

HOMEWORK 2

Instructions:

There are two **dimensionality reduction** questions in this assignment. Both **feature extraction** and **feature selection** are used for dimensionality reduction which is key to reducing machine learning (ML) model complexity and overfitting.

- Upload your jupyter solution notebook(s) and dataset(s). Don't forget to put code descriptions (markdown or comments), mention about your references/sources in your notebook.
- Non-coding questions may be answered with power point slides (less than 5 slides)
- Each group will share their final visualization in class via jupyter notebook. Please plan for a demo of up to 8 - 10 minutes. Assignments not presented in the class will not be graded.

1. Feature Extraction

Feature extraction is about extracting/deriving information from the original features set to create a new features subspace. Feature extraction is a part of the dimensionality reduction process, in which, an initial set of the raw data is divided and reduced to more manageable groups.

- a. **List at least 3 feature extraction methods** to process an image and compare them. Explain methods and differences between methods.
- b. Write a code snippet using python to denote **extracted features** from the image(s). You may use any image (or images) you like for this question.

2. Feature Selection

Feature selection is about selecting a subset of features out of the original features in order to reduce ML model complexity, enhance the computational efficiency of the ML models and reduce generalization error introduced due to noise by irrelevant features.

You may use any dataset(s) you wish for this question. You may select your dataset from kaggle, UCI Machine Learning Repository or another publicly available datasets. You can find many others on the Web. Pick something that interests you.

- c. **List at least 3 feature selection methods** to process your dataset and compare them. Explain methods and differences between methods.
- d. Write a code snippet using python to denote all original features and **selected features** from your dataset.