MSPR 3 Principal Component Analysis Exercises (Due: Sunday 27.9.2015 12h pm (noon))

Dr. Hendrik Purwins, Assistant Professor
Jan Stian Banas, TA
Dept. Architecture, Design and Media Technology, Aalborg University Copenhagen
A.C. Meyers Vænge 15, DK-2450 Copenhagen SV, Denmark

1. (Feedback) Please give us feedback on the last lecture and homework: `http://goo.gl/forms/E9kB38QZk5` Thanks!

2. (Correlation) Load the 100 points in `cloud.mat`. Using Matlab, project the 100 points on the XY- plane, the YZ-plane, and the plane spanned by $\begin{bmatrix}1\\1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\\1\end{bmatrix}$. Calculate the

$$\tau = \frac{\sum_{j=1}^{100} \|\mathbf{P}\mathbf{x}_j\|^2}{\sum_{j=1}^{100} \|\mathbf{x}_j\|^2}$$

   for all three projections, as a measure of quality of these projections, i.e. to what extend the projections preserve the inertia of the points. (50 P)

3. Eigenvalue decomposition

   Use MATLAB to find the eigenvectors and eigenvalue decomposition of $\mathbf{A}$.

$$A = \begin{bmatrix} \frac{3}{2} & \frac{-1}{2} \\ \frac{-1}{2} & \frac{3}{2} \end{bmatrix} \tag{1}$$

   (10 P)

4. In Matab, generate a grid of `x` values from -10 ascending in steps of 0.5 to 10. Then, generate a matrix `X` in the following way. Put `x` in the first row of `X`. Using Matlab and arguments `x`, generate the function $f(x) = x$ and write its values into the second row of `X`. Fill rows 3-6 of matrix `X` with the function values for $f(x) = -x$, $f(x) = x^2$, $f(x) = \sin(x)$ and the last row with a random vector of the length of `x`, using `rand` in Matlab. Plot all 5 functions in Matlab. Then calculate the Pearson correlation coefficient for all pairs of functions and display the results in a 6×6 table. Discuss the results. (20 P)

5. Analyze the `adult` dataset. Use features age, education-num, sex, capital-gain, capital-loss, hours-per-week, and income ('>50k','<=50k'). Convert the categorical variables sex and income into a number (0,1), using the Matlab function `strcmp`.

(a) Calculate the correlation matrix between all 7 features. Based on the correlation, which feature of the first 6 features would be best suited to predict the income? (5P)

6. Self Assessment: Check the exercises that you have seriously worked on.

2	3	4	5 a