UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: Unik 4590 - Mønstergjenkjenning
Eksamensdag: 14.12.2010
Tid for eksamen: 4 timer
Oppgavesettet er på 3 sider
Vedlegg: -
Tillatte hjelpemidler: Håndskrevne og trykte hjelpemidler. Godkjent kalkulator.

Kontroller at oppgavesettet er komplett
før du begynner å besvare spørsmålene.

Oppgave 1: Innledning

Gi en kort besvarelse på følgende spørsmål:

a) Hva er formålet med faget mønstergjenkjenning?

b) Skisser et typisk klassifiseringssystem.

c) Hva menes med egenskapsuttrekking?

d) Hva ligger i begrepet trening av en klassifikator?

e) Hvordan kan man evaluere en klassifikator?

f) Hva er forskjellen på ledet og ikke-ledet læring?

Oppgave 2: Bayesisk desisjonsteori

I et todimensjonalt problem med tre klasser ω_1, ω_2 og ω_3 er klassene normalfordelte med felles kovariansmatrise gitt ved:

$$\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

Forventningsvektorene til de tre klassene er gitt ved: $\bar{\mu}_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}^t$, $\bar{\mu}_2 = \begin{bmatrix} 2 & 3 \end{bmatrix}^t$ og $\bar{\mu}_3 = \begin{bmatrix} -1 & 2 \end{bmatrix}^t$.
Klassene har like á priori sannsynligheter.

a) Lag diskriminantfunksjoner for dette klassifiseringsproblemet, basert på minimum feilrate-prinsippet.

b) Klassifiser objektet representert ved egenskapsvektoren $\bar{x} = \begin{bmatrix} 1.6 & 2.5 \end{bmatrix}^t$.

c) Hvilken form har desisjonsgrensene? Begrunn svaret.

d) Skisser forventningsvektorene og desisjonsgrensene.

e) Skisser kurvene med konstant Mahalanobisavstand omkring forventningsvektorene.

f) Anta at á priori sannsynligheten til klasse ω_2 er 2/3, mens de to andre klassene har lik á priori sannsynlighet på 1/6 hver. Hvordan blir desisjonsgrensene i dette tilfellet? Begrunn svaret.

Oppgave 3: Parameterestimering

Anta at tetthetsfunksjonen til en klasse er gitt ved Erlang-fordelingen:

$$p(x|\theta) = \theta^2 x \exp(-\theta x) u(x)$$

der θ er den ukjente parameteren og funksjonen $u(x)$ har verdien 1 for x>0 og verdien 0 for $x \leq 0$.
Parameteren skal estimeres ved bruk av ”Maximum likelihood” (ML) metoden basert på et

treningssett $x_1, x_2, ..., x_n$ med n én-dimensjonale sampler.

a) Gi en generell beskrivelse av ML-metoden.

b) Vis at ML-estimatet av den ukjente parameteren er gitt ved:

$$\hat{\theta} = \frac{2n}{\sum_{k=1}^{n} x_k}$$

Oppgave 4: Ikke-parametriske metoder

Anta at vi har et treningssett bestående av følgende todimensjonale sampler: $\overline{x}_1 = [1 \quad 1]^t$, $\overline{x}_2 = [2 \quad 1]^t$, $\overline{x}_3 = [0.5 \quad 1.5]^t$, $\overline{x}_4 = [2.5 \quad 0.5]^t$, $\overline{x}_5 = [1.5 \quad 0.5]^t$.

Samplene er hentet fra én og samme klasse, dvs. den klassen det skal foretas tetthetsestimering for.

a) Beskriv vindumetoden med utgangspunkt i en hyperkubisk vindufunksjon.

b) Lag en skisse av det todimensjonale egenskapsrommet med samplene ovenfor plottet inn. Tegn inn vindufunksjonene omkring samplene med side gitt ved *h=1*.

c) Finn tetthetsestimatet i punktet:

$$\overline{x}_0 = [0.7 \quad 1.3]^t$$

Bruk vindustørrelsen *h=1*.

d) Hva blir tetthetsestimatet dersom *h=0.5*?

e) Gi et eksempel på en generalisert vindufunksjon.

Oppgave 5: Klyngeanalyse

a) Hva er formålet med klyngeanalyse?

b) Gi en kort beskrivelse av hovedtyper av metoder.

c) Beskriv den *agglomerative (samlende)* hierarkiske metoden og gi et eksempel på et avstandsmål mellom klynger.

d) Skisser et *dendrogram* basert på det én-dimensjonale samplesettet:

$$\chi = \{1.51, \quad 1.70, \quad 2.05, \quad 3.30, \quad 3.45, \quad 3.66, \quad 3.80, \quad 4.80, \quad 4.97, \quad 5.10\}$$

Bruk avstandsmålet $d_{\min}(\chi_i, \chi_j)$, dvs. minimum avstand.

e) Bruk informasjonen i dendrogrammet til å angi et rimelig antall klynger.