

Tilastollinen hahmontunnistus (kuulustelija: Petri Koistinen)
Tentti 20.4.2004

Kirjoita vastauspaperiin nimesi lisäksi joko sosiaaliturvatunnus tai opiskelijanumero.

1. Selitä lyhyesti seuraavat käsitteet.

- a) Virhetodennäköisyyden takaisinsijoitusestimaattori. (2 p)
- b) Sekoiteotanta. (2 p)
- c) Bayesin luokitin. (2 p)

2. Satunnaismuuttujaparilla (X, Y) on jakauma, jonka tiheysfunktio on

$$f_{XY}(x, y) = \frac{e^{x+y}}{c^2} \mathbb{1}_{[0,1]}(x) \mathbb{1}_{[0,1]}(y),$$

jossa $c = e - 1$. Diskreetti satunnaismuuttuja J saa arvon yksi, jos $Y \leq 1 - X$ ja muussa tapauksessa arvon kaksi. Tilanteessa havaitaan ainoastaan satunnaismuuttujan X arvo, jonka perusteella satunnaismuuttujan J arvo pitää arvata.

- a) Laske seuraavat suureet: luokkien prioritodennäköisyydet P_1 ja P_2 , luokkien tiheysfunktiot f_1 ja f_2 , luokkien posterioritodennäköisyydet $P(1 | x)$ ja $P(2 | x)$ sekä piirteen X tiheysfunktio. (4 p)
 - b) Määritä Bayesin luokitin g^* sekä kirjoita sen virhetodennäköisyydelle integraalilauseke a-kohdassa laskettujen suureiden avulla. (Integraalilausekkeen arvoa ei tarvitse laskea eikä sieventää.) (2 p)
3. Tässä tehtävässä käytetään sekoiteotannalla saatua opetusaineistoa.
- a) Selosta lyhyesti, miten k -lähinaapuriluokitin toimii. (2 p)
 - b) Miten k :n arvo valitaan k -lähinaapuriluokittimessa v -kertaisella ristiinvalidoinnilla? (4 p)

4. Käytävissä on aineisto $(\mathbf{X}_1, J_1), \dots, (\mathbf{X}_n, J_n)$, joka on saatu sekoiteotannalla c :n luokan hahmontunnistustehtävistä. Aineiston alustavan tarkastelun jälkeen ollaan päätetty, että luokkien jakaumia voidaan approksimoida normaalijakaumilla $N(\boldsymbol{\mu}_j, C_j)$, jossa luokkien kovarianssimatriiseja ei voida pitää samoina. Tehtävänäsi on selittää, miten tässä tilanteessa muodostetaan estimoitu (normaaliseen malliin perustuva) kvadraattinen luokitin.

Kirjoita selitys siten, että matemaattiseen notatioon sekä matemaattisten ohjelmistojen käyttöön harjaantunut toverisi (joka ei tiedä mitään tilastotieteestä tai hahmontunnistuksesta) voi kuvauksen perusteella kirjoittaa tietokoneohjelman, joka toteuttaa kyseisen luokittimen. (6 p)

5.

- a) Määrittele käsitteet *ortogonaalinen matriisi* sekä matriisin *singulaariarvohajotelma*. (3 p)
- b) Olkoon mitattuna D -ulotteiset hahmovektorit $\mathbf{Z}_1, \dots, \mathbf{Z}_n$, ja olkoon tavoitteena laskea niistä d -ulotteiset piirrektorit $\mathbf{X}_1, \dots, \mathbf{X}_n$ pääkomponenttianalyysillä, kun pääkomponentit määritellään hahmovektorien \mathbf{Z}_i otoskovarianssimatriisin d :tä suurinta ominaisarvoa vastaavien ominaisvektorien avulla. Selitä perustellusti, miten piirrektorit voidaan laskea ilman otoskovarianssimatriisin muodostamista, jos sen sijaan lasketaan matriisin

$$B = [\mathbf{Z}_1 - \hat{\boldsymbol{\mu}}, \dots, \mathbf{Z}_n - \hat{\boldsymbol{\mu}}]$$

singulaariarvohajotelma, missä $\hat{\boldsymbol{\mu}}$ on hahmovektorien \mathbf{Z}_i otoskeskiarvo. (3 p)