Artificial neural networks

Assigment 4: PCA and SOM

Lood, Cédric
Master of Bioinformatics

June 18, 2016

1 Context

In this exercise, I explored techniques of unsupervised learning using neural networks. Specifically Principal Component Analysis (PCA) and Self-Organizing Maps (SOM).
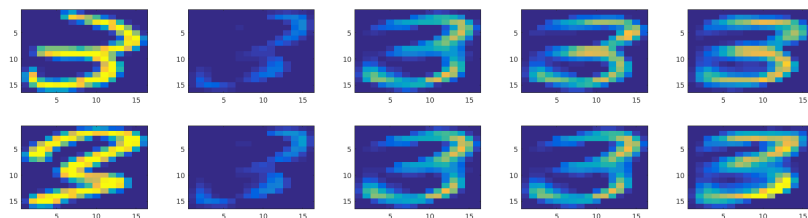
2 Principal component analysis

The idea behind PCA is to reduce the dimensionality of the data in the input space with the hope that the lower dimensional space captures most of the structure of the data. This is done by projecting the data onto the eigenvectors of the covariance matrix. Once the lower dimensionality space has been computed, another question one might try to solve is to see whether a correct reconstruction of the datapoints is possible.

Here we will work with a subset of the US postal service database that consists of digitized versions of handwritten digits (16 by 16 pixels). In particular, we'll focus on the digit "3". On the top left-hand side of the figure below, we view the mean "3" over the 500 examples we have. The right-hand side plot shows the magnitude of the eigenvalues. The bottom left-hand side shows the plots of the eigenvectors associated with the highest eigenvalues. As can be seen on the "EV 256" plot, the curvy feature typical of a 3 seems to be associated with the highest eigenvalue.
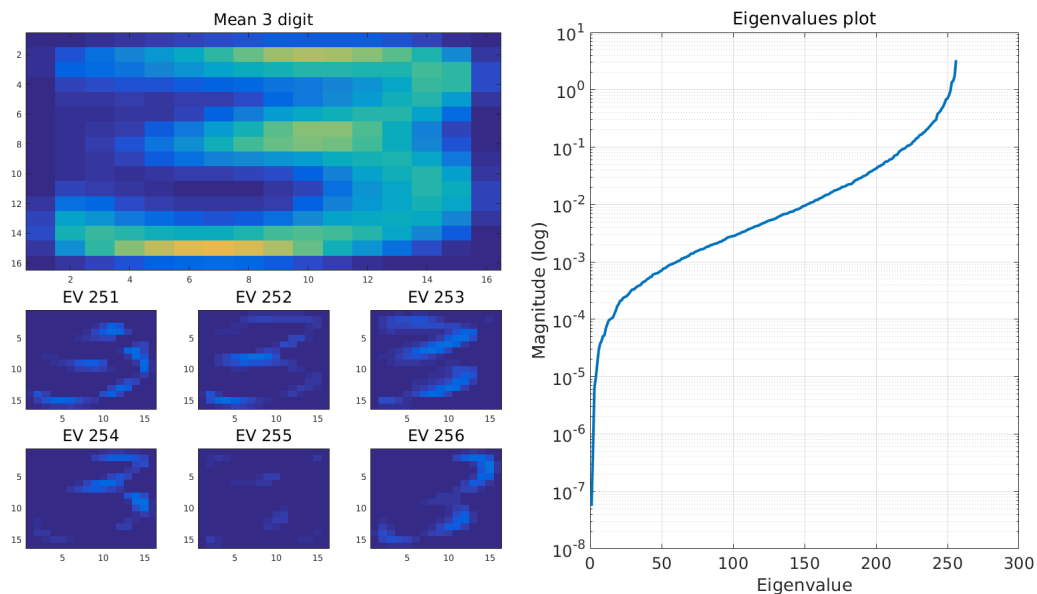
In the next figure, you can observe the reconstruction process for 2 digits from the dataset. The rows start with the digit as it exists in the dataset, then the next are a reconstruction of the digit based on the first 4 principal components. As can be observed, the reconstructed digits are much more similar than the two original digits.

3 SOM