ANALYTICS FOR INDUSTRIAL INTERNET 2018

EXERCISE 4
Deadline 22.10.2018, submit at https://abacus.abo.fi/ro.nsf/

Part A:
Principal Component Analysis

Principal Component Analysis (PCA) is a methodology to express a dat set using projections P, projecting the orginal data X ,$Y = PX$ which maximize the variance of data (first principal components). The first basis in the projection is freely selectable (representing the major dynamics in the system), the second is orthogonal to the first one, the third one is again orthogonal to the previos ones.
The objetive is to find a projection P that diagnolize the covariance matrix. As the covariance matrix express the dependency of the variables in Y, a diagonal covariance matrix express that there is no dependency between variables, hence the data can be expressed linearly along one basis in the projection.
Mathematically this can be done using Singlar Vale Decomposition (SVD) or using eigenvalues of covariance. Good general overview is found in `http://www.cs.cmu.edu/~elaw/papers/pca.pdf`.

(i) Use the position data from `Position.mat` (from same location as this exercise, use `Position.longitude` and `Position.latitude`. Form a measurement matrix X from these values, and then perform PCA using the code in the paper `http://www.cs.cmu.edu/~elaw/papers/pca.pdf`. The projection P is now the principal component of vehicle movement. Now express the vehicle movement using the principal components. Plot the original measurements and the first principal component in the same graph.

(ii) Also use the altitude data `Position.altitude`, and repeat what is done in the previous task. Comment on the results.

Part B:
Independent Component Analysis

Independent component analysis (ICA) can be used to separate source signals from measurements on mixed signals. The assumptions in ICA is that the source signals are independent, i.e. the values of one signals give no information on the values of the other.
One example of independent signals is the cocktail-party party, where sound signals from different sound sources (people) x_i are mixed in the observed signals (microphones) m_i. If the vectors of source signals is expressed as a matrix X, the vectors of observations is represented in matrix M, and A represent the mixing matrix, the relation can be expressed $M = AX$. In order to find the original source signals, the matrix B, $\hat{X} = BM$ should be estimated. This can be made finding a matrix B that minimize the independence between vectors in $\hat{X}$. The package FastICA performs this minimization.

(i) Download the package FastICA from `https://research.ics.aalto.fi/ica/fastica/code/dlcode.shtml`. Use the matrix M in file `soundmix.mat` as 2 measurements (sample rate 44,1 kHz) and apply FastICA to the measurements. How well is the sound signals separate? What is the estimate of the mixing matrix A?

(ii) The files `w1.dat` and `w2.dat` contain 11025 Hz audio signals $w1(n)$ and $w2(n)$, which consist of two mixtures of a sound signal and a heavy noise signal (such as that from a vacuum cleaner). Use ICA for extracting the sound signal buried in the noise. Verify that the sound signal has been found by listening to it (note the sampling frequency!), and by comparing the result with the original signal, which is given in the file `sound50000to50100.dat` for the range n = 50000 - 50100. Note again, that scaling and possible sign inversion is needed (cf. Remark).