

Vastata voi suomeksi, ruotsiksi tai englanniksi.

Kurssipalautetta antamalla voit ansaita yhden ylimääräisen pisteen!

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

- (i) XOR-ongelma
- (ii) Coverin teoreema
- (iii) mikstuurimallit
- (iv) graafien homo- ja isomorfismit
- (v) Fisherin diskriminantti
- (vi) Softmax-skaalaus skalaarijoukolle $\{x_1, x_2, \dots, x_N\}$

2. Kirjoita essee aiheesta Tukivektorikone (*support vector machine*). Käsittele ainakin seuraavia kysymyksiä: (i) Motivaatio ja pääperiaate. (ii) Lineaarisesti separoituvat luokat. (iii) Kustannusfunktio ja sen duaaliohjelma. (iv) Lineaarisesti separoitumattomat luokat. (v) Epälineaarinen yleistys eli *kernel trick*. (vi) Kantafunktiot. 6p.

3. (i) Millainen on lineaarisen diskriminanttifunktion yleinen muoto? Miksi piirrekuvoreihin yleensä lisätään ylimääräinen vakiokomponentti? (ii) Millainen on, miten toimii ja mihin soveltuu perseptroni-algoritmi? (iii) Olkoon datajoukko $\{\mathbf{x}_1 = (-1, 0), \mathbf{x}_2 = (0, 1), \mathbf{x}_3 = (0, -1), \mathbf{x}_4 = (1, 0)\}$, joista $\mathbf{x}_1 \in \omega_1, \mathbf{x}_2 \in \omega_1, \mathbf{x}_3 \in \omega_2$ ja $\mathbf{x}_4 \in \omega_2$. Piirrä kuva datapisteistä ja kerro, onko tehtävä lineaarisesti separoituva vai ei. (iv) Olkoon perseptronin alkutila $\mathbf{w}(0) = [0, 0, 0]^T$ ja opetuskerroin $\rho = 1$. Esitä askeleet, joilla perseptroni löytää ratkaisun, kun opetuksessa käytetään yhtä vektoria kerrallaan järjestyksessä $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_1, \dots$ (v) Kommentoi perseptronin toimintaa ja saamaasi tulosta. 6p.

4. (i) Selitä *dynamic time warping* (DTW) -algoritmin yleinen toimintaperiaate. (ii) Mikä suhde DTW-algoritmilla on Bellmanin optimaalisuusperiaatteeseen? (iii) Selitä DTW-algoritmiin perustuva symbolijonojen vertailussa käytettävä (Levenshteinin) "editointietäisyys". (iv) Oletetaan, että editointietäisyyden laskemisessa symbolin vaihdolla, lisäyksellä ja poistolla on sama yksikkökustannus. Päättele tällöin symbolijonojen KISSA, KOIRA, KUKKA ja KIRJA väliset kaikki etäisyydet. (v) Oletetaan sitten, että symbolin vaihdon kustannus onkin kaksi yksikköä, mutta lisäyksen ja poiston edelleen yksi. Millaisiksi edellisen kohdan etäisyydet nyt muodostuvat? (vi) Arvioi saamiasi tuloksia. Kumpi editointietäisyyden toteutus vastasi paremmin intuitiotasi tai on muutoin mielestäsi perustellumpi? 6p.

5. Tarkastellaan alla näkyvää matriisia P , jonka lukuarvot p_{ij} ilmaisevat kahden klusterin välisen etäisyyden. (i) Käytä yksinkertaista linkkialgoritmia (*single link*) hierarkkisen *bottom-up* klusteroinnin toteuttamiseksi. Selitä algoritmi ja näytä P -matriisin kehitys ratkaisun kaikissa vaiheissa. (ii) Piirrä klusterointia kuvaava dendrogrammi ja merkitse siihen kuhunkin klustereiden yhdistymiseen johtanut klustereiden välinen etäisyys. (iii) Toteuta hierarkkinen *bottom-up* klusterointi käyttäen nyt täydellistä linkkialgoritmia (*complete link*). (iv) Piirrä myös tätä klusterointia kuvaava dendrogrammi etäisyystietoineen. (v) Vertaile ja kommentoi tuloksiasi. Kerro lisäksi, mikä suhde niillä on minimiviritäjäpuuhun (*minimum spanning tree*). (vi) Millä tavoin etäisyysmatriisin päivittämiseen perustuvat klusterointimenetelmät eroavat kasaavan, hierarkkisen klusterointialgoritmin yleistetystä muodosta?

6p.

$$P = \begin{bmatrix} 0 & 4 & 9 & 6 & 5 \\ 4 & 0 & 1 & 8 & 7 \\ 9 & 1 & 0 & 3 & 2 \\ 6 & 8 & 3 & 0 & 1 \\ 5 & 7 & 2 & 1 & 0 \end{bmatrix}$$