# DAT10 SF: HOMEWORK 4 ASSIGNMENT

**Assigned:** Saturday, November 1
**Due:** Tuesday, November 4th, midnight.

The purpose of this homework is to gain deep, hands-on experience with dimension reduction and clustering techniques.

## DATA & CONTEXT

For this assignment, we will use the Olivetti faces dataset that we saw in the lecture. This is a very well known dataset in the machine learning world and can be obtained here:

http://scikit-learn.org/stable/datasets/olivetti_faces.html#olivetti-faces

## HOMEWORK QUESTIONS

1. Implement PCA on the dataset, using the sklearn. What do the principal components tell? (is scaling necessary?)

2. Change from PCA to RandomizedPCA. Do you obtain the same principal components? Why? Why not?

3. Plot the datapoints in the plane of the two principal components. Are any clusters visible?

4. Implement k-means clustering, using sklearn, with k = number of people in the dataset. Does the clustering work? Does each cluster map correctly to one person?

5. Vary the number k between 1 and <number of datapoints> and plot the Silhouette coefficient as a function of k. What do you observe? What conclusions can you draw?