Datasta Tietoon, Autumn 2008

EXAM 14. 1. 2009 (note: problems in Finnish on the reverse side)

1.
A convolution filter is given using the formula

$$g_k = \sum_{m=-\infty}^{\infty} f_m h_{k-m},$$

where f_m is the (discrete) input signal, h_n is the filter sequence, and g_k is the output signal. Derive the output signal g_k when

$$f_0 = f_1 = 1, \ f_m = 0 \text{ otherwise;} \quad (1)$$
$$h_n = a^n, n \geq 0, \ h_n = 0 \text{ otherwise} \quad (2)$$

where a is a constant, $0 < a < 1$. Plot the output signal when $a = 0.5$.

2.
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.

3.

a) Explain how the c-means clustering algorithm works.

b) We are given the following data matrix:

$$X = \begin{bmatrix} 1 & 1 & 3 & 3 & 6 & 8 \\ 2 & 4 & 2 & 4 & 1 & 1 \end{bmatrix}$$

Perform clustering of its columns into 2 clusters using the c-means algorithm. Initially, set the values of the cluster mean vectors as

$$\mathbf{m}_1 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}; \ \mathbf{m}_2 = \begin{bmatrix} 8 \\ 4 \end{bmatrix}.$$

Show graphically (by picture) how the algorithm now works.

4.
(a) Define the frequent set of 0-1 data. Give an example of a small 0-1 data set and list its frequent sets using some suitable threshold value N.
(b) Describe the principle of the levelwise algorithm for finding frequent sets.

5.
Answer one of the following essay questions that are associated with the Matlab exercise:
A) "Eigenfaces" and the use of eigenvalues for clustering face images
B) k-nearest neighbor classifier.