



UPPSALA UNIVERSITET

Signaler och System

TENTAMEN I DATORISERAD MÖNSTERIGENKÄNNING, 3P

12 DECEMBER 2002, 13.00-18.00

Examinator: Björn Hammarberg, tel 070 - 603 9623

Hjälpmedel: Mathematical Handbook, Mathematics Handbook (BETA), räknare, formelsamling från kursens hemsida

Betygsgränser: 3: 18p 4: 26p 5: 32p

**Observera att alla svar måste motiveras! Svar utan motivering ger poängavdrag.
Skriv bara på ena sidan av papperet!**

Lycka till!

1. (6p)

Tillämpningar inom datoriserad mönsterigenkänning kan designas med olika metoder, t ex prototypbaserade och linjära modellbaserade. Vilken metod som är lämpligast är olika från fall till fall men vissa generella drag skiljer dem åt.

- a) Vad är skillnaden mellan de två angivna metoderna? Fördelar/nackdelar?
- b) Nämn ett exempel på en prototypbaserad metod och förklara kort dess princip.
- c) Nämn ett exempel på en linjär modellbaserad metod och förklara kort dess princip.

2. (6p)

Du skall konstruera en klassificerare som kan skilja på objekt som härrör från två olika klasser som är lika vanliga. För detta ändamål har du tillgång till sensorer som mäter två värden per objekt. Du vet att mätsignalerna är normalfördelade och har följande väntevärden och kovariansmatriser

$$\mu_1 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mu_2 = \begin{pmatrix} 3 \\ -3 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$$

- a) Ange a priori-sannolikheten för respektive klass.
- b) Beräkna diskrimantfunktionen för respektive klass.
- c) Vilken ekvation ges den optimala beslutsytan av? Skriv på formen

$$\mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c = 0$$

Ledning: Vad gäller för diskriminantfunktionerna på denna yta?

- d) Beräkna den optimala beslutsytan och ange den explicit i de ingående variablerna (x_1 och x_2 eller x och y). Vilken geometrisk form har den? Rita figur där bl a nivåytor och beslutsytor ingår.

3. (4p)

För att visualisera data och/eller hantera stora datamängder i ett klassificeringsproblem kan man använda sig av klustring eller principalkomponentanalys (PCA).

- a) För resp metod antas olika saker om datamängden. Hur skiljer sig dessa antaganden åt?
- b) Vad är huvudprincipen med resp metod och vilka är ev nackdelar/risker?

4. (6p)

Beskriv de delsteg som krävs för en framgångsrik tillämpning av datoriserad mönsterigenkänning i praktiken. Motivera även betydelsen för varje delsteg.

5. (5p)

Ställ upp det problem som man vill lösa vid linjär regression.

- a) Förklara varför OLS-metoden inte alltid är pålitlig för att lösa detta problem.
- b) Ge två exempel på metoder som kan vara bättre lämpade att lösa problemet.

6. (6p)

Givet N st mätdata $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ med dimension d där $\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$.

- a) Härled en maximum likelihood-skattning (ML-skattning) av väntevärdet.
- b) Nämn två andra estimat av väntevärdet. Beskriv kortfattat hur dessa skiljer sig från ML-skattningen.
- c) Ange kort för- och nackdelar med de olika metoderna.

7. (4p)

Civ har fått i uppdrag att designa en klassificerare som skall skilja på två typer av objekt baserat på deras vikt. Civ har tillgång till viktfordelningsfunktionerna för respektive klass och vet även a priori-sannolikheterna. Hon har tyvärr missat att läsa kursen Datoriserad mönsterigenkänning och ber därför om din hjälp.

- a) Ange Bayes besluts Kriterium för optimal klassificering.
- b) Använd Bayes sats för att skriva om ovanstående uttryck uttryckt i fördelningsfunktionerna och a priori-sannolikheterna.
- c) I vilken mening är denna klassificerare optimal?