CS-C3110 Datasta tietoon
Hollmén
Tentti, 3.4.2017

Information for English speakers: You can answer in English, however, the questions are only available in Finnish.

1. Vastaa seuraaviin väittämiiin joko TOSI tai EPÄTOSI.

a) Kun tasoittaista algoritmia ajetaan 0-1 -datalle, kaikki kandidaattijoukot ovat kattavia.
b) Vain osa kattavan joukon osajoukoista on kattavia.
c) PageRank -algoritmillä haetaan kattavia joukkoja.
d) Mahdollisten kattavien joukkojen määrä d-ulotteiselle 0-1 -datalle on 2^d.
e) c-means -algoritmissä klusterikeskukset esitetään vektoreina data-avaruudessa.
f) Hierarkisen klusteroinnin tuloksena saadaan datan ryhmittely optimaaliseen määrään ryhmiä.
g) Hierarkinen klusterointi antaa kaikki klusterointitulokset yhdestä klusterista d klusteriin, jossa d on datavektoreiden dimensio.
h) k:n lähimmän naapurin luokittelijassa valitaan k:n arvoksi aina parillinen kokonaisluku.
i) k:n lähimmän naapurin luokittelijassa parhaan tuloksen antaa aina kaikkein pienimmät k:n arvot.
j) PageRank -algoritmi esittää verkon solmujen relevanssin keskuspainojen ja auktoriteettipainojen avulla.
k) Keskukset ja auktoriteetit -algoritmi esittää verkon solmujen relevanssin kahdella erillisellä painovektorilla.
l) Bayesin kaavassa prioritermi riippuu datasta.
m) MAP -estimoinnissa priorijakauman vaikutus on sitä pienempi, mitä enemmän mittausdataa on käytettävissä.
n) Bayesin kaavassa prioritermi kuvaa käsitystä parametrin todennäköisyysjakaumasta mittauksen jälkeen.
o) lineaarinen suodin voidaan toteuttaa konvoluution avulla
p) Kahden signaalin konvoluutio aikatasossa voidaan toteuttaa niiden taajuusmuunnosten avulla tulona taajuustasossa.
q) Kun yksiulotteisen SOM-kartan yksiköt ovat kerran järjestyneet moniulotteisessa data-avaruudessa, ne eivät voi enää mennä epäjärjestykseen.
r) Itseorganisoiva kartta -algoritmi on kehitetty Suomessa.

2. d-ulotteiset datavektorit ovat tasaisesti jakautuneita hyperkuutioon, jonka sivun pituus on s. Määritellään sisäpisteiksi ne, joiden etäisyys hyperkuution pinnalta on vähintään $\epsilon > 0$. Osoita, että sisäpisteiden joukon kokonaistodennäköisyys (tasainen tiheysjakauma integroituna sisäpisteiden joukon yli) menee nollaan kun $d \to \infty$, toisin sanoen hyvin suurissa dimensioissa melkein kaikki datapisteet ovat hyperkuution pinnalla.

3. Laske suurimman uskottavuuden estimaatti eksponentiaalijakauman

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

parametrille λ kun suureesta x on olemassa otos $x(1), ..., x(n)$.