Nico Espinosa Dice
Math189R SP19
Homework 6
Monday, March 9, 2020

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, we begin with the result of the M step presented in Murphy Equation 11.30 (Section 11.4.2.3):

$$\begin{aligned} l(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) &= -\frac{1}{2}\sum_i r_{ik}[\log|\boldsymbol{\Sigma}_k| + (\mathbf{x}_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)] \\ &= -\frac{1}{2}\sum_i r_{ik}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k). \end{aligned}$$

Next, we will find the optimal values for $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ by finding their critical points. Taking the derivative of $l(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ with respect to $\boldsymbol{\mu}_k$, we have:

$$\begin{aligned} \frac{\partial l(\mu_k, \boldsymbol{\Sigma}_k)}{\partial \boldsymbol{\mu}_k} &= \frac{\partial}{\partial \boldsymbol{\mu}_k}\left(-\frac{1}{2}\sum_i r_{ik}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\right) \\ &= \frac{\partial}{\partial \boldsymbol{\mu}_k}\left(-\frac{1}{2}\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\right) \\ &= \frac{\partial}{\partial U}\left(-\frac{1}{2}\sum_i r_{ik}U^\top \boldsymbol{\Sigma}_k^{-1}U\right)\frac{dU}{d\boldsymbol{\mu}_k}, \text{ where } U = (\mathbf{x}_i - \boldsymbol{\mu}_k), \\ &= \left(-\frac{1}{2}\sum_i r_{ik}\boldsymbol{\Sigma}_k^{-1}U - \frac{1}{2}\sum_i r_{ik}\boldsymbol{\Sigma}_k^{-1}U\right)(-1), \text{ since } \frac{\partial U^T x}{\partial U} = \frac{\partial x^T U}{\partial U} = x, \\ &= \sum_i r_{ik}\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k). \end{aligned}$$

Now we must find the critical point for $\boldsymbol{\mu}_k$. Setting the derivative of $l(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ with respect to $\boldsymbol{\mu}_k$ to 0 and solving for $\boldsymbol{\mu}_k$, we have:

$$
\begin{aligned}
\sum_i r_{ik}\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k) = 0 \\
\sum_i r_{ik}\boldsymbol{\Sigma}_k^{-1}\mathbf{x}_i &= \sum_i r_{ik}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k \\
\sum_i r_{ik}\mathbf{x}_i &= \sum_i r_{ik}\boldsymbol{\mu}_k, \text{ since the covariance matrix is a linear operator} \\
\sum_i r_{ik}\mathbf{x}_i &= \boldsymbol{\mu}_k \sum_i r_{ik}, \text{ since } \boldsymbol{\mu}_k \text{ does not depend on } i, \\
\sum_i r_{ik}\mathbf{x}_i &= \boldsymbol{\mu}_k r_k, \text{ so} \\
\boldsymbol{\mu}_k &= \frac{\sum_i r_{ik}\mathbf{x}_i}{r_k}.
\end{aligned}
$$

We see that this yields the value of $\boldsymbol{\mu}_k$ presented in Murphy, Equation 11.31.

Next, we will take the derivative of $l(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ with respect to $\boldsymbol{\Sigma}_k$:

$$
\begin{aligned}
\frac{\partial l(\mu_k, \boldsymbol{\Sigma}_k)}{\partial \boldsymbol{\Sigma}_k} &= \frac{\partial}{\partial \boldsymbol{\Sigma}_k}\left(-\frac{1}{2}\sum_i r_{ik}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\right) \\
&= \frac{\partial}{\partial \boldsymbol{\Sigma}_k}\left(-\frac{1}{2}\sum_i r_{ik}\log|\boldsymbol{\Sigma}_k|\right) - \frac{\partial}{\partial \boldsymbol{\Sigma}_k}\left(\frac{1}{2}\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\right) \\
&= -\frac{1}{2}\sum_i r_{ik}\boldsymbol{\Sigma}_k^{-1} - \frac{\partial}{\partial \boldsymbol{\Sigma}_k}\left(\frac{1}{2}\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\right) \\
&= -\frac{1}{2}\sum_i r_{ik}\boldsymbol{\Sigma}_k^{-1} - \frac{1}{2}\sum_i r_{ik}(-\boldsymbol{\Sigma}_k^{-\top})(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top(\boldsymbol{\Sigma}_k^{-\top}) \\
&\quad \text{since } \frac{\partial \mathbf{a}^\top X^{-1}\mathbf{b}}{\partial X} = -X^{-\top}\mathbf{a}\mathbf{b}^T X^{-\top}.
\end{aligned}
$$

Now we must find the critical point for $\boldsymbol{\Sigma}_k$. Setting the derivative of $l(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ with respect to $\boldsymbol{\Sigma}_k$ to 0 and solving for $\boldsymbol{\Sigma}_k$, we have:

$$\begin{aligned}
0 &= -\frac{1}{2}\sum_i r_{ik}\boldsymbol{\Sigma}_k^{-1} - \frac{1}{2}\sum_i r_{ik}(-\boldsymbol{\Sigma}_k^{-\top})(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top(\boldsymbol{\Sigma}_k^{-\top}) \\
-\frac{1}{2}\sum_i r_{ik}\boldsymbol{\Sigma}_k^{-1} &= \frac{1}{2}\sum_i r_{ik}(-\boldsymbol{\Sigma}_k^{-\top})(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top(\boldsymbol{\Sigma}_k^{-\top}) \\
r_k\boldsymbol{\Sigma}_k^{-1} &= \sum_i r_{ik}(\boldsymbol{\Sigma}_k^{-\top})(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top(\boldsymbol{\Sigma}_k^{-\top}) \\
r_k\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Sigma}_k &= \sum_i r_{ik}\boldsymbol{\Sigma}_k^{-\top}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top\boldsymbol{\Sigma}_k^{-\top}\boldsymbol{\Sigma}_k \\
r_k &= \sum_i r_{ik}\boldsymbol{\Sigma}_k^{-\top}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top \\
\boldsymbol{\Sigma}_k r_k &= \sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top, \\
&\text{since } \boldsymbol{\Sigma}_k \text{ is a linear operator, so} \\
\boldsymbol{\Sigma}_k &= \frac{1}{r_k}\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top.
\end{aligned}$$

We see that this yields the value of $\boldsymbol{\Sigma}_k$ presented in Murphy, Equation 11.32. ■

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

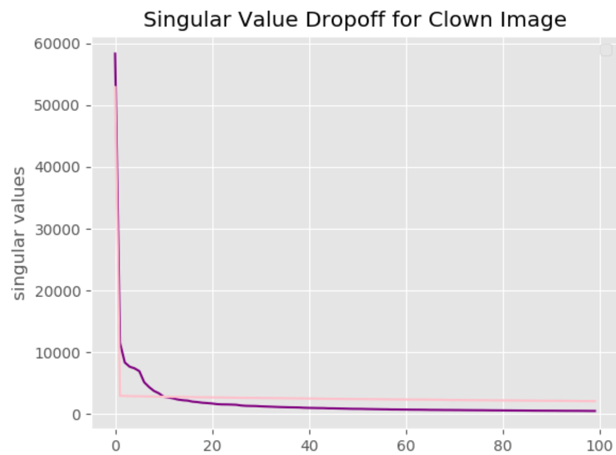

Figure 1: The purple line corresponds to the original image. The pink line corresponds to the shuffled image.