TAMPEREEN TEKNILLINEN YLIOPISTO
Ohjelmistotekniikan laitos

OHJ-2550 **Tekoäly** Laskuharjoitus 6
prof. Tapio Elomaa 2. 4. 2009

1. Tekstin luokittelussa annetut dokumentit luokitellaan yhteen ennalta määrätyistä luokista sisältönsä perusteella. Naiivia Bayes -mallia käytettäessä dokumentin luokka on kyselymuuttuja ja havaintomuuttujia ovat sanojen sisältyminen/ sisältymättömyys dokumenttiin. Tausta-ajatus on sanojen riippumaton esiintyminen dokumenteissa siten, että dokumentin luokka määrää esiintymien frekvenssin.

   (a) Kuinka malli voidaan muodostaa, kun on annettu opetusaineisto, joka koostuu joukosta valmiiksi luokiteltuja dokumentteja?

   (b) Miten uusia dokumentteja voidaan luokitella?

   (c) Onko naiivin Bayes -mallin riippumattomuusoletus sovelias tähän tehtävään?

2. Ydinvoimalassa on hälytin, joka aistii milloin lämpötila-anturi mittaa annettua kynnysarvoa korkeamman lämmön. Anturi mittaa ytimen lämpötilaa. Totuusarvoisia muuttujia ovat H (hälytys), RH (hälytin on rikki) ja RA (anturi on rikki). Moniarvoisia puolestaan ovat A (anturin lukema) ja T (ytimen lämpötila).

   (a) Piirrä Bayes-verkko ydinvoimalalle, kun anturi hajoaa sitä helpommin, mitä korkeampi ytimen lämpötila on.

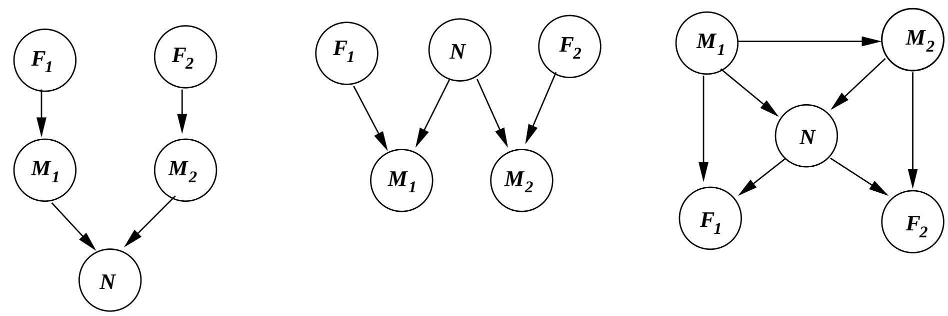
   (b) Olkoot myös A ja T binääriarvoisia arvoinaan normaali ja korkea. Mikä on solmun A ehdollinen todennäköisyystaulu, kun anturin todennäköisyys antaa oikea lukema on x ehjänä ja y rikkinäisenä?

3. Edelliseen liittyen.

   (a) Oletetaan, että hälytin toimii oikein jos se on ehjä. Muutoin se ei koskaan anna hälytystä. Mikä on solmun H ehdollinen todennäköisyystaulu?

(b) Oletetaan, että hälytin ja anturi ovat ehjiä kun hälytys tulee. Laske todennäköisyys, että ytimen lämpötila on liian korkea, verkon muin ehdollisin todennäköisyyksin ilmaistuna.

4. Kaksi astronomia eri puolilla maailmaa tekevät teleskoopeillaan mittaukset M_1 ja M_2 pienen alueen tähtien määrästä N. Yleensä on olemassa pieni mahdollisuus e korkeintaan yhden tähden virhearvioon. Teleskoopit voivat kuitenkin hyvin pienellä todennäköisyydellä f olla fokusoimattomia (tapahtumat F_1 ja F_2), jolloin tähtitieteilijä laskee tähtien määrän alakanttiin ainakin kolmella tähdellä. Tarkastellaan seuraavia Bayes-verkkoja tähän tehtävään.



(a) Mitkä verkoista ovat oikeita (eivät välttämättä tehokkaita) kuvauksia tälle ongelmalle?

(b) Mikä verkoista on paras? Perustele.

(c) Kirjoita parametrien e ja f funktiona ehdollinen jakauma $\underline{\mathbf{P}}(M_1 \mid N)$, kun $N \in \{1, 2, 3\}$ ja $M_1 \in \{0, 1, 2, 3, 4\}$

(d) Olkoot $M_1 = 1$ ja $M_2 = 3$. Jos tähtien lukumäärää N ei ole ennalta rajoitettu, niin mitkä ovat näiden havaintojen perusteella mahdollisia lukumääriä?

5. Tarkastellaan yo. kuvan keskimmäistä verkkoa ja oletetaan molempien teleskooppien toimivan identtisesti. $N \in \{1, 2, 3\}$ ja $M_1 \in \{0, 1, 2, 3, 4\}$ ja solmujen ehdollisten todennäköisyyksien taulukot ovat edellisessä tehtävässä määrätyt. Laske todennäköisyysjakauma $\underline{\mathbf{P}}(N \mid M_1 = 2, M_2 = 2)$.