1 PCA application: eigenfaces

Ability to recover low-dimensional representation of the high dimensional data set is essential in overcoming what is called a curse of dimensionality (in simple terms this means that in order to reason well about the data we need exponentially many data points than the data dimension). To illustrate the point consider the image 256×256 pixels, since each of them is a separate data point our data space is $\mathbb{R}^{256^2}$. Luckily points are correlated with each other and most often lay on low dimensional space.

One of examples of such low-dimensional data is human face. PCA was one of the first tools used for face recognition. Idea is simple: create a correlation matrix from the training dataset and find the best projection of the new training image onto the eigenvectors of the correlation matrix. Since eigenvectors represent images of the face they were named *eigenfaces*. In this lab you will learn how to apply reconstruct faces via PCA.

1.1 Assigment [1]

1) Take a look at lines $1 - 23$ in the `test.m`. Those lines in the code do the following: for every image read it in from the directory `data/pict`, resize it to be 100×100 image, reshape it as a column vector of size $\mathbb{R}^{10000 \times 1}$ after that resulting column is added to the data matrix `data`. You task is to normalize this data matrix to have zero mean and unit variance. You will need to implement function `normalizeData.m`. Let D denote our data matrix we have $D \in \mathbb{R}^{n \times m}$, where $n = 10000, m$- number of images, at first you calculate mean face , which is given by:

$$meanFace_i = \frac{1}{m}\sum_{j=1}^{m} D_{ij} \quad \forall i = 1, ..., n \tag{1}$$

Then mean face should be subtracted from each column of the data matrix. To make our data have a unit variance the following transformation should be made:

$$\sigma_j^2 = \frac{1}{m}\sum_{i=1}^{n} D_{ij}^2$$
$$D_{ij} = D_{ij}/\sigma_j \quad \forall j = 1, ..., m \quad i = 1, ..., n \tag{2}$$

After normalization you should be able to run lines $26 - 40$.

2) Implement `myPCA.m` without using built-in MATLAB functions `pca,princom,cov`. Implementation details are given in the file.

3) Plot different eigenvectors. To do that eigenvectors should be normalized to be in range $0 - 255$. Reshape the column vector to size 100×100.

4) How many eigenvectors explain 60% variance? 80%? 95% ?

5) Implement the function `projectPCA.m` which should project new image onto the data matrix. New image should be projected as a linear combination of the

[1] Materials for this assignment can be downloaded from here

eigenvectors. Project test image , B (lines $66 - 70$). What is the person's name on the test picture? Reproject test image using different using 10, 30, 60 first eigenvectors.