

# Datasta Tietoon, Autumn 2009

EXAM

13. 1. 2010

(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  $x(1), \dots, x(n)$  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  $\mu$  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  $\mu$  and interpret your result when the variance  $\sigma^2$  changes from a small to a large value.

3.

Assume 5 vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_5$ , whose mutual distances  $d(\mathbf{x}_i, \mathbf{x}_j)$  are given as the following matrix

$$D = \begin{bmatrix} 0 & 4 & 9 & 6 & 5 \\ 4 & 0 & 1 & 8 & 7 \\ 9 & 1 & 0 & 2 & 3 \\ 6 & 8 & 2 & 0 & 1 \\ 5 & 7 & 3 & 1 & 0 \end{bmatrix}$$

Perform hierarchical clustering for the vectors and plot the clustering tree, assuming that the distance between two clusters is equal to the distance between the two closest vectors in the two clusters. What is the best clustering into 3 clusters?

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.