

SGN-2500 Johdatus hahmontunnistukseen, Kevät 2010
Tentti 20.9.2010 / Jari Niemi

Vastaa kaikkiin viiteen tehtävään 1-5. Jokaisen maksimipistemäärä on 6 pistettä. Ei kirjallisuutta. Tarvittavat kaavat annetaan tehtävien yhteydessä. Oma funktiolaskin sallittu. Ohjelmoitavaa/graaafista laskinta ei saa käyttää.

Tehtävät:

1. 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 $\varepsilon = 0$.

Perceptron-algoritmi:

Aseta $t \leftarrow 0$, Alusta $\mathbf{a}(0), \eta, \varepsilon$.

```
while  $-\sum_{\mathbf{y}: \mathbf{a}(t)^T \mathbf{y} < 0} \mathbf{a}(t)^T \mathbf{y} > \varepsilon$  do
     $\mathbf{a}(t+1) = \mathbf{a}(t) + \eta \sum_{\mathbf{y}: \mathbf{a}(t)^T \mathbf{y} < 0} \mathbf{y}$ 
    Aseta  $t \leftarrow t+1$ 
```

end while

Palauta $\mathbf{a}(t)$.

Muista jatkaa ja normalisoida piirrevektorit. Ota kolme askelta eli laske $\mathbf{a}(3)$. (4 p.)

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

Näytä kaikki laskujen vaiheet ja selitä ne lyhyesti.

2. Kaksi korttia vedetään satunnaisesti normaalista 52 kortin pakasta^{*}.

- a) Määrittele sopiva todennäköisyysavaruus tälle kokeelle (otosavaruus, tapahtumat, todennäköisyys). (2 p.)
- b) Mikä on todennäköisyys, että toinen kortti on ässä olettaen, että ensimmäinen oli ässä? (2 p.)
- c) Mikä on todennäköisyys, että toinen kortti on ässä? (2 p.)

^{*} Normaalissa 52 kortin pakassa on 13 korttia (Ässä, 2, 3, ..., 10, Prinssi, Kuningatar, Kuningas) jokaista neljää maata (Pata, Risti, Hertta, Ruutu) kohti.

Näytä kaikki laskujen vaiheet ja selitä ne ja saamasi tulokset lyhyesti.

Käännä →

3. On saatu seuraava opetusdata (sekoiteotannalla) luokista ω_1, ω_2 ja ω_3 :

$$\omega_1: [4,3]^T, [3,4]^T, [16,8]^T, \quad \omega_2: [7,3]^T, [14,9]^T, [0,5]^T, [3,6]^T, \quad \omega_3: [5,10]^T, [3,9]^T, [12,6]^T.$$

Opetä Parzen-luokittaja ja luokita sillä piirre $[7,7]^T$ (perustuen Bayesin minimivirheluokitussääntöön). Muista estimoida prioritodennäköisyydet. (6 p.)

Ohje: Käytä ikkunafunktiona $\varphi(\mathbf{u}) = \frac{1}{2\pi} e^{-\frac{1}{2}\mathbf{u}^T\mathbf{u}}$ ja leveysparametrille arvoa 1.

Hyödyllisiä kaavoja: ($|\omega|$ = ”joukon ω alkioden lukumäärä”.)

$$\hat{P}(\omega_i) = \frac{|\omega_i|}{\sum_{i=1}^c |\omega_i|}, P(\omega|\mathbf{x}) = \frac{P(\omega)p(\mathbf{x}|\omega)}{p(\mathbf{x})}, p_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{V_n} \varphi\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_n}\right)$$

Näytä kaikki laskujen vaiheet ja selitä ne ja saamasi luokitustulos lyhyesti.

4. Allaolevat yhtälöt ja epäyhtälöt perustelevat k :n lähimmän naapurin luokittajan mielekkääksi. Mutta niistä puuttuvat luonnollisen kielen selitykset. Täydennä esitys sanoin selkeäksi niin, että lukija voi ymmärtää k :n lähimmän naapurin luokittajan olevan hyvin perusteltu luokitusmenetelmä. Yhtälöitä ja epäyhtälöitä ei siis tarvitse todistaa, vaan riittää selittää, mitä ne tarkoittavat. (6 p.)

$$p_n(\mathbf{x}, \omega_i) = \frac{k_i}{nV}; \quad P_n(\omega_i|\mathbf{x}) = \frac{p_n(\mathbf{x}, \omega_i)}{p_n(\mathbf{x})} = \frac{p_n(\mathbf{x}, \omega_i)}{\sum_{j=1}^c p_n(\mathbf{x}, \omega_j)} = \frac{k_i}{\sum_{j=1}^c k_j} = \frac{k_i}{k};$$

$$\alpha_{knn}(\mathbf{x}) = \arg \max_{\omega_i} P_n(\omega_i|\mathbf{x}); \quad E(\alpha_{bayes}(\mathbf{x})) \leq E(\alpha_{knn}(\mathbf{x})) \leq E(\alpha_{bayes}(\mathbf{x})) + \sqrt{\frac{2E(\alpha_{nn}(\mathbf{x}))}{k}}.$$

5. Kirjoita essee (n. 300-1000 sanaa) otsikolla *Ohjattu ja ohjaamaton luokitus*. (6 p.)

Ohje: Käsittele ainakin seuraavat asiat: ohjatun ja ohjaamattoman luokituksen määritelmät, kummankin hyödyt ja haitat toisiinsa verrattuina sekä kummallekin esimerkkejä erityyppisistä estimointi- ja luokitusmenetelmistä (käy läpi lyhyesti: suurin uskottavuus, Parzen, k lähintä naapuria, k -means, sekoitemallit ja näiden virhelähteet).