UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: Unik4590 Mønstergjenkjenning
Eksamensdag: 06.12.2014
Tid for eksamen: 09:15 – 13:15
Oppgavesettet er på 4 sider
Vedlegg: 0
Tillatte hjelpemidler: Ingen trykte og håndskrevne hjelpemidler tillatt. Bestemt, enkel kalkulator tillatt.

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

Oppg. 1

Innledning

a) Forklar hva som menes med begrepene *egenskapsuttrekking* og *klassifisering* og redegjør for hvordan disse prosessene inngår i et typisk mønstergjenkjenningssystem.

b) Forklar hva som menes med begrepene *ledet læring* og *ikke-ledet læring*.

c) Gi en kortfattet beskrivelse av hvilke størrelser og begreper som inngår i Bayesisk beslutningsteori (desisjonsteori).

d) Forklar hva som menes med klassifiseringsprinsippene *minimum feilrate* og *minimum risk*.

Oppg. 2

Beslutningsteori

a) *Betinget risk* for en gitt handling α_i kan uttrykkes ved hjelp av kostnader og á posteriori sannsynligheter ved uttrykket:

$$R(\alpha_i|\vec{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\vec{x}), \quad i = 1,\ldots,a$$

der a er antall mulige handlinger og $\lambda(\alpha_i|\omega_j) = \lambda_{ij}$ er kostnaden (tapet) forbundet med handling α_i når sann klasse er ω_j. Forklar hvilket valg av handling som leder til minimum *total risk* (minimum forventet tap) og formulér den tilhørende beslutningsregelen.

b) Vis hvordan et spesielt valg av kostnader leder til *minimum feilrate* klassifisering og formulér beslutningsregelen i dette tilfellet. Anta at $a = c$ (antall handlinger lik antall klasser).

c) I et univariat (éndimensjonalt) klassifiseringsproblem med to klasser er tetthetsfunksjonene for de to klassene gitt ved normalfordelingene:

$$p(x|\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right], \ i = 1,\ 2$$

Konstruér en minimum feilrate beslutningsregel for det tilfellet at klassene har like á priori sannsynligheter, dvs. $P(\omega_1) = P(\omega_2) = 1/2$.

d) Hva blir beslutningsregelen dersom standardavvikene til fordelingene er like, dvs. $\sigma_1 = \sigma_2 = \sigma$, og hva blir desisjonsgrensen mellom klassene i dette tilfellet?

Oppg. 3

Parametriske metoder

a) Beskriv maksimum likelihood (ML) metoden for estimering av parametre i en antatt tetthetsfunksjon ved ledet læring.

b) Finn ML-estimatet av parameteren θ i den univariate eksponentialfordelingen gitt ved:

$$p(x|\theta) = \begin{cases} \theta\, \mathrm{e}^{-\theta x} & x \geq 0 \\ 0 & \textit{ellers} \end{cases}$$

der $\theta > 0$ og treningssettet er gitt ved $\mathscr{X} = \{x_1, ..., x_n\}$.

c) Anta et toklasseproblem der klassene er univariat eksponentialfordelte (som i deloppgave 3b) med parameterestimatene $\hat{\theta}_1$ og $\hat{\theta}_2$ for de to klassene. Utled en toklasse diskriminantfunksjon for dette problemet uttrykt ved disse parameterverdiene og á priorisannsynlighetene $P(\omega_1)$ og $P(\omega_2)$ for de to klassene.

Oppg. 4

Ikke-parametriske metoder

a) Forklar hva som skiller ikke-parametriske metoder fra parametriske metoder.

b) Sett opp et estimat for tettheten i et vilkårlig punkt $\vec{x}$ i egenskapsrommet basert på et treningssett med n sampler, og gjør rede for hvordan dette estimatet kan brukes til klassifisering av ukjente sampler.

c) Beskriv *nærmeste-nabo regelen* (NNR) og *k-nærmeste-nabo regelen* (k-NNR).

d) For et todimensjonalt klassifiseringsproblem med to klasser ω_1 og ω_2 er treningssettet gitt ved:

$$\mathscr{X}_1 = \left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 6 \\ 2 \end{bmatrix} \right\} \quad \text{(sampler fra } \omega_1\text{)}$$

og

$$\mathscr{X}_2 = \left\{ \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \begin{bmatrix} 5 \\ 7 \end{bmatrix}, \begin{bmatrix} 6 \\ 5 \end{bmatrix}, \begin{bmatrix} 7 \\ 1 \end{bmatrix}, \begin{bmatrix} 7 \\ 3 \end{bmatrix}, \begin{bmatrix} 8 \\ 6 \end{bmatrix}, \begin{bmatrix} 9 \\ 2 \end{bmatrix} \right\} \quad \text{(sampler fra } \omega_2\text{).}$$

Klassifisér det ukjente samplet gitt ved egenskapsvektoren $[4,5]^t$ med NNR og k-NNR når $k = 3$.

e) Hva skjer med det ikke-parametriske tetthetsestimatet (se deloppgave 4b) i det asymptotiske tilfellet (når $n \to \infty$)?

Oppg. 5

Trening av lineær diskriminantfunksjon

a) Sett opp en toklasse lineær diskriminantfunksjon uttrykt ved *utvidet vektvektor* og *utvidet egenskapsvektor* og forklar hva som inngår i disse vektorene.

b) Forklar kort hvilke metoder som kan brukes til å komme frem til en *løsningsvektor* og et tilhørende separerende hyperplan for et lineært separabelt sett.

c) Beskriv "*Fast inkrement regelen*" og angi betingelsen for konvergens til en løsningsvektor.

d) La treningssettet i et éndimensjonalt toklasseproblem bestå av samplene $\mathscr{X} = \{1, 2, 6, 8\}$ der de to første samplene er fra klasse ω_1 og de to siste fra ω_2. Bruk fast inkrement regelen til å finne en løsningsvektor. Sett startvektoren til $\vec{a}_0 = [0, 0]^t$.