

Valitse jompikumpi aiheista 1 ja 2. Vaihtoehtoisesti voit ehdottaa omaa aihettasi. Oma aihe kannattaa hyväksyttää luennoijalla (ma. prof. Jaakko Hollmén) ennen töihin ryhtymistä.

Harjoitustyön laajuus on 2 opintoviikkoa. Työselostuksen olisi suotavaa olla enintään kymmenen sivun pituinen. Palauta selostus kurssiassistentille (Jouni Seppänen) tammi-kuun 2002 loppuun mennessä.

1. Rakenneoppiminen

Kurssin kotisivulla on datatiedosto, joka sisältää otoksen neljän satunnaismuuttujan A , B , C ja D yhteisjakaumasta. Jokaisen muuttujan arvojoukko on $\{1, 2\}$. Tavoitteena on löytää Bayes-verkkorakenne, joka kuvaa mahdollisimman hyvin muuttujien ehdollisia riippumattomuuksia.

Montako mahdollista verkkorakennetta on? Montako rakenteiden luokkaa on, kun samaan luokkaan sijoitetaan riippumattomuusien suhteen ekvivalentit verkot (kirjan s. 332)? Valitse rakenteille jokin prioriksi, joka suosii yksinkertaisempia rakenteita.

Määritellään verkkorakenteelle S hyvyyskriteeri eräänlaisen ristiinvalidoinnin avulla seuraavasti. Jaa data kymmeneen osaan, etsi S :n paikalliset jakaumat maksimiuskottavuusmenetelmällä käyttäen yhdeksää datan osaa ja laske jäljelle jäävän osan todennäköisyys verkolle S löydettyillä parametreilla. Toista tämä kaikille kymmenelle datan osalle ja käytä todennäköisyyksien keskiarvoa S :n "uskottavuutena". Kerro tämä priorilla. Voit myös käyttää jossain mielessä parempaa hyvyyskriteeriä.

- (a) Käy tyhjentävällä haulla läpi kaikki verkkorakenteet ja valitse paras.
- (b) Toteuta ahne haku: aloita esimerkiksi tyhjästä tai täydestä verkosta ja tarkastele yhden kaaren lisäämistä tai poistamista kerrallaan.

Vertaile menetelmiä. Kuinka suurella muuttujien määrällä tyhjentävä haku kävisi mahdolltomaksi? Jos haluat, voit myös generoida itse moniulotteisempaa dataa ja kokeilla ohjelmaasi sillä.

2. Monte Carlo

Tarkastellaan Lauritzenin ja Spiegelhalterin ASIA-verkkoa (kirjan s. 17). Liitetään siihen todennäköisyyksiä seuraavasti:

Aasian-matkan prioritodennäköisyys on 0,01. Aasiassa käymättömällä on tuberkuloosi todennäköisyydellä 0,01; Aasiassa käyneellä todennäköisyydellä 0,05. Tupakoinnin prioritodennäköisyys on 0,5. Tupakoijalla on keuhkosyöpä todennäköisyydellä 0,1 ja keuhkoputkentulehdus (*bronchitis*) todennäköisyydellä 0,6; tupakoimattomalla todennäköisyydet ovat 0,01 ja 0,3. Jos potilaalla on tuberkuloosi tai syöpä, röntgentutkimus (*X-ray*) antaa positiivisen tuloksen todennäköisyydellä 0,98. Jos potilaalla ei ole kumpaakaan tautia, positiivisen tuloksen todennäköisyys on 0,05.

Hengitysvaikeuksien (*dyspnoea*) todennäköisyys määräytyy seuraavasti: Jos potilaalla on sekä keuhkoputkentulehdus että joko syöpä tai tuberkuloosi, todennäköisyys on 0,9. Jos potilaalla on joko syöpä tai tuberkuloosi mutta ei keuhkoputkentulehdusta, todennäköisyys on 0,7. Jos potilaalla on keuhkoputkentulehdus mutta ei syöpää eikä tuberkuloosia, todennäköisyys on 0,8. Jos potilaalla ei ole mitään näistä taudeista, todennäköisyys on 0,1.

Ratkaise seuraavat kysymykset sekä tarkan päättelyn että jonkin Monte Carlo -menetelmän avulla.

- (a) Potilaalla on hengitysvaikeuksia ja hän on käynyt Aasiassa. Mitkä ovat tautien marginaalitodennäköisyydet?
- (b) Edellisen lisäksi tiedetään, että potilas tupakoi. Mitkä ovat todennäköisyydet nyt?
- (c) Edellisen kohdan tilanteessa lääkäri tekee röntgentutkimuksen. Miten negatiivinen tai positiivinen tulos vaikuttaa todennäköisyyksiin?

Kuinka pitkään Monte Carlo -ajoa täytyy jatkaa, jotta tulokset ovat kahden merkitsevän numeron tarkkuudella oikeat?

Jos haluat, voit myös kokeilla antaa esimerkiksi hengitysvaikeussolmulle jonkin jatkuvan jakauman.

Bayes Net Toolbox

BNT (<http://www.cs.berkeley.edu/~murphyk/Bayes/bnt.html>) on Matlab 5:ssä toimiva kirjasto, josta voi olla hyötyä harjoitustyön tekemisessä. Informaatiotekniikan laboratorion koneilla BNT:n saa käyttöön antamalla Matlabissa seuraavat komennot:

```
addpath /r/matlab
add_BNT_to_path
```

Jälkimmäinen komento on nimetty hieman väärin, sillä polkuasetusten lisäksi se alustaa useita BNT:n käyttämiä globaaleja muuttujia. Ensimmäisen komennon voi turvallisesti lisätä Matlabin `startup.m` -tiedostoon. Käyttöohjeita on BNT:n WWW-sivuilla.