

Vastata voi suomeksi, ruotsiksi tai englanniksi.

Answers can be given in English, Finnish or Swedish.

1. Selitä lyhyesti, noin 20–40 sanalla tai matemaattisella määritelmällä, seuraavat käsitteet tai lyhenteet: 6p.

- (i) Bayesin kaava *a posteriori* -todennäköisyyden laskemiseksi
- (ii) Parzen-ikkunat
- (iii) perseptroni
- (iv) Tiling-algoritmi
- (v) graafien vertailumenetelmät
- (vi) *C-means*-algoritmi

2. Kirjoita essee aiheesta Lauserakennekieliopit (*phrase-structure grammars*) hahmontunnistuksessa. Käsittele ainakin seuraavia kysymyksiä: (i) Motivaatio ja pääperiaate. (ii) Yleinen määritelmä. (iii) Lauserakennekielioppien tyypit ja niiden määritelmät ja suhteet toisiinsa. (iv) Symbolijonojen tunnistus lauserakennekieliopilla. (v) *Top-down*-jäsenitys esimerkin avulla. (vi) *Bottom-up*-jäsenitys esimerkin avulla. 6p.

3. Kahden luokan keskiarvot ovat $\mathbf{m}_1 = [-2 \ -2]^T$ ja $\mathbf{m}_2 = [2 \ 2]^T$. Luokkien sisäiset sirontamatriisit (scatter matrices) ovat (α on parametri):

$$\mathbf{S}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{ja} \quad \mathbf{S}_2 = \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix}.$$

(i) Hahmottele luokkien tiheysfunktiot olettaen luokkien olevan (approksimatiivisesti) Gaussisesti jakautuneita ja apriorisesti yhtä todennäköisiä.

(ii) Laske Fisherin lineaarisen diskriminantin suunta käyttäen kaavaa

$$\hat{\mathbf{w}} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2), \text{ missä } \mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2.$$

(iii) Määritä Fisherin lineaarisen diskriminantin suunta laskemalla matriisin $\mathbf{S}_W^{-1}\mathbf{S}_B$ ominaisvektorit, kun $\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$.

(iv) Vertaa kahden edellisen kohdan tuloksia ja kommentoi.

(v) Hahmottele Fisherin lineaarinen diskriminantti tapauksissa, joissa $\alpha = 1$, $\alpha = 4$, $\alpha \rightarrow \infty$ ja $\alpha \rightarrow 0$.

(vi) Arvioi saamiasi tuloksia ja luokittelijoiden toimivuutta parametrin α eri arvoilla. 6p.

4. (i) Esitä lineaariseen diskriminanttifunktioon perustuvan luokittimen yleinen muoto. (ii) Selitä ja piirrä XOR-ongelma ja osoita, että sille ei ole lineaarista ratkaisua. (iii) Esitä yleistetyn lineaarisen luokittimen yleinen muoto, jossa alkuperäiselle piirrevektorille \mathbf{x} on suoritettu epälineaarit kuvaukset $\mathbb{R}^l \rightarrow \mathbb{R}$ uuden piirrevektorin $\mathbf{y} \in \mathbb{R}^k$ komponenttien laskemiseksi. (iv) Esitä edellisen kohdan mallin mukainen epälineaarinen luokitin, jossa kantafunktioina käytetään toisen asteen polynomeja. (v) Muodosta XOR-ongelmassa uudet piirrevektorit käyttäen kantafunktioita $\mathbf{y} = [1, x_1, x_2, x_1x_2]$ ja osoita, että ne separoituvat lineaarisesti hypertason $\mathbf{w} = [-1/4, 1, 1, -2]^T$ avulla. (vi) Hahmottele alkuperäiseen XOR-ongelman kuvaan edellä mainitun epälineaarisen diskriminanttifunktion päätöspinta $g(\mathbf{x}) = 0$. 6p.

5. Tutkitaan yksinkertaista sekventiaalista klusterointia (*Basic Sequential Algorithmic Scheme*, BSAS) käyttäen yksidimensionaalista havaintojoukkoa: $X = \{\mathbf{x}_1 = 0, \mathbf{x}_2 = 2, \mathbf{x}_3 = 5\}$. (i) Luettele kaikki mahdolliset tavat jakaa havaintojoukko X klustereiksi. (ii) Mitkä ovat BSAS-algoritmin kaksi parametria q ja Θ ? Entä algoritmiin liittyvä funktio $d(\mathbf{x}_i, C_j)$? (iii) Selitä BSAS-algoritmin toiminta askel askeleelta. (iv) Käytetään nyt $d(\mathbf{x}_i, C_j) = d(\mathbf{x}_i, \mathbf{m}_j)$, missä $d(\cdot, \cdot)$ on eukliidinen etäisyys ja \mathbf{m}_j on C_j :n havaintojen keskiarvo. Kokeile kaikki havaintojoukon läpikäyntijärjestykset ja selvitä kullekin niistä, millä parametrien arvovalinnoilla algoritmi tuottaa erilaiset klusterointitulokset. (v) Esitä edellisen kohdan tuloksista graafinen yhteenveto, joka kertoo, montako klusteria eri Θ -parametrin arvoilla keskimäärin muodostui, kun $q = 3$. (vi) Kommentoi tuloksiasi ja BSAS-algoritmin hyviä ja huonoja puolia. 6p.