Homework 2 – Solutions

September 7, 2017

Due: September 14, 2017, 11:59 PM

Instructions

Your homework submission must cite any references used (including articles, books, code, websites, and personal communications). All solutions must be written in your own words, and you must program the algorithms yourself. If you do work with others, you must list the people you worked with. Submit your solutions as a PDF to the E-Learning at UF (http://elearning.ufl.edu/).

Your programs must be written in either MATLAB or Python. The relevant code to the problem should be in the PDF you turn in. If a problem involves programming, then the code should be shown as part of the solution to that problem. If you solve any problems by hand just digitize that page and submit it (make sure the problem is labeled).

If you have any questions address them to:

- Catia Silva (TA) – catiaspsilva@ufl.edu
- Sheng Zou (TA) – shengzou@ufl.edu

Question 1 – 15 points

Suppose $\mathbf{x}$ is a random vector with covariance matrix $\Sigma = \mathbf{x}^T\mathbf{x} = \frac{1}{4}\begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 7 \end{bmatrix}$. Take the *eigendecomposition* of the matrix Σ in order to:

1.1 Find the eigenvalues, λ_1 and λ_2, of Σ. The eigenvalues of Σ can be found by solving:

$$\begin{aligned} \det(\Sigma - \lambda\mathbf{I}) &= 0 \\ \det\left(\frac{1}{4}\begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 7 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) &= 0 \\ \det\left(\begin{bmatrix} \frac{5}{4} - \lambda & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{7}{4} - \lambda \end{bmatrix}\right) &= 0 \\ (\frac{5}{4} - \lambda)(\frac{7}{4} - \lambda) - \frac{3}{16} &= 0 \\ \lambda^2 - \frac{12}{4}\lambda + \frac{32}{16} &= 0 \\ 16\lambda^2 - 48\lambda + 32 &= 0 \\ \lambda = 1 \vee \lambda &= 2 \end{aligned} \tag{1}$$

The eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = 1$.

1.2 Verify that $\mathbf{v}_1 = \begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix}$ are eigenvectors of Σ. The equation to find the eigenvectors is:

$$\Sigma\mathbf{v} = \lambda\mathbf{v} \tag{2}$$

$$\Sigma\mathbf{v}_1 = \frac{1}{4}\begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 7 \end{bmatrix}\begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix} = \begin{bmatrix} \frac{5}{4} + \frac{3}{4} \\ \frac{\sqrt{3}}{4} + \frac{7\sqrt{3}}{4} \end{bmatrix} = \begin{bmatrix} 2 \\ 2\sqrt{3} \end{bmatrix} = 2\begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix} = \lambda_1\begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}$$

$$\Sigma\mathbf{v}_2 = \frac{1}{4}\begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 7 \end{bmatrix}\begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{5\sqrt{3}}{4} - \frac{\sqrt{3}}{4} \\ \frac{3}{4} - \frac{7}{4} \end{bmatrix} = \begin{bmatrix} -\sqrt{3} \\ -1 \end{bmatrix} = 1\begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix} = \lambda_2\begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix}$$

This proves that $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of $\Sigma = R_x$. Moreover, the eigenvector $\mathbf{v}_1$ is associated with the eigenvalue $\lambda_1 = 2$, and the eigenvector $\mathbf{v}_2$ is associated with the eigenvalue $\lambda_2 = 1$.

Note that Matlab provides *normalized eigenvectors*. These are called the *orthonormal* coordinate system and are given as:

$$\begin{aligned} V_1 &= \frac{\begin{bmatrix} 1 & \sqrt{3} \end{bmatrix}^T}{\|v_1\|} \\ &= \frac{\begin{bmatrix} 1 & \sqrt{3} \end{bmatrix}^T}{\sqrt{(1)^2 + (\sqrt{3})^2}} \\ &= \frac{\begin{bmatrix} 1 & \sqrt{3} \end{bmatrix}^T}{2} \\ &= \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}^T \end{aligned} \tag{3}$$

and

$$\begin{aligned} V_2 &= \frac{\begin{bmatrix} \sqrt{3} & -1 \end{bmatrix}^T}{\|v_2\|} \\ &= \frac{\begin{bmatrix} \sqrt{3} & -1 \end{bmatrix}^T}{\sqrt{(\sqrt{3})^2 + (-1)^2}} \\ &= \frac{\begin{bmatrix} \sqrt{3} & -1 \end{bmatrix}^T}{2} \\ &= \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}^T \end{aligned} \tag{4}$$

