This homework is not to be handed in, but you should still do it, as practice for Exam C.

Principal component analysis is covered in Section 7.5 of our textbook. We have discussed the following aspects, some of which are not in that section:

1. sample mean and covariance
2. diagonalizing the sample covariance to get the spreading directions
3. transforming data vectors $\vec{x}$ to vectors $\vec{w}$ in which those spreading directions are prioritized
4. application: visualizing data, including detecting multiple modes
5. application: data compression
6. doing all of this using the SVD instead of diagonalizing

On Exam C, I might ask you about aspects 1, 2, 3, including explicit calculations in which the data are 2D or 3D vectors. I will not ask you to do explicit calculations with higher-dimensional data. I will not ask you to calculate aspects 4, 5, 6 explicitly (although SVD is a separate topic that is covered by the exam).

The following two problems help you practice with aspects 1, 2, 3 above. If you feel that you need more practice, then talk to me or the prefect. Better yet, make up a problem, solve it, and get your study partner to do it too, as a check.

Problem A: Here is a data set of five 2D vectors $\vec{x}_1, \ldots, \vec{x}_5$:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

Compute the first two PCA vectors $\vec{w}_1, \vec{w}_2$.

Problem B: A data set of n 3D vectors has sample mean and covariance

$$\vec{m} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \quad S = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

The first PCA vector is

$$\vec{w}_1 = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}.$$

What was the first data vector $\vec{x}_1$?