

SGN-2500 Johdatus hahmontunnistukseen
Tentti 3 5.10.2009 /Jussi Tohka

Huomaa, että laskutehtävissä oikeat vastaukset ilman selitystä kuinka ne ollaan saatu tuottavat nolla pistettä, joten selitä miten laskit vastaukset! Muista myös määritellä KAIKKI vastauksissa esiintyvät symbolit. Laskimen käyttö on sallittua.

1. Hahmontunnistusjärjestelmän perusrakenne. Esitä ja selitä viisi luentomonisteessa mainittua hahmontunnistusjärjestelmien perusvaihetta. (6p)
2. (a) Määrittele Bayes luokitin c :n luokan luokitusongelmalle kun luokkatiheysfunktiot $p(\mathbf{x}|\omega_i)$ ja prioritodennäköisyydet $P(\omega_i)$ tunnetaan. (1p)
 (b) Määrittele erotinfunktiot ja lineaariset erotinfunktiot. (2p)
 (c) Oletetaan, että yhden piirteen luokitusongelmassa

$$p(x|\omega_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu_i}{\sigma^2}\right)^2\right], i = 1, \dots, c.$$

(Siis varianssi σ^2 on kaikille luokille sama.) Johda Bayes luokitinta vastaavat lineaariset erotinfunktiot. (3p)

3. Ajatellaan kolmen kategorian Parzen-ikkuna-luokitinta ja seuraavia kaksipiirteisiä luokiteltuja harjoitusnäytteitä

ω_1	(7,4)	(14,9)	(16,2)
ω_2	(4,3)	(3,4)	(0,6)
ω_3	(5,10)	(3,9)	(1,7)



Opeta luokitin ja luokita piirrevektorit (5,5) ja (3,8). Ikkunafunktioksi valitaan

$$\varphi(\mathbf{u}) = \frac{1}{(2\pi)^{d/2}} \exp[-0.5\mathbf{u}^T \mathbf{u}]$$

ja ikkunanleveys $h_n = 1$. Prioritodennäköisyydet $P(\omega_i) = 1/3$ jokaiselle luokalle. (6p)

4. Henkilö A epäilee olevansa allerginen matematiikalle ja tekee testin tämän selvittämiseksi. Testin tiedetään olevan 95%:sti oikeassa. Tämä tarkoittaa, että 95% positiivisista testituloksista on oikeassa ja 95% negatiivisista testituloksista on oikeassa. Henkilö A kuuluu 100000:n henkilön joukkoon, joista 4000 on allergisia matematiikalle. Testi on positiivinen. Mikä on todennäköisyys, että henkilö A on allerginen matematiikalle? (**Vinkki:** Bayesin lause...) (6p)
5. K-means klusterointi. Vastauksestasi tulisi käydä ilmi mihin K-means klusterointia käytetään, mitä kriteerifunktiota käytetään, ja mikä on itse algoritmi. Lyhyt kuvaus algoritmin/kriteerifunktion huonoista puolista vaaditaan myös kuuteen pisteeseen. (6p)