Introduction to Machine Learning and Artificial Intelligence, Summer 2023 (ET1550)

Project 3

In two parts, this project investigates applications for Principal Component Analysis (PCA) and k-Means Clustering algorithms as unsupervised learning methods.

In the first part, the Principal Component Analysis (PCA) algorithm will be utilized for dimensionality reduction of an economic dataset. The purpose of this part is to utilize PCA in order to reduce the dataset dimensions to 2D for visualization. This helps to gain some intuition regarding the relationships between different indicators.

In the second part, the k-Means Clustering algorithm will be employed for clustering on a handwritten digits dataset. The idea is to study the capability of the k-means clustering algorithm in grouping data points based on the distribution of their features.

The Datasets

In the first part, a dataset from the World Bank, an international financial institution is used that contains several economic development indicators for the EU countries from the year 2015.

In the second part, a handwritten digits dataset from the University of California Irvine Machine Learning repository is utilized which has 1797 images with 64 features.

The Code

The project may be implemented in both Google Colab and Jupyter Notebook (please refer to the Project IDE document).

The Report

For the project report, you only need to provide answers to the questions. In addition, include your written codes (from the places that are specified in the code) for each question (if required). The project report should be a PDF file containing your answers, codes and possible outputs from the specified parts.

The deadline for submission is **6 August 2023**.