



Systems and Biomedical Engineering  
SBE 4032 - Advanced Topics in Medical Informatics (2)  
Spring 2024



### Assignment 1

For this assignment, you're required to use **Python** to apply the following requirements.

#### Requirement 1:

You're required to perform data **sphering/whitening** using the steps discussed in course.

Please apply the following:

- 1- Generate a 2-D normally distributed random dataset with mean  $[2, 2]$  and covariance matrix  $\begin{bmatrix} 10 & 5 \\ 5 & 10 \end{bmatrix}$ , containing 100 samples.
- 2- Create a function that applies data sphering using the steps discussed in the course (you can't use a library function to directly perform the sphering).
- 3- Plot the data before and after the sphering transformation and comment on the results.
- 4- Verify that the transformed data has a 2-D mean of 0 and an identity covariance matrix.

#### Requirement 2:

You're required to perform the PCA technique on the Iris dataset.

Please apply the following:

- 1- Create a function that performs PCA using the steps discussed in the course (you can't use a library function to directly perform PCA).
- 2- Use your function to apply PCA on the Iris dataset. The function should return the sorted eigenvectors, their corresponding eigenvalues and the transformed dataset.
- 3- Verify that the eigenvectors are orthonormal.

- 4- In one figure, plot all the principal components against each other (PC1 vs PC2, PC1 vs PC3, ... . You will have 4 principal components, so a 4x4-plot figure). Each class should be represented with a unique color on the plots.  
Comment on the plots.  
Which 2 PCs look like they separate the classes best? Does it make sense?
- 5- Implement each of the following 3 methods to determine the number of PCs to be selected:  
Cumulative Percentage of Variance Explained - Scree plot - Size of variance.

What number of PCs would you select based on each method?

### Requirement 3:

**Use Python** to solve Exercise 2.1 from the course textbook “Exploratory data analysis using Matlab”.

### Instructions

- 1- The assignment is **individual**. Sharing code is not accepted. Cheating will result in a zero grade.
- 2- Add comments that describe your code.
- 3- Your conclusions and comments on the results can be included in markdown cells within your code.
- 4- **Late submissions** will be penalized.
- 5- **Submission** will be through providing me with a link to your code on Google Colab. The link should be submitted and turned in on Google classroom. Please remember to provide me with access to the code as an Editor.
- 6- Editing your code after the submission deadline is not accepted.