

Instructor Yusuf Yaslan
Assistant Kiymet Kaya

- **Please do your homework on your own.** You are encouraged to discuss the questions with your classmates, but the code and the homework you submitted must be your own work. Cheating is highly discouraged as it could mean a zero or negative grade from the homework.
- Submissions are expected to include a **pdf file prepared with Latex** that contains the solutions and a Jupyter Notebook file with Python for the coding-related questions.

Please read the instructions below for coding-related questions

- Install a **Conda** environment if you do not have it already.
- Install **Jupyter Notebook**.
- Questions should be done in Jupyter Notebook.
- Do not forget to format your code and leave comments for non-trivial sections.

Make sure that you read Chapter 6-7 from E. Alpaydin, Introduction to Machine Learning book, and our lecture slides.

You will be using the **PenDigits** and **FashionMnist** datasets for all the questions. Only use pendigits.tra and fashion-mnist-train.csv files from the datasets.

Question 1

Implement Principal Component Analysis(PCA) to project the digits into TWO and THREE dimensional space. Plot the transformed data points on a scatter plot. Use a unique color for each class. Evaluate how many dimensions are better at representing the data through PCA.

Question 2

Implement Linear Discriminant Analysis(LDA) and project the data points into 2D and 3D space. This time, use the class labels given within the dataset. Plot the transformed data points on a scatter plot using a unique color for each class. Compare and interpret the transformations between LDA and PCA.

Question 3

Implement k-Means Clustering. Run the algorithm for the data given within the homework files. Plot the data points on a scatter plot using a unique color for each cluster. Repeat the procedure for $K \in 2, 3, 4$. Which one seems to form better clusters?