# Optimization and Data Analytics
# Project

## Abstract
In this Project you will implement and compare the performance of five classification schemes. Implementation should be in MATLAB, C/C++ or Python. Finally, you will write a report following the standard scientific writing style. For the submission of the project, you will need to include code source files and the report in a .zip file. Submission will be through black board

## Data
You can find the following two datasets in the blackboard page of the course:
ORL face data set: a set of 400 (vectorized) 40x30 facial images depicting 40 persons.
MNIST data set: a set of 70k (vectorized) 28x28 pixel images depicting hand-written numbers.

## Experimental setup
For the MNIST data set, use the already provided (60k/10k images) train/test splits. For the ORL data set, randomly split each of the 40 classes in 70% training and 30% test images. In both data sets, use the training data to determine the best values for the hyper-parameters of each method. Then, using the best hyper-parameter values, train the methods on the entire training set and evaluate their performance on the test set.

## Methods
Implement and evaluate the performance of the following methods:

1. Nearest class centroid classifier
2. Nearest sub-class centroid classifier using number of subclasses in the set {2,3,5}
3. Nearest Neighbor classifier
4. Perceptron trained using Backpropagation
5. Perceptron trained using MSE (least squares solution)

Apply the above classifiers using:

- the original data (784D for MNIST and 1200D for ORL)
- 2D data obtained after applying PCA

Visualise the classified 2D data (from PCA) using methods we learnt in lectures, such as 2D scatterplots with varying size, shape, and/or colour of points.

Instructions: In this task, you should implement the functions needed in order to apply each classifier, e.g. a structure similar to the one provided in the following. Then, write the code of the experimental setup in one file that calls the classifier functions. The code should be well-commented.

## Example code structure
```
lbls = nc_classify(Xtrain,Xtest,train_lbls);
lbls = nsc_classify(Xtrain,K,Xtest,train_lbls);
W = train_perceptron_backprop(Xtrain,train_lbls,eta);
lbls = perceptron_bp_classify(W,Xtest);
```

## Report

Write a technical report (pdf format) which includes the following:

- Introduction to the Image Classification problem and description of the basic and more advanced approaches.
- Description of the classification schemes used in your experiments.
- Use at least two different visualization techniques to help communicate your results
- Databases description.
- Experimental protocols and results.
- Discussion, where you will provide (qualitative and quantitative) comparisons between the competing methods.
- Conclusions

References (make sure that you reference properly the methods and packages mentioned in the report).

The report should be written using the IEEE paper format (double column, single space, 10pt font size, up to 8 pages – both Word and Latex templates can be found online). This report should have the standard structure (i.e. Abstract, Introduction, Methods, Results and Conclusion). Make sure that you reference properly the methods and packages mentioned in the report.