Q1: Is the following set of vectors an orthonormal basis of $\mathbb{R}^3$? Explain why or why not.

$$\begin{pmatrix}3\\4\\0\end{pmatrix}, \begin{pmatrix}4\\-3\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}$$

<u>Solution:</u>

$$v_1 = [3\ 4\ 0]^T$$

$$v_2 = [4\ -3\ 0]^T$$

$$v_3 = [0\ 0\ 1]^T$$

The vectors are orthogonal on one another since:

$$\langle v_1, v_2 \rangle = v_1^T v_2 = 3(4) + 4(-3) + 0(0) = 0$$

$$\langle v_1, v_3 \rangle = v_1^T v_3 = 3(0) + 4(0) + 0(1) = 0$$

$$\langle v_2, v_3 \rangle = v_2^T v_3 = 4(0) + -3(0) + 0(1) = 0$$

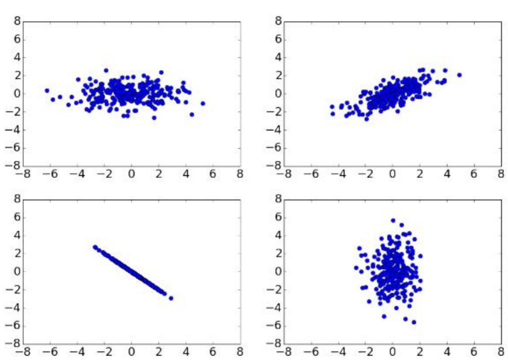
However, they are not normal since the Euclidian norm for the first and second vector is not equal to unity:

$$||v_1||_2 = \sqrt{3(3) + 4(4) + 0(0)} = 5 \neq 1$$

$$||v_2||_2 = \sqrt{4(4) + (-3)(-3) + 0(0)} = 5 \neq 1$$

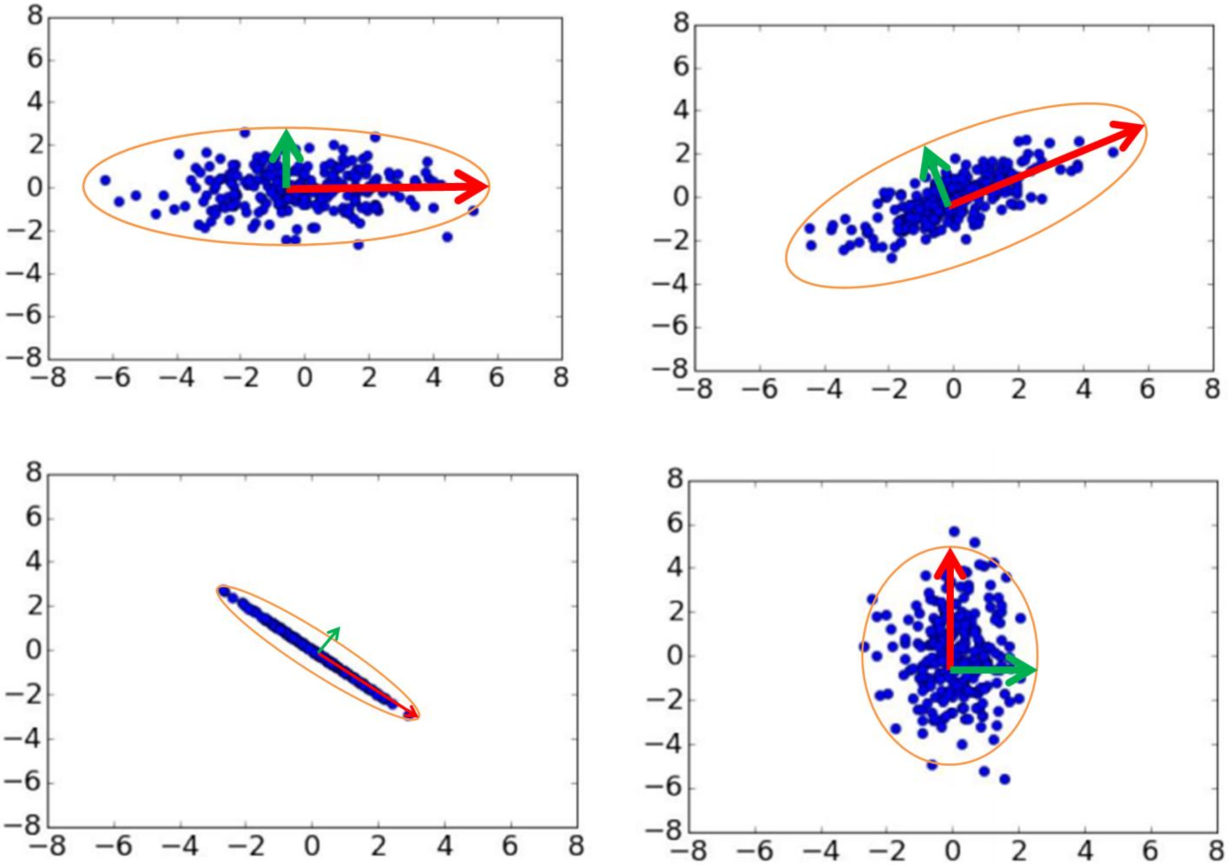
Therefore, the set $\{v_1, v_2, v_3\}$ is not an orthonormal basis set of $\mathbb{R}^3$

