -\frac{1}{2}\sum_i r_{ik}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\mu}_k}\big((\mathbf{x}_i - \boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\big).$$

Note that

$$(\mathbf{x}_i - \boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k) = \mathbf{x}_i^T\boldsymbol{\Sigma}_k^{-1}\mathbf{x}_i - \mathbf{x}_i^T\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^T\boldsymbol{\Sigma}_k^{-1}\mathbf{x}_i + \boldsymbol{\mu}_k^T\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k$$

so

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\mu}_k}\big((\mathbf{x}_i - \boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\big) = 2\big(\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\big)$$

which implies

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

To find the optimal solution, we need to set this equal to zero, and solve for $\boldsymbol{\mu}_k$ which gives us

$$0 = \sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k) \qquad \Rightarrow \boldsymbol{\mu}_k = \frac{\sum_i r_{ik}\mathbf{x}_i}{r_k} \quad \text{where} \quad r_k \equiv \sum_i r_{ik}.$$

Now for the gradient with respect to $\mathbf{\Sigma}_k$, we have

$$\begin{aligned}
\frac{\mathrm{d}l}{\mathrm{d}\mathbf{\Sigma}_k} &= -\frac{1}{2}\sum_i r_{ik}\frac{\mathrm{d}}{\mathrm{d}\mathbf{\Sigma}_k}\big(\log(\det(\mathbf{\Sigma}_k)) + (\mathbf{x}_i - \boldsymbol{\mu}_k)^T\mathbf{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\big) \\
&= -\frac{1}{2}\sum_i r_{ik}\bigg(\mathbf{\Sigma}_k^{-T} + \frac{\mathrm{d}}{\mathrm{d}\mathbf{\Sigma}_k}(\mathbf{x}_i - \boldsymbol{\mu}_k)^T\mathbf{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\bigg) \\
&= -\frac{1}{2}\sum_i r_{ik}\big(\mathbf{\Sigma}_k^{-1} - \mathbf{\Sigma}_k^{-T}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T\mathbf{\Sigma}_k^{-T}\big) \\
&= -\frac{1}{2}\sum_i r_{ik}\big(\mathbf{\Sigma}_k^{-1} - \mathbf{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T\mathbf{\Sigma}_k^{-1}\big)
\end{aligned}$$

which we set to zero and solve for $\mathbf{\Sigma}_k$. Doing so, we get

$$0 = \sum_i r_{ik}\big(\mathbf{\Sigma}_k^{-1} - \mathbf{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T\mathbf{\Sigma}_k^{-1}\big)$$

implying the next few steps of algebra:

$$\begin{aligned}
\sum_i r_{ik}\mathbf{\Sigma}_k^{-1} &= \sum_i \big(r_{ik}\mathbf{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T\mathbf{\Sigma}_k^{-1}\big) \\
\mathbf{\Sigma}_k\mathbf{\Sigma}_k^{-1}\sum_i r_{ik} &= \mathbf{\Sigma}_k\sum_i \big(r_{ik}\mathbf{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T\mathbf{\Sigma}_k^{-1}\big) \\
r_k\mathbf{I} &= \sum_i \big(r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T\mathbf{\Sigma}_k^{-1}\big) \\
r_k\mathbf{\Sigma}_k &= \sum_i \big(r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T\mathbf{\Sigma}_k^{-1}\mathbf{\Sigma}_k\big)
\end{aligned}$$

and thus, gives us our final answer

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

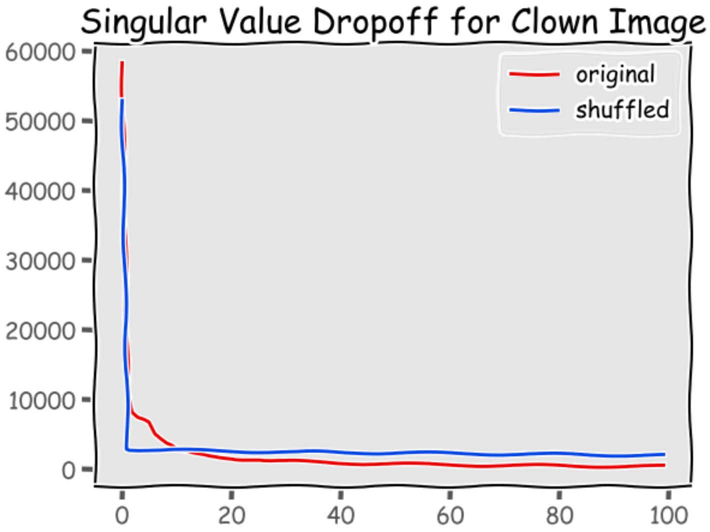
as desired.

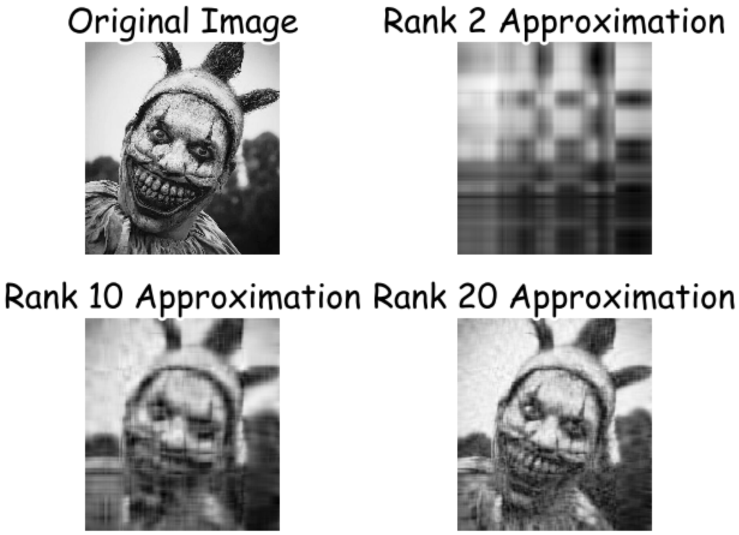
∎

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

Dropoff and reconstruction plots are below.





Pretty neat to be honest. It's encouraging that the eigenvalues of the dropoff drops off so much for the shuffled data. Makes sense logically but still neat. ■