1.3 Find an *orthogonal* matrix $\mathbf{U}$ such that $\Sigma = \mathbf{U}\mathbf{D}\mathbf{U}^T$. *(Hint: use the spectral theorem).* Using the spectral theorem, we can take the matrix $\mathbf{U}$ to be the matrix containing the *orthonormal* eigenvectors its columns:

$$\begin{aligned} \mathbf{U} &= \begin{bmatrix} V_1 & V_2 \end{bmatrix} \\ \mathbf{U} &= \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \\ \mathbf{U} &= \frac{1}{2} \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix} \end{aligned} \tag{5}$$

And the matrix $\mathbf{D}$ is the matrix containing the eigenvalues associated with the eigenvectors in its diagonal elements:

$$\begin{aligned} \mathbf{D} &= \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \\ \mathbf{D} &= \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{6}$$

So, now,

$$\begin{aligned} \Sigma &= \mathbf{U}\mathbf{D}\mathbf{U}^T \\ \Sigma &= \frac{1}{4} \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}^T \end{aligned} \tag{7}$$

1.4 Suppose $\mathbf{y}$ is a random vector obtained by the principal component transform of $\mathbf{x}$. Write the formula for computing $\mathbf{y}$ from $\mathbf{x}$. The principal component transform of $\mathbf{x}$ is

$$\begin{aligned}\mathbf{y} &= \mathbf{x}\mathbf{U} \\ \mathbf{y} &= \mathbf{x}\frac{1}{2}\begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}\end{aligned} \tag{8}$$

where $\mathbf{U}$ is the matrix containing the orthonormal eigenvectors, and $\mathbf{x}$ is our N-sample 2-dimensional data ($\mathbf{x}$ is $N \times 2$ and $\mathbf{y}$ is $N \times 2$, where N is the number of samples).

1.5 What is the covariance matrix of $\mathbf{y}$? The covariance matrix of $\mathbf{y}$ is given as:

$$\begin{aligned}\Sigma_y &= \mathbf{y}^T\mathbf{y} \\ &= (\mathbf{x}\mathbf{U})^T(\mathbf{x}\mathbf{U}) \\ &= \left(\mathbf{U}^T\mathbf{x}^T\right)(\mathbf{x}\mathbf{U}) \\ &= \mathbf{U}^T\left(\mathbf{x}^T\mathbf{x}\right)\mathbf{U} \\ &= \mathbf{U}^T\Sigma_x\mathbf{U} \text{ from prob. 1.3, we know that } \Sigma_x = \mathbf{U}\mathbf{D}\mathbf{U}^T \\ &= \mathbf{U}^T\left(\mathbf{U}\mathbf{D}\mathbf{U}^T\right)\mathbf{U} \\ &= \left(\mathbf{U}^T\mathbf{U}\right)\mathbf{D}\left(\mathbf{U}^T\mathbf{U}\right), \ \mathbf{U} \text{ contains orthonormal vectors (linear independent), so } \mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I} \\ &= \mathbf{D}, \text{ where } \mathbf{D} \text{ is the matrix whose diagonal elements are the eigenvalues of } \Sigma_x \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\end{aligned} \tag{9}$$

The principal component transformation, *uncorrelates* the data in x by finding the best *rotation* matrix $\mathbf{U}$. You can see that in the covariance matrix of $\mathbf{y}$, the elements outside the diagonal are zero. This means that the two dimensions (or features) are (at least) uncorrelated.

1.6 Plot the new coordinate system on the following graph and plot the curve describing all points with a Mahalanobis distance of 1 from the origin.