Q2: The following four figures show different 2-dimensional data sets. In each case, make a rough sketch of an ellipsoidal contour of the covariance matrix and indicate the directions of the first and the second eigenvectors (mark which is which).



Solution:

In the following plots, the ellipsoidal contour of the covariance matrix is shown in orange-colored contour. Furthermore, the red arrow indicates the first eigenvector (the one associated with the larger eigenvalue and thus variance). On the other hand, the blue arrow indicates the second eigenvector (the one associated with the smaller eigenvalue).



Q3: Let $u_1, u_2 \in \mathbb{R}^p$ be two vectors with $\|u_1\| = \|u_2\| = 1$ and $u_1 . u_2 = 0$. Define

$$U = \begin{pmatrix} \uparrow & \uparrow \\ u_1 & u_2 \\ \downarrow & \downarrow \end{pmatrix}$$

a) What are the dimensions of each of the following?
- U
- U^T
- UU^T
- $u_1 u_1^T$

b) What are the differences, if any, between the following four projections?

- $x \longmapsto (u_1.x, u_2.x)$
- $x \longmapsto (u_1.x)u_1 + (u_2.x)u_2$
- $x \longmapsto U^T x$
- $x \longmapsto UU^T x$

Solution:

As for the projections:

$x \mapsto (u_1.x, u_2.x)$: In this case we only find the magnitude of components of x along the two vectors with no account for the directions.

$x \mapsto (u_1.x)u_1 + (u_2.x)u_2$: In this case projection is performed onto the sub-space spanned by both vectors since they are orthonormal

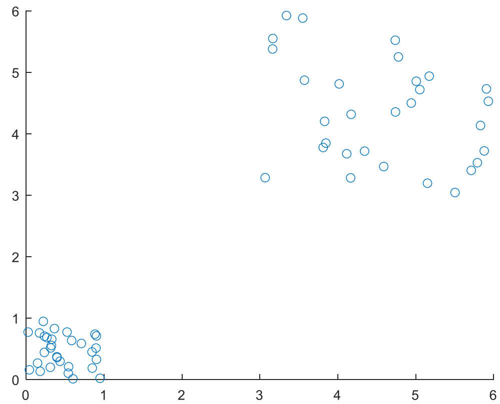
$x \mapsto U^T x$: In this case we only find the magnitudes of the components of x along both vectors as in the first case, $U^T x \equiv\mapsto (u_1.x, u_2.x)^T$

$x \mapsto UU^T x$: This is the same as in the second case, where we perform proper projection with the direction. Both expressions are mathematically equivalent with the latter written in matrix format.

