Machine Learning and Data Mining
Summer 2015
Exercise Sheet 10

Presentation of Solutions to the Exercise Sheet on the 08.07.2015

Exercise 10-1   PCA

a) Please describe what a PCA aims for and under what circumstances it is most helpful.

b) Which possibly negativ consequences might arise when applying PCA to a dataset of unknown structure?

Exercise 10-2   PCA

Consider the $\mathbf{X} \in \mathbb{R}^{M \times N}$ matrix containing six data points $\mathbf{x}_i \in \mathbb{R}^2$. Note that in contrast to the conventional representation, the patterns are held in columns here.

dim 1	1	2	3	5	6	7
dim 2	0	0	0	6	6	6

Conduct a PCA on the given data. Please state the eigenvectors, eigenvalues, covariance matrix and visualize the data before and after the PCA.

Exercise 10-3   Eigenfaces

The term *eigenfaces* describes the eigenvectors of a normalized covariance matrix.

a) Find the eigenfaces of the number dataset from the `numberMatrix.RTable`.

b) How many principal components are required to reconstruct the dataset? Are the eigenfaces sufficient for this purpose? Are all patterns reconstructable with equal quality?

Exercise 10-4   Soccer Ball PCA

The Caltech 101 dataset consists of more than 9000 images which have each been assigned to one out of 102 classes. We consider down-scaled images ((32×32)-thumbnails) taken from the classes `soccer ball` and `faces easy`.

a) Conduct a PCA on the 64 `soccer ball` images. Can the images be reconstrcted losslessly with only part of the principal components?

b) Now consider the `faces easy` dataset, consisting of 435 image. Can this dataset be adequately reconstructed using the principal components from part a)?

c) Now consider the dataset of part a) and the principal components of the dataset of part b). Does reconstructing the soccer balls from the faces' components work?