Artificial Neural Networks: Exercise session 3

Stijn Staring (r0620003)

May 21, 2021

1 Principal Component Analysis

Principal component analysis is a method for dimensionality reduction. A principal component analysis uses the eigenvectors of the covariance matrix of the original data as principal components to conduct a linear transformation of the original data to a lower dimension. The eigenvectors that correspond to the largest eigenvalues are considered the most important principal components that determine data appearance the most.

$$\boldsymbol{z} = \boldsymbol{E}^T \boldsymbol{x}. \tag{1}$$

In Eq. 1, $\boldsymbol{x}$ is an original data point with dimensions $\mathbb{R}^p$ and $\boldsymbol{z}$ is the corresponding data point with reduced dimensions $\mathbb{R}^q$. $\boldsymbol{E}^T$ is the used transformation matrix with dimensions $q \times p$ and with the q largest eigenvectors of the covariance matrix as rows. The new dimensions of the data point equals the amount of chosen eigenvectors that span the new subspace. If all eigenvalues are distinct, $\boldsymbol{E}$ is orthogonal and the original data point can be retrieved by multiplying 1 by $\boldsymbol{E}$ because of the orthogonal property: $\boldsymbol{E} \cdot \boldsymbol{E}^T = \boldsymbol{I}$.

The first task performed is the comparison of the dimensionality reduction of random data in comparison to highly correlated data. It is found that there are less eigenvectors needed when the same RMSE is considered. The second task is to perform PCA on handwritten images of the digit 3 taken from the US Postal Service database. In this dataset every image consists out of 16×16 pixels which is represented by an array of 256 floating numbers. The dataset consists out of a total of 500 images of the digit 3. When the mean array is calculated, also the mean digit 3 is obtained which is displayed in Figure 2a. The covariance matrix is calculated for the data set and the eigenvalues derived are shown by Figure 2b. The influence of the amount of eigenvectors is demonstrated in Figure 1c. When only one eigenvector is used, the reduced images are very similar. This means that the eigenvector with the largest eigenvalue focusses on the intrinsic shape of a 3 and in what distinguishes it from an other number e.g. a five. when more eigenvectors are used and the dimensionality of the spanned subspace increases, different variations in the digit three becomes clear.

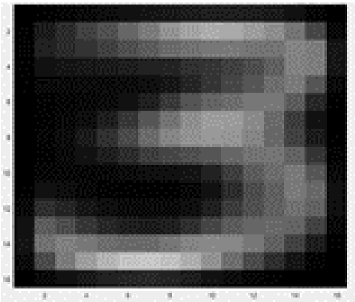
- Logical that when more eigenvectors are included the representation error decreases because the dimensionality is less reduced. However when the derivative of the function is assessed, it can be seen that in the beginning the function decreases much faster than at the end. This means that by a small amount of eigenvectors already a good representation is possible of the original data.

- the quality of the contribution of an eigenvector corresponds to the size of the eigenvalue. It can be seen in the figure that the quality of the the eigenvectors rapidly decreases. This again stresses the fact that already good representations of the digit 3 are possible with a small amount of eigenvectors.

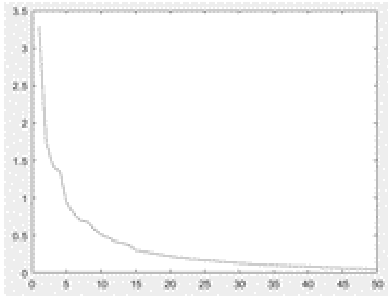
- reconstruction error should be machine precision

- The quality of the reduction depends on how close the sum of the largest q eigenvalues is to the sum of all p eigenvalues –¿ this means that the eigenvectors corresponding to large eigenvalues contribute more to the quality –¿ more fundamental principal component.

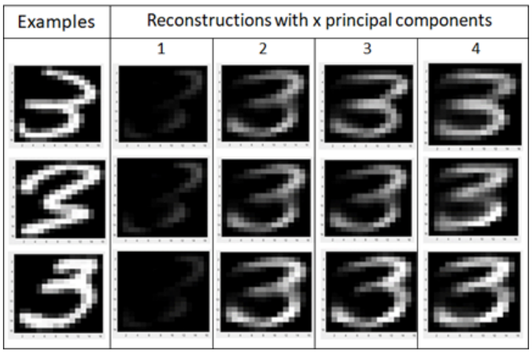
- Reason that E transpose is same as inverse is because the matrix consisting of eigenvectors is orthogonal matrix –¿ E*ET gives I.



(a) Mean image of digit 3

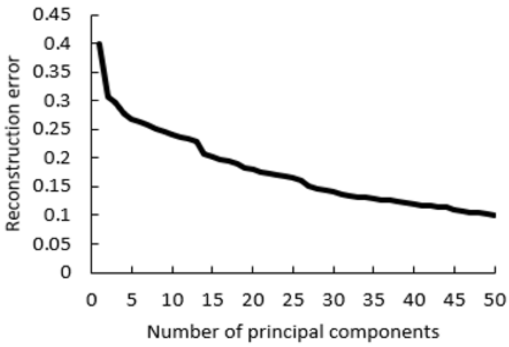


(b) Eigenvalues

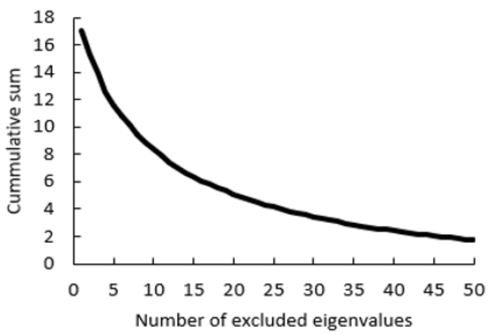


(c) Increasing amount of eigenvectors

Figure 1: Results of applying PCA on the US Postal Service database.



(a)



(b)

Figure 2: The influence of the amount of eigenvectors used.