Machine Learning - Exercise 8
PCA - Eigenfaces
Goal of the exercise

- The goal of the exercise is to implement the **Eigenfaces** algorithm on a subset of the **Labeled Faces into the Wild** (LFW) dataset.

http://vis-www.cs.umass.edu/lfw

- The original dataset has 13233 images for 5749 people (=classes).
- Here we focus on 5 persons with the most images (1140 in total):
  - Colin Powell
  - Donald Rumsfeld
  - George W Bush
  - Gerhard Schroeder
  - Tony Blair

![Predicted vs. True Images](image)
Classification on raw data

- The images are loaded using scikit-learn and reduced to 62*47 black and white vectors.
- A simple MLP (with ReLU and dropout) is used to classify the data.

1. Read the script Eigenfaces.py and identify its different parts (data loading, PCA, MLP, training, analysis).

2. Run the script. Is the performance satisfying? Is the network overfitting?

3. What do the classification report and the confusion matrix represent?
Classification using PCA

4. Identify the different steps of the PCA algorithm.

5. Switch on PCA by setting `use_pca = True`.


7. Observe the resulting principal components. Why are they distorted?

8. Vary the number of principal components and observe their effect.

9. Modify the MLP (linear, more neurons, more layers) for an optimal performance.