

Datasta Tietoon, Autumn 2013

EXAM, 16.12.2013

(note: problems in Finnish on the reverse side)

1. Assume that d dimensional data vectors are uniformly distributed in a hyperball with radius 1. Let us define as inner points those whose distance from the center point of the hypersphere is at most $1 - \epsilon < 1$. Show that the relative volume of the set of inner points tends to zero as $d \rightarrow \infty$, in other words, in very high dimensions almost all data points are on the surface of the hyperball. (Auxiliary result: The volume of a d -dimensional hyperball with radius r is $V_d(r) = C_d r^d$ where the constant C_d does not depend on the radius r .)

2. Derive the maximum likelihood estimate for the parameter λ of the exponential probability density

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

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

3. Let us consider a 1-dimensional SOM map with three units, whose weights and inputs are scalars on the interval $[0,1]$. The neighbor of unit 1 is 2, the neighbor of unit 3 is 2, and the neighbors of unit 2 are 1 and 3. Initially, the weights are $m_1 = 0.5$, $m_2 = 0.25$ ja $m_3 = 0.75$. Once a new input x has been chosen, the nearest unit is found and the weights of itself and its neighbors are updated according to

$$m_i^{new} = m_i + 0.5(x - m_i).$$

Choose an input x in such a way that after the update the new weights will be in increasing order:

$$m_1^{new} < m_2^{new} < m_3^{new}.$$

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 .
4. (b) Describe the principle of the levelwise algorithm for finding frequent sets.
5. Write an essay about hierarchical clustering. Present the principle and the computing steps of the algorithm. You can use as an example 2-dimensional data and gray scale digital face photos.