

ICS-C3000 Datasta tietoon

Hollmén

Tentti, 4.4.2016

Information for English speakers: You can answer in English, however, the questions are only available in Finnish. För svenskspråkiga studerande: Du får också svara på svenska, frågorna finns dock endast på finska.

1. Vastaa seuraaviin väittämiin joko TOSI tai EPÄTOSI.

- Kun tasoittaista algoritmia ajetaan 0-1 -datalle, kaikki kandidaattijoukot ovat kattavia.
- Kattavan joukon kaikki osajoukot ovat kattavia.
- PageRank -algoritmilla haetaan kattavia joukkoja.
- PageRank -algoritmi esittää verkon solmujen relevanssin keskuspainojen ja auktoriteettipainojen avulla.
- Bayesin kaavassa prioritermi riippuu datasta.
- Hierarkisen klusteroinnin tuloksena saadaan datan ryhmittely optimaaliseen määrään ryhmiä.
- c-means -algoritmissä klusterikeskukset esitetään vektoreina data-avaruudessa.
- k:n lähimmän naaprin luokittelijassa valitaan k:n arvoksi aina pariton kokonaisluku.
- k:n lähimmän naapurin luokittelijassa parhaan tuloksen antaa aina kaikkein pienimmät k:n arvot.
- MAP -estimoinnissa priorijakauman vaikutus on sitä pienempi, mitä enemmän mitausdataa on käytettävissä.
- Mahdollisten kattavien joukkojen määrä d -ulotteiselle 0-1 -datalle on 2^d .
- Kahden signaalin konvoluutio aikatasossa voidaan toteuttaa niiden taajuusmuunnosten avulla tulona taajuustasossa.

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

3. Laske suurimman uskottavuuden estimaatti eksponentiaalijakauman

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

parametrille λ kun suuresta x on olemassa otos $x(1), \dots, x(n)$.

4. Oletetaan 5 datavektoria x_1, x_2, \dots, x_5 , joiden etäisyydet toisistaan $d(x_i, x_j)$ on annettuna seuraavana matriisina

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

Tee datavektoreille hierarkinen ryhmittely ja muodosta ryhmittelypuu, olettaen että kahden ryhmän etäisyys on niihin kuuluvien lähimpien vektoreiden etäisyys. Mikä on paras ryhmittely 3 ryhmään?