Laboratory 3 Report
Principal component analysis

14-06-2019

Diogo Correia s264324
diogo.valacorreia@studenti.polito.it

1 Theoretical Cheatsheet

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of linearly uncorrelated variables called principal components. [1]

1.0.1 Algorithm

- Mean normalization and scaling
- Reduce data from n-dimensions to k-dimensions

To reduce the data to k-dimensions: compute *covariance matrix* following equation 1 that will be an n by n matrix [2].

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(x^{(i)})(x^{(i)})^T \tag{1}$$

After that, the *eigenvectors* of matrix Σ need to be computed. This can be done via the **eig** command in matlab, like [W, lambda] = eig(sigma).

With the eigenvectors we can calculate the vector of latent variables z_i like in equation 2.

$$z_i = W^T x_i \tag{2}$$

An approximation of x_i, $\hat{x_i}$, can also be calculated like showed in equation 3

$$\hat{x_i} = W z_i \tag{3}$$

2 Implementation & Results

2.1 Exercise 1

In this exercise, we have to employ spectral vectors belonging to two classes in the Indian Pines dataset. I've choose classes 2 and 6.

The objective is to reduce the dimensionality of the spectral vectors of the two classes using PCA.

2.1.1 Results

All the steps described in the practical guide were concluded.

In figure 1 can be seen the z_i representation in 2D for $k = 2$ PCAs.

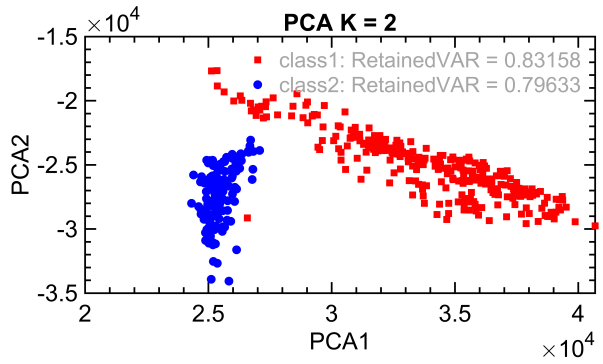

Figure 1:

The computation of the error (mean square error - MSE) can be seen in figure 2. Attention that the MSE was divided by the variance of each class to better see the results.

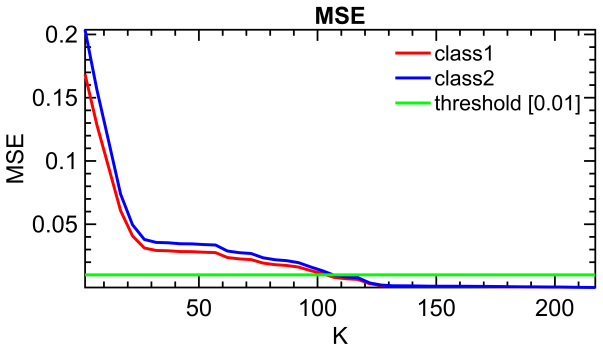

Figure 2:

The plot of the eigenvectors corresponding to the 3 largest eigenvalues can be seen in figure 3.

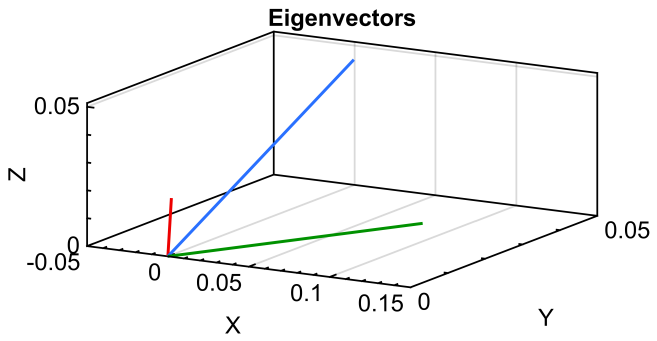

Figure 3:

3 Appendix

References

[1] Principal component analysis. https://en.wikipedia.org/wiki/Principal_component_analysis. Accessed: 2019-05-10.

[2] Enrico Magli. Slides: Statistical learning and neural, ict for smart societies.