The Mahalanobis distance is given as:

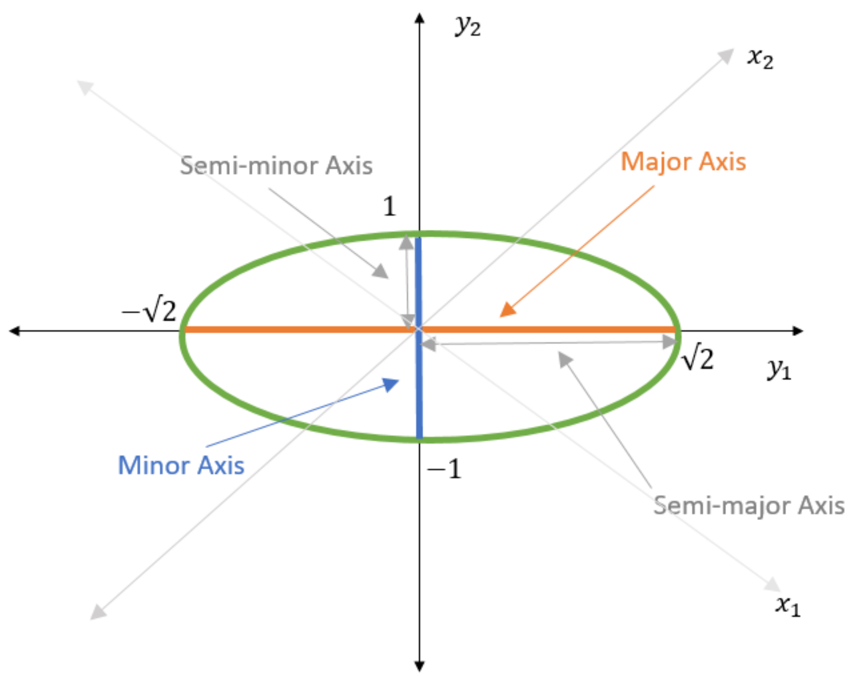
$$D_M(\overrightarrow{y}) = \sqrt{(\overrightarrow{y} - \overrightarrow{\mu})S^{-1}(\overrightarrow{y} - \overrightarrow{\mu})} \tag{10}$$

where $\overrightarrow{\mu} = 0$ and $S = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ is the covariance matrix of $\mathbf{y}$. The inverse matrix of S is given as: $S^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}$.

Thus,

$$
\begin{aligned}
D_M(\overrightarrow{y}) &= 1 \\
\sqrt{\begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}} &= 1 \text{, squaring both sides} \\
\begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} &= 1 \\
\frac{1}{2} y_1^2 + y_2^2 &= 1 \\
\left(\frac{y_1 - 0}{\sqrt{2}} \right)^2 + \left(\frac{y_2 - 0}{1} \right)^2 &= 1
\end{aligned}
\tag{11}
$$

Which is the equation of an ellipse, centered at $(0, 0)$ with semi-major axis length of $\sqrt{2}$ and semi-major length of 1.



Question 2 – 10 points

Compute a formula for $\mathbf{A}^k$, i.e., the matrix power operation defined as the matrix product of k copies of A, using *eigendecomposition*. Show the necessary steps of your derivation.

Let A be an $N \times N$ matrix. Consider $\mathbf{V}$ to be the matrix containing the *orthonormal* eigenvectors of $\mathbf{A}$ in its columns. And the matrix $\mathbf{D}$ to the the matrix whose diagonal elements contain the eigenvalues of matrix $\mathbf{A}$. Now we can write:

$$\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$$
$$\mathbf{A}^2 = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}\mathbf{V}\mathbf{D}\mathbf{V}^{-1} = \mathbf{V}\mathbf{D}\mathbf{I}\mathbf{D}\mathbf{V}^{-1} = \mathbf{V}\mathbf{D}^2\mathbf{V}^{-1},\ \mathbf{V}\mathbf{V}^{-1} = \mathbf{I} \text{ because } \mathbf{V} \text{ contains linearly independent columns}$$
$$\mathbf{A}^3 = \mathbf{A}^2\mathbf{A} = \mathbf{V}\mathbf{D}^2\mathbf{V}^{-1}\mathbf{V}\mathbf{D}\mathbf{V}^{-1} = \mathbf{V}\mathbf{D}^2\mathbf{I}\mathbf{D}\mathbf{V}^{-1} = \mathbf{V}\mathbf{D}^3\mathbf{V}^{-1}$$

