SGN-2500 Johdatus hahmontunnistukseen
Tentti 1 9.5.2011 /Jukka-Pekka Kauppi

Vastaa kaikkiin viiteen tehtävään 1-5. Jokaisen tehtävän maksimipistemäärä on 6 pistettä. Tentissä ei saa olla mukana kirjallista materiaalia. Oman laskimen käyttö on sallittua. Huomaa, että laskutehtävissä pelkät oikeat vastaukset ilman perustelua tuottavat nolla pistettä, joten muista esittää laskujen kaikki keskeiset välivaiheet ja selvittää mitä olet laskemassa. Muista myös määritellä vastauksissasi esiintyvät symbolit. Tenttipaperia ei tarvitse palauttaa.

Tehtävät:

1. Kirjoita essee aiheesta *hahmontunnistusjärjestelmä*. Esitä ja selvitä viisi luentomonisteessa mainittua hahmontunnistusjärjestelmien perusvaihetta. Hyvä vastaus sisältää myös lyhyesti pohdintaa siitä, minkälaisia asioita mielestäsi keskeisimpien vaiheiden suunnittelussa tulisi ottaa (yleisellä tasolla) huomioon. Kokonaispituus noin 300-1000 sanaa. (6p)

2. Henkilö A epäilee olevansa allerginen matematiikalle ja tekee testin tämän selvittämiseksi. Testin tiedetään olevan 98%:sti oikeassa. Tämä tarkoittaa, että 98% positiivisista testituloksista on oikeassa ja 98% negatiivisista testituloksista on oikeassa. Henkilö A kuuluu 200000:n henkilön joukkoon, joista 7000 on allergisia matematiikalle. Testi on positiivinen. Mikä on todennäköisyys, että henkilö A on allerginen matematiikalle? (**Vinkki**: Bayesin kaava...) (6p)

3. On saatu seuraava data (jonka piirrevektorit siis kuuluvat joko luokkaan ω_1 tai ω_2 kuten alla on esitetty):

$$\omega_1 : [0,1]^T, [1,0]^T, \quad \omega_2 : [0,0]^T, [1,1]^T.$$

   (a) Kokeile perceptron-algoritmia käyttäen $\mathbf{a}(0) = [0,1,1]^T, \eta = 1$ ja $\epsilon = 0$.

   Perceptron-algoritmi:

   Aseta $t \leftarrow 0$, alusta $\mathbf{a}(0)$, η, ϵ.
   while $-\sum_{j:\mathbf{a}(t)^T\mathbf{y}_j \leq 0} \mathbf{a}(t)^T\mathbf{y}_j > \epsilon$ **do**
   $\quad \mathbf{a}(t+1) = \mathbf{a}(t) + \eta \sum_{j:\mathbf{a}(t)^T\mathbf{y}_j \leq 0} \mathbf{y}_j$
   $\quad$ Aseta $t \leftarrow t+1$
   end while
   Palauta $\mathbf{a}(t)$.

   Ota kolme askelta eli laske $\mathbf{a}(3)$. (4p)

   (b) Miten a)-kohdassa nähdään arvosta $\mathbf{a}(3)$, että kolmannen askeleen jälkeen on hyödytöntä jatkaa iterointia? (1p)

   (c) Miksei Perceptron-algoritmi suppene a)-kohdassa? (1p)

4. Tarkastellaan ohjaamatonta oppimista.

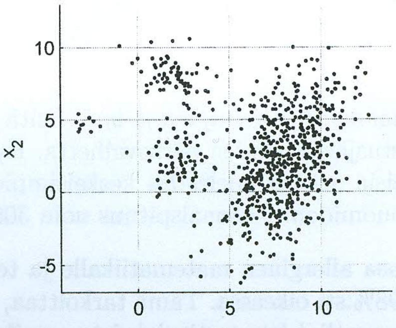
   (a) Olkoon $\mathbf{x}_1 = [1,2,1]^T, \mathbf{x}_2 = [2,1,0]^T, \mathbf{x}_3 = [2,2,1]^T, \mathbf{x}_4 = [2,1,0.5]^T$, ja ajatellaan seuraavia partitioita:

      i. $\mathcal{D}_1 = \{\mathbf{x}_1, \mathbf{x}_2\}, \mathcal{D}_2 = \{\mathbf{x}_3, \mathbf{x}_4\}$
      ii. $\mathcal{D}_1 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}, \mathcal{D}_2 = \{\mathbf{x}_4\}$

   Laske k-means-kriteerin (eli neliösummakriteerin) arvot näille partitioille. Kumpaa partitiota k-means kriteeri suosii? (4p)

Hyödyllisiä kaavoja: $J(D_1, D_2, ..., D_c) = \sum_{i=1}^{c} \sum_{\mathbf{x} \in D_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2$ ja $\boldsymbol{\mu}_i = \frac{1}{|D_i|} \sum_{\mathbf{x} \in D_i} \mathbf{x}$.

(b) Tavoitteenasi on klusteroida Kuvassa 1 esitetty data neljään luokkaan. Käytössäsi on k-means algoritmi sekä EM-algoritmi (EM = expectation maximization) sekoitemallille. Kumman menetelmän valitset tehtävään? Perustele miksi! (2p)



Kuva 1: Datajoukko tehtävässä 4(b).

5. Päätöspinta ja päätösalueet.

(a) Ajatellaan kahden luokan luokitusongelmaa, jossa $P(\omega_1) = 0.9$, $P(\omega_2) = 1 - P(\omega_1)$, $p(x|\omega_1) = \frac{1}{\sqrt{2\pi}} e^{-0.5(x-0.5)^2}$ ja $p(x|\omega_2) = \frac{1}{\sqrt{2\pi}} e^{-0.5(x-2)^2}$. Etsi päätösalueet Bayes-luokittimelle. (2p)

(b) Havainnollista tilannetta graafisesti. (1p)

(c) Onko kyseinen luokitin lineaarinen? Perustelu! (1p)

(d) Ajatellaan kahden luokan lineaarista luokitinta, jonka erotinfunktio on $g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$. Osoita, että päätöspinnan normaali on painovektori w:n suuntainen. (2p)