Math 2LA3 Assignment 6

1. Given the following dataset, compute its sample covariance matrix. Determine the variance of the first variable and the covariance between the second and third variables.

$$\begin{bmatrix}2\\3\\3\end{bmatrix}, \begin{bmatrix}4\\5\\1\end{bmatrix}, \begin{bmatrix}4\\4\\1\end{bmatrix}, \begin{bmatrix}5\\2\\3\end{bmatrix}, \begin{bmatrix}3\\4\\1\end{bmatrix}.$$

2. Let T be the last three digits of your student number. For example, if your student number is 400123456, then $T = 456$.

   Give an example of a covariance matrix whose total variance is T and for which:

   - The variance of the first variable is $N/4$,
   - The variance of all the other variables is nonzero,
   - The second and fourth variables have no correlation and all other pairs of variables have nonzero correlation.

3. Given the following sample covariance matrix, what percentage of the data's total variance is explained by its first principal component?

$$S = \begin{bmatrix}3 & 2 & 1\\2 & 3 & 1\\1 & 1 & 4\end{bmatrix}$$

4. Let S be a sample covariance matrix with orthogonal diagonalization $S = PDP^T$ where

$$P = \begin{bmatrix}\frac{1}{\sqrt{14}} & 0 & \frac{13}{9\sqrt{14}} & \frac{-31}{9}\sqrt{\frac{2}{287}} & -10\sqrt{\frac{2}{287}}\\ 0 & \frac{3}{5} & \frac{-4}{3}\sqrt{\frac{2}{7}} & \frac{-10}{3}\sqrt{\frac{2}{287}} & \frac{-2}{5}\sqrt{\frac{14}{41}}\\ \sqrt{\frac{2}{7}} & 0 & \frac{-4}{9}\sqrt{\frac{2}{7}} & \frac{169}{9\sqrt{575}} & \frac{-5}{\sqrt{574}}\\ 0 & \frac{4}{5} & \sqrt{\frac{2}{7}} & \frac{5}{\sqrt{575}} & \frac{3}{5}\sqrt{\frac{7}{82}}\\ \frac{3}{\sqrt{14}} & 0 & \frac{1}{9\sqrt{14}} & \frac{-46}{9}\sqrt{\frac{2}{287}} & 5\sqrt{\frac{2}{287}}\end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix}83.1 & 0 & 0 & 0 & 0\\ 0 & 83.1 & 0 & 0 & 0\\ 0 & 0 & 42.0 & 0 & 0\\ 0 & 0 & 0 & 25.6 & 0\\ 0 & 0 & 0 & 0 & 5.67\end{bmatrix}.$$

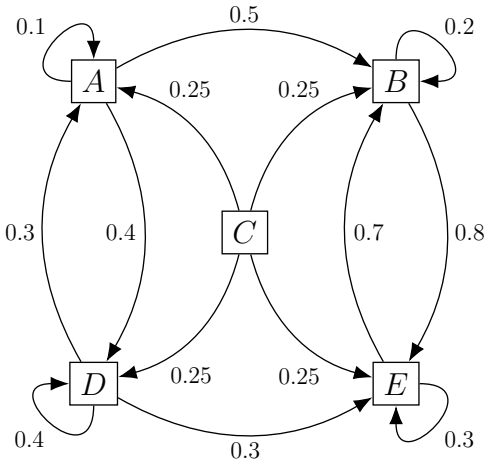
   What is the fewest number of principal components needed that together explain at least 90% of the total variance? What are these components?

5. Find any picture on Wikipedia. Find the dimensions (in pixels) of the image (e.g., upload it to FotoForensics and click "Digest" under Analysis).

   What is the value of k whose best rank k approximation gives a compression ratio of approximately 50% for this image?

   Include your image in your submission (e.g., as one of the pages you upload).

6. Construct the stochastic matrix associated to the following discrete system. Then find an equilibrium vector for the associated Markov chain.



7. Give an example of a Markov chain that in the long run converges to the vector

$$\vec{q} = \begin{bmatrix} 1/2 \\ 3/8 \\ 1/8 \end{bmatrix}.$$

That is, find a stochastic matrix M such that for any initial condition $\vec{x}_0$ and by defining $\vec{x}_t = M\vec{x}_{t-1}$ for all t, we have that

$$\vec{q} = \lim_{t\to\infty} \vec{x}_t = \lim_{t\to\infty} M^t \vec{x}_0.$$

Show your work deriving such a matrix M and justify your work.

8. Let $p = 4/5$. Compute the Google matrix associated to the following network of webpages. Find the PageRank of this network.

You may use technology to compute eigenvalues and eigenvectors. Indicate in your work where you have done so.