Statistical Machine Learning – Week 13

Sascha Jecklin

HS 2022/23

1 Task

This week we will implement Principal Component Analysis (PCA) from scratch. PCA is a linear dimensionality reduction technique. As we will see, there exists also an inverse transformation which maps from the low-dimensional space back into high-dimensional space. However, due to the process of reducing dimension, part of the information is lost. At best, this lost information is irrelevant noise. Thus, the PCA can also be used as a method for noise reduction by eliminating irrelevant dimensions in the transformed space and then transforming the data back into the original space.

- Try to get an overview of the code provided in `week13_exercise.py`. The script contains three functions, each implementing an sample application of PCA. Try to understand them.
  - example_1: This function performs PCA on the `USArrest` data which was introduced in class. It prints and displays the explained variance ratio, which can be calculated from the eigenvalues of the covariance matrix.
  - example_2: This function reproduces Figure 10.1 on page 378 in the textbook. It shows the first two principal components for each State of the USArrest dataset. Moreover, it shows the direction of the predictors `UrbanPop`, `Rape`, `Assault` and `Murder` in this new coordinate system.
  - example_3: This function uses PCA as a noise reduction method. We add some random noise to the images of handwritten digits (MNIST dataset). In a second step we try to get rid of this noise by transforming the images into a new space (spanned by the principal axes), removing the least important dimensions and transforming it back to the original space.
- These functions are given and require no additional implementation. Your task is to complete the implementation of the Python class `MyPca` in the module `my_pca_ex.py`. Complete the functions `fit`, `transform` and `inverse_transform` which calculate the transformation matrix, the projection on the principal axes and the backtransfomation respectively.
- Test your PCA implementation by running the example functions in the module `week13_exercise.py`.

Comments

- Note that there are several ways to implement PCA. The book uses sum and matrix notation.