By induction

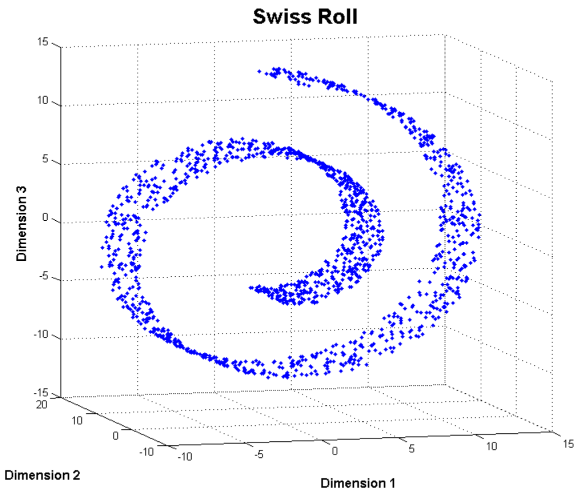
$$\mathbf{A}^k = \mathbf{V}\mathbf{D}^k\mathbf{V}^{-1} \tag{12}$$

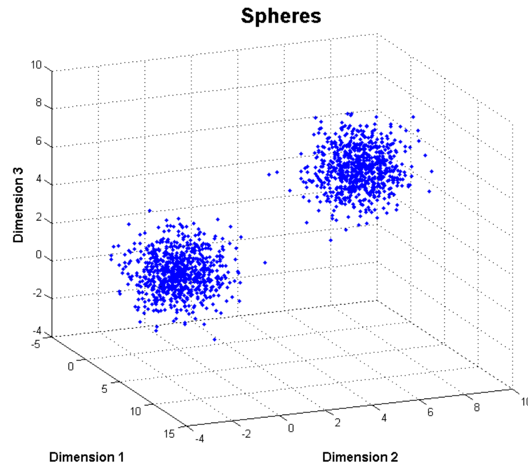
Note: $V^{-1} = V^T$ if and only if A is a symmetric matrix. That would be the case for a covariance matrix.

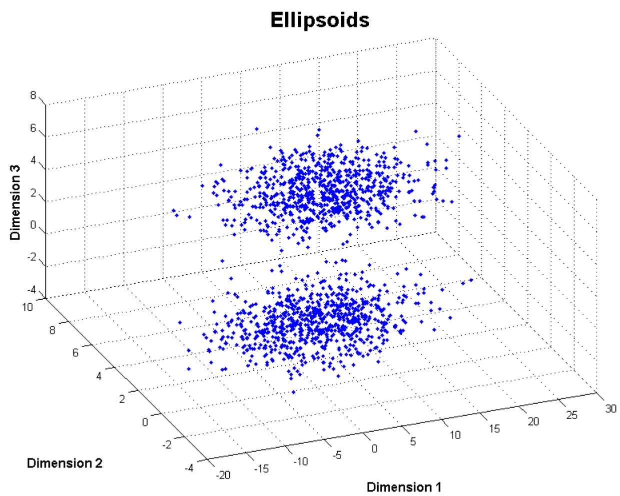
Question 3 – 15 points

For this question, please download all 3 datasets provided along with this assignment: *swissroll.txt*, *spheres.txt*, and *ellipsoid.txt*. Import these files into your programming software.
All datasets have 1500 samples and 3 features/dimensions. Now, consider $\mathbf{X}$ to be a data set (so $\mathbf{X}$ is of size 1500×3). For example, in MATLAB, you can use the following script to generate a plot of the dataset $\mathbf{X}$:

```
figure,plot3(X(:,1),X(:,2),X(:,3),'.');grid on;
xlabel('Dimension 1','FontSize',12,'FontWeight','bold');
ylabel('Dimension 2','FontSize',12,'FontWeight','bold');
zlabel('Dimension 3','FontSize',12,'FontWeight','bold');
title('Swiss Roll','FontSize',20,'FontWeight','bold');
```







For each dataset:

3.1 Find the covariance matrix.

Spheres: The covariance matrix of the "spheres" data set is

$$R_{spheres} = \begin{bmatrix} 10.0368 & 8.9743 & 9.0941 \\ 8.9743 & 9.9246 & 9.0678 \\ 9.0941 & 9.0678 & 10.2626 \end{bmatrix} \tag{13}$$

