Datasta Tietoon, Autumn 2007

EXERCISE PROBLEMS 2 [Nov 16th 2007, Nov 21st 2007]

H2 / 1. (Principal component analysis)
We have the following data matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{pmatrix} 2 & 5 & 6 & 7 \\ 1 & 3 & 5 & 7 \end{pmatrix}$$

a) Plot the columns of $\mathbf{X}$ in the (x_1, x_2) - coordinate system
b) Normalize $\mathbf{X}$ to zero mean by subtracting from the columns their mean vector
c) Compute the covariance matrix $\mathbf{C}$ and the eigenvector corresponding to its largest eigenvalue. Plot the eigenvector in the coordinate system of item a). How would you interpret the results according to PCA?

H2 / 2. (Principal component analysis)
Assume that $\mathbf{x}$ is a zero mean random vector and we have a sample $\mathbf{x}(1), ..., \mathbf{x}(n)$. Assume $\mathbf{w}$ is a unit vector (such that $\|\mathbf{w}\| = 1$) and define $y = \mathbf{w}^T\mathbf{x}$. We want to maximize the variance of y given as $E\{y^2\} = E\{(\mathbf{w}^T\mathbf{x})^2\}$. Prove that it will be maximized when $\mathbf{w}$ is the eigenevctor of the matrix $E\{\mathbf{x}\mathbf{x}^T\}$ corresponding to its largest eigenvalue.

H2 / 3. (On-line-principal component analysis)
In the lectures, the so-called SGA algorithm was introduced for computing the eigenvector defined in the previous Exercise 2:

$$\mathbf{w} \leftarrow \mathbf{w} + \gamma[y\mathbf{x} - y^2\mathbf{w}].$$

For simplicity, assume that vector $\mathbf{x}$ is a constant, $y = \mathbf{w}^T\mathbf{x}$ and the learning rate γ is a (small) constant. Where does the vector $\mathbf{w}$ converge in this algorithm?

H2 / 4. (ML-estimation)
Derive the maximum likelihood estimate for the parameter λ of the exponential probability density

$$p(x|\lambda) = \lambda e^{-\lambda}$$

when there is available a sample $x(1), ..., x(n)$ of the variable x.

H2 / 5. (Bayesian estimation)
We are given a sample $x(1), ..., x(n)$ of a variable x known to be normally distributed

$$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

We have good reason to assume that the average value μ is close to zero. Let us code this assumption into a prior density

$$p(\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mu^2}.$$

Derive the Bayes MAP estimate for the value μ and interpret your result when the variance σ^2 changes from a small to a large value.