

ICS-C3000 Datasta tietoon

Hollmén

Tentti, 18.2.2016

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

2. d -ulotteiset datavektorit ovat tasaisesti jakautuneita hyperkuution, 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 nollian kun $d \rightarrow \infty$, toisin sanoen hyvin suurissa dimensioissa melkein kaikki datapisteet ovat hyperkuution pinnalla.

3. Oletetaan kaksi luokkaa skalaarimuuttujalle x . Luokkien tiheysfunktiot $p(x|\omega_1)$ ja $p(x|\omega_2)$ ovat normaalijakautuneita siten että molempien keskiarvo on 0 mutta hajonnat σ_1 ja σ_2 ovat erisuuret. Prioritodennäköisyydet ovat $P(\omega_1), P(\omega_2)$. Piirrä tiheysfunktiot. Johda Bayes-luokittimen luokkarajat.

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