Ellipsoids: The covariance matrix of the "ellipsoids" data set is

$$r_{ellipsoids} = \begin{bmatrix} 57.8661 & 14.6934 & 7.3664 \\ 14.6934 & 9.8630 & 4.5264 \\ 7.3664 & 4.5264 & 3.3570 \end{bmatrix} \tag{14}$$

Swiss Roll: The covariance matrix of the "swiss roll" data set is

$$r_{swissroll} = \begin{bmatrix} 43.2882 & 0.1535 & 4.4555 \\ 0.1535 & 10.5800 & 0.1544 \\ 4.4555 & 0.1544 & 47.1548 \end{bmatrix} \quad (15)$$

3.2 Find the eigenvectors and eigenvalues of the covariance matrix.

Spheres: The eigenvalues and orthonormal eigenvectors of the covariance of the "spheres" data set are:

$$\begin{aligned} \lambda_1 = 28.1668,\ v_1 &= \begin{bmatrix} 0.5760 \\ 0.5731 \\ 0.5829 \end{bmatrix} \\ \lambda_2 = 1.0571,\ v_2 &= \begin{bmatrix} 0.6207 \\ 0.1572 \\ -0.7681 \end{bmatrix} \\ \lambda_3 = 1.0001,\ v_3 &= \begin{bmatrix} 0.5318 \\ -0.8043 \\ 0.2652 \end{bmatrix} \end{aligned} \quad (16)$$

Ellipsoids: The eigenvalues and orthonormal eigenvectors of the covariance of the "ellipsoids" data set are:

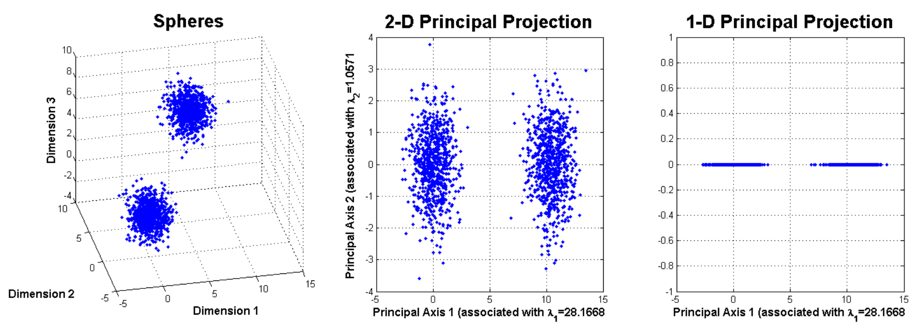
$$\begin{aligned} \lambda_1 = 63.1653,\ v_1 &= \begin{bmatrix} -0.9518 \\ -0.2741 \\ -0.1380 \end{bmatrix} \\ \lambda_2 = 6.8856,\ v_2 &= \begin{bmatrix} -0.3068 \\ 0.8433 \\ 0.4413 \end{bmatrix} \\ \lambda_3 = 1.0352,\ v_3 &= \begin{bmatrix} 0.0046 \\ -0.4623 \\ 0.8867 \end{bmatrix} \end{aligned} \quad (17)$$

Swiss roll: The eigenvalues and orthonormal eigenvectors of the covariance of the "swiss roll" data set are:

$$
\lambda_1 = 50.0795,\ v_1 = \begin{bmatrix} 0.5486 \\ 0.0054 \\ 0.8361 \end{bmatrix}
$$
$$
\lambda_2 = 40.3647,\ v_2 = \begin{bmatrix} 0.8361 \\ 0.0015 \\ -0.5486 \end{bmatrix} \tag{18}
$$
$$
\lambda_3 = 10.5788,\ v_3 = \begin{bmatrix} 0.0042 \\ -1.0000 \\ 0.0037 \end{bmatrix}
$$

3.3 Find (and plot) the projection of the data points into the 2-D and 1-D principal components. After projecting the data into 2-D and 1-D, provide a short discussion (2-3 sentences) of the results for each data set that answers the following question: Does the projection preserve the "important" or "most informative" structure of the original data? Why or why not?

Discussion: Principal Components (PC) are orthogonal directions that capture most of the variance in the data. In PCA, lower-dimensional projections (1) are linear, and (2) only preserve the Euclidean distances between sample points. However, some data sets may contain non-linear structures that linear projections cannot preserve.



