NTNU
Fakultet for informasjonsteknologi, matematikk og elektroteknikk
Institutt for teknisk kybernetikk

Faglig kontakt under eksamen:
Idar Dyrdal
Tlf: 99579753

EKSAMEN I TTK4205 MØNSTERGJENKJENNING

Onsdag 7. desember 2011
Tid: 09.00-12.00

Sensur vil foreligge innen tre uker.
Hjelpemiddelkombinasjon D: Kalkulator med tomt minne tillatt. Ingen trykte eller håndskrevne hjelpemidler tillatt.

Oppgave 1: Innledning

a) Gi en kort beskrivelse av formålet med mønstergjenkjenning.

b) Hva menes med klassifisering?

c) Hvilke komponenter består et typisk mønstergjenkjenningssystem av?

Oppgave 2: Bayesisk desisjonsteori

a) Forklar begrepene klassebetinget tetthetsfunksjon, á priori sannsynlighet og á posteriori sannsynlighet, og sett opp et uttrykk for sammenhengen mellom dem (Bayes regel).

b) Forklar begrepene handlinger (actions) og kostfunksjoner.

c) Sett opp et uttrykk for betinget risk og begrunn hvorfor man ønsker å velge den handling som minimaliserer denne.

d) Utlede minimum feilrateprinsippet ved å finne et uttrykk for betinget risk, innsatt en null-én kostfunksjon (null kostnad for riktig klassifisering og lik kostnad for alle typer feilklassifiseringer).

Oppgave 3: Diskriminantfunksjoner

a) Definer begrepet diskriminantfunksjoner og forklar hvordan disse brukes.

b) Utled en diskriminantfunksjon for minimum feilrate klassifikatoren for to klasser uttrykt ved tetthetsfunksjonene og á priori sannsynlighetene for de to klassene.

c) I et todimensjonalt klassifiseringsproblem med to klasser, er klassene multivariat normalfordelte med tetthetsfunksjonene:

$$p(\bar{x}|\omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(\bar{x}-\bar{\mu}_i)^t \Sigma_i^{-1}(\bar{x}-\bar{\mu}_i)\right] \quad, i=1,2 .$$

Klassene har like kovariansmatriser gitt ved:

$$\Sigma_1 = \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad ,$$

mens forventningsvektorene til klassene er gitt ved henholdsvis:

$$\bar{\mu}_1 = [1.0 , 3.0]^t \quad \text{og} \quad \bar{\mu}_2 = [3.0 , 2.0]^t .$$

Anta at klassene har like á priori sannsynligheter, slik at $P(\omega_1) = P(\omega_2) = \frac{1}{2}$.

Lag diskriminantfunksjoner for dette klassifiseringsproblemet, basert på minimum feilrateