# T1 - Principal component analysis

After a labwork session, you have 7 days to complete the exercises:

- Write a report of at least 2 pages to describe your work (figure and table), discussion, and analysis of labwork.
- Submit in PDF format to Google Classroom (link in Moodle).
- Source code is optional in the report.

Note:

- You can work in a pair to complete all the assignments.
- Remember to provide YOUR student id and full name in the report.

## 1 Study the dataset

- Choose 2 datasets from UCI Machine Learning Repository with more than 3 dimensions. (http://archive.ics.uci.edu/ml/)
- For each dataset, determine which feature is discrete or continuous? is it quantitative or qualitative? Explain.
- Calculate mean, variance, covariance, correlation of the selected datasets.
- Find the most corelated couple of features of each dataset. Comment on the results.

# 2 PCA

- Apply PCA on the two selected datasets
- How much of the total variation in the data is explained by the first two principal components?
- How well are the individual classes separated in the case we use two principal components? Explain.
- Increase the number of principal components used, analyze and comment on the obtained results.
- Visualize data distribution in 2D.