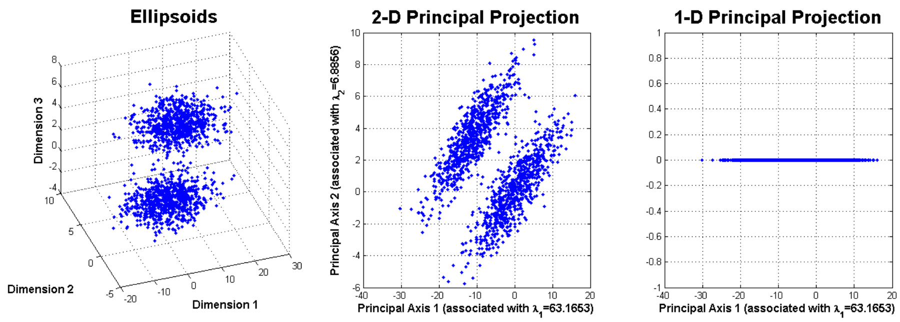
Spheres: Based on the eigenvalues of the covariance matrix, we can see that the 2-D projection preserves approximately

$$
\frac{28.1668 + 1.0571}{28.1668 + 1.0571 + 1.0001} \approx 0.9669 \Rightarrow 96.69\%
$$

of the variance of the original data. Similarly, the 1-D projection preserves approximately

$$
\frac{28.1668}{28.1668 + 1.0571 + 1.0001} \approx 0.9319 \Rightarrow 93.19\%
$$

of the variance of the data. Visually, we see that the 2-D and 1-D projections preserve a lot of the original variance of the data while keeping the separation of the two spherical clusters; therefore, one can conclude that the 2-D and 1-D projections preserve the most informative structure of the data.



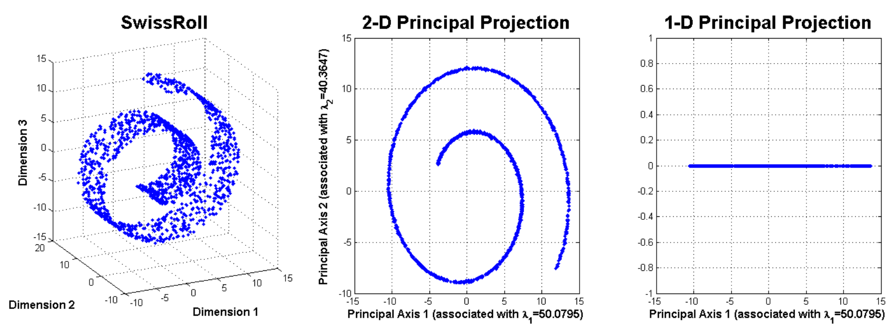
Ellipsoids: Based on the eigenvalues of the covariance matrix, we can see that the 2-D projection preserves approximately

$$\frac{63.1653 + 6.8856}{63.1653 + 6.8856 + 1.0352} \approx 0.9854 \Rightarrow 98.54\%$$

of the variance of the original data. Similarly, the 1-D projection preserves approximately

$$\frac{63.1653}{63.1653 + 6.8856 + 1.0352} \approx 0.8886 \Rightarrow 88.86\%$$

of the variance of the data. Both the 2-D and 1-D projections preserve a good amount of the original variance; however, it should be noted that in the 1-D projection, one can no longer distinguish between the two ellipsoids because the data points are all clumped together. In the 2-D representation, one can still see a separation between the two ellipsoids (thereby, seeing some structure that is not preserved in the 1-D case).



Swiss roll: Based on the eigenvalues of the covariance matrix, we can see that the 2-D projection preserves approximately

$$\frac{50.0795 + 40.3647}{50.0795 + 40.3647 + 10.5788} \approx 0.8953 \Rightarrow 89.53\%$$

of the variance of the original data. Similarly, the 1-D projection preserves approximately

$$\frac{50.0795}{50.0795 + 40.3647 + 10.5788} \approx 0.4957 \Rightarrow 49.57\%$$

of the variance of the data. We see that the 1-D projection does not preserve the original variance of the data. At the same time, it fails to preserve the original underlying shape of the data because it condenses everything roughly into one connected line; the 2-D projection is better at preserving the variance and the most important structure.

See the attached file named "*HW2_problem3.m*" to find an example Matlab code to solve this problem.