

# Datasta Tietoon, syksy 2008

TENTTI

20. 12. 2008

(note: problems in English on the reverse side)

1.

Tiivistämättömän digitaalisen kuvan kunkin kuvapisteen (pikselin) väri esitetään kolmen luvun R (red), G (green) ja B (blue) yhdistelmänä (3-ulotteisena vektorina) jossa kullakin kolmesta luvusta on 8 bitin pituinen binääriesitys.

a) Montako erillistä väriä voidaan esittää?

2<sup>8·3</sup> värejä

b) Tehdään pikseleiden värihistogrammi, jossa jokaisella erillisellä värillä on oma lokeronsa. Halutaan että kuhunkin lokeroon osuu keskimäärin 10 pikseliä. Kuinka suuri digitaalinen kuva tähän tarvitaan?

(huom: likiarvot sallitaan).

2.

Oletetaan kaksi luokkaa skalaarimuuttujalle  $x$ . Luokkien tiheysfunktiot  $p(x|\omega_1), p(x|\omega_2)$  ovat normaalijakautuneita siten että molempien keskiarvo on 0 mutta hajonnat  $\sigma_1, \sigma_2$  ovat erisuuret. Prioritodennäköisyydet ovat  $P(\omega_1), P(\omega_2)$ . Piirrä tiheysfunktiot. Johda Bayes-luokittimen luokkarajat.

3.

Oletetaan 5 vektoria  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_5$ , joiden etäisyydet toisistaan  $d(\mathbf{x}_i, \mathbf{x}_j) = D_{ij}$  on annettuna seuraavana matriisina

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

Tee vektoreille hierarkkinen ryhmittely ja muodosta ryhmittelypuu, olettaen että kahden ryhmän etäisyys on niihin kuuluvien lähimpien vektoreiden etäisyys. Mikä on paras ryhmittely 3 ryhmään?

4.

Etsittäessä relevantteja sivuja verkosta käytetään joskus keskus- ja auktoriteettipainoihin (hubs and authorities) perustuvaa menetelmää. Kuvaava tämän menetelmän toimintaperiaate ja anna laskukaavat, joilla painot lasketaan.

5.

Vastaa jompaan kumpaan seuraavista esseeaiheista, jotka liittyvät Matlab-harjoitustehtävään:

(a) Kirjoita essee aiheesta äänenkäsittelyn perusteet. Luetaan (esimerkiksi) Matlabin sisälle äänisignaali, jossa on puheen seassa häiriöllinen sinisignaali taajuudella 2400 Hz. Millaisia työkaluja ja visualisointeja signaalin analysointiin on tarjolla? Mitä voidaan tehdä häiriön poistamiseksi?

(b) Kirjoita essee aiheesta ominaiskasvot ("eigenfaces"). Olkoon käytettävissä joukko digitaalisia harmaasävyllisiä kasvokuvia, jotka on keskitetty yhtenevästi ja kuvattu samanlaisissa valaistusominaisuuksissa. Esitä pääkomponenttianalyysin (PCA) periaate ja laskennan vaiheet. Miten PCA:ta voidaan käyttää kasvokuvien kanssa?

HUOM! Annathan palautetta käymistäsi kurseista osoitteessa:

<<http://www.cs.hut.fi/Opinnot/Palaute/kurssipalaute.html>>

OBS! Var så god och ge kursfeedback i följande address:

<<http://www.cs.hut.fi/Opinnot/Palaute/kurssipalaute.html>>

# Datasta Tietoon, Autumn 2008

EXAM

20. 12. 2008

(note: problems in Finnish on the reverse side)

1.

The color of each pixel in an uncompressed digital image is represented as a combination of three numbers R (red), G (green) and B (blue) (a three dimensional vector) in which each number is represented by an 8 bit binary number.

- a) How many distinct colors can be represented?
- b) Let us make a color histogram for the pixels in which each distinct color has its own bin. We want to have at least 10 pixels on the average to go to each bin. How large a digital image is needed?

(note: approximations are allowed).

2.

Assume two classes for a scalar variable  $x$ . The class densities  $p(x|\omega_1), p(x|\omega_2)$  are gaussian such that both have mean value 0 but different standard deviations  $\sigma_1, \sigma_2$ . The prior probabilities are  $P(\omega_1), P(\omega_2)$ . Plot the densities. Derive the decision boundaries of the Bayes classifier.

3.

Assume 5 vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_5$ , whose mutual distances  $d(\mathbf{x}_i, \mathbf{x}_j) = D_{ij}$  are given as the following matrix

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

Perform hierarchical clustering for the vectors and plot the clustering tree, assuming that the distance between two clusters is equal to the distance between the two closest vectors in the two clusters. What is the best clustering into 3 clusters?

4.

When searching for relevant pages from the Web, a method based on hubs and authorities is sometimes used. Describe the basic principle of this method and give the formulas for computing the weights.

5.

Answer one of the following essay questions that are associated with the Matlab exercise:

A) Write an essay about basics of signal processing. Assume that a speech signal is read into Matlab. The signal contains a disturbing sine signal at frequency 2400 Hz. What kind of tools and visualizations for signal analysis do exist? What can be done in order to remove the noise signal?

B) Write an essay about eigenfaces. Assume that you have a set of digital grayscale photos of faces, which are aligned and taken in similar illumination conditions. What is the purpose of the principle component analysis (PCA), and show the phases of computation. How can PCA be used with face photos?

NOTE! Please give feedback on the courses at the Web address

<<http://www.cs.hut.fi/Opinnot/Palaute/kurssipalaute.html>>.