

Datasta Tietoon, syksy 2007

HARJOITUSTEHTÄVÄT 5

[pe 7.12.2007, ke 12.12.2007]

H5 / 1. (Kattavat joukot)

Tarkastellaan 0-1 -havaintojoukkoa

a	0	1	1	1	1	0	0	1	1	1
b	1	1	0	1	0	1	1	0	0	1
c	0	0	0	1	1	1	1	1	1	1
d	1	1	0	1	1	1	1	0	0	1

Tässä siis muuttujat ovat a, b, c, d ja havaintoja on 10. Etsi kattavat muuttujajoukot, kun kynnysarvo $N = 4$.

H5 / 2. (Tasoittainen algoritmi)

Mikä on tasoittaisen algoritmin aikavaativuus data koon ja tarkasteltavien ehdokasjoukkojen määrän funktiona?

H5 / 3. (Tšernovin raja)

Tutki luennolla mainittua Tšernovin rajaa. Miten raja käyttäytyy eri parametrien funktiona?

H5 / 4. (Segmentointi)

Tarkastellaan yksiulotteista aikasarjaa jonka 10 pistettä ovat 1, 3, 6, 2, 4, 2, 8, 9, 7, 8. Etsi paras tapa segmentoida tämä aikasarjaa kahteen segmenttiin niin että virhe

$$\sum_{j=1}^k \sum_{i=b_j}^{e_j} (y_i - w_j)^2$$

minimoituu.

H5 / 5. (Segmentointi)

Osoita, että yhden segmentin paras esitys on sen alkoiden keskiarvo, ts. että virhe

$$\sum_{i=b_j}^{e_j} (y_i - w_j)^2$$

minimoituu kun w_j on alkoiden y_{b_j}, \dots, y_{e_j} keskiarvo. Jos minimoitaisiinkin virhettä

$$\sum_{i=b_j}^{e_j} |y_i - w_j|,$$

niin mikä olisi paras valinta arvoksi w_j ?