

Project Name: High dimensionality reduction in data stream for Adaptive Machine Learning

Instructors: Dr. Attila Kiss, Hayder K. Fatlawi

Pre-request knowledge: Fundamentals of Machine learning, Python

Framework: Python (Anaconda, Spyder) including: pandas, numpy, sklearn, skmultiflow.

Summery:

Machine learning is concerned with improving the learning ability of the computer to make (intelligent) decisions and to distinguish different patterns based on the given data [1]. Stream data represents a significant source for a vast amount of data. In Adaptive Machine Learning AML, instead of building a fixed classifier during the learning process, AML continuously modifies the classifier as a response to the change in the distribution of a data stream. Although the reduction of the required time and space due to mining a limited number of instances in a specific time, the high dimensionality still represents a challenge for AML[2]. Principal Component Analysis PCA is one of the most popular methods for reducing the number of attributes that can be used for this task.

Main Objective: Improve the efficiency of an adaptive classifier using PCA without delay the response of the classifier and preserving the accuracy of the classification process.

References:

[1] Han J, Pei J, Tong H. Data mining: concepts and techniques. Morgan kaufmann; 2022 Jul 2.

[2] Joao Gama. Knowledge discovery from data streams. CRC Press, 2010

Useful links:

- <https://www.kaggle.com/datasets>.
- <https://physionet.org>
- <https://archive.ics.uci.edu/datasets>