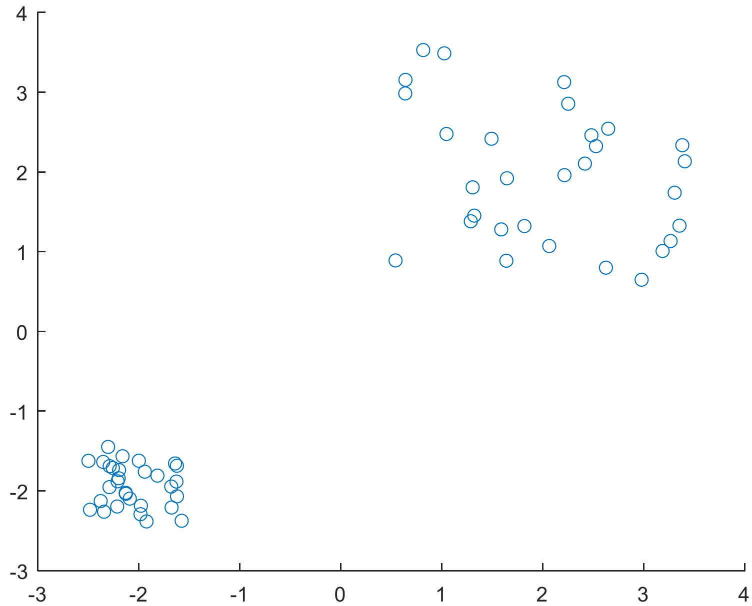
Q4: For the 2-D data file "data2.xlsx". Determine the PCA.

Solution:

In this question, the data has 60 instances with 2 features, visualization of the data on a two-dimensional plane is in the following figure.



In order to find the PCA, we need to find the covariance matrix. First we center the data around zero but subtracting the mean as follows:



The mean vector is $[2.53 \quad 2.40]^T$

The covariance matrix is:

4.71	3.91
3.91	4.21

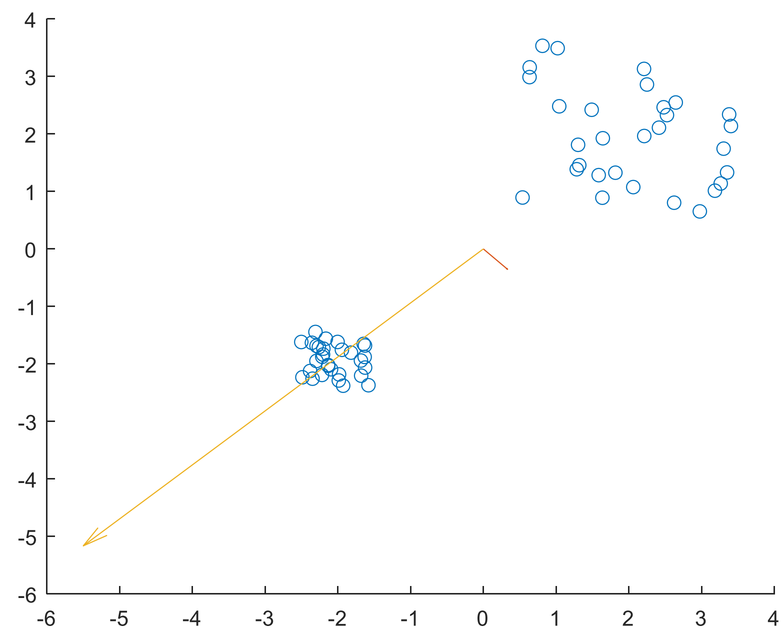
Using Matlab to perform Eigen decomposition for the covariance matrix, we obtain the following eigen-pairs:

$$v_1 = \begin{pmatrix} -0.73 \\ -0.69 \end{pmatrix}, \lambda_1 = 8.38$$

$$v_2 = \begin{pmatrix} 0.69 \\ -0.73 \end{pmatrix}, \lambda_2 = 0.541$$

As it can be seen, both vectors are orthonormal as expected. The first eigen value is much larger than the second one, which indicates that the first principal component of the data can be effectively used.

Visualization of the eigenvectors indicating the principal components is shown in the following figure:



Visualization of the transformed data is given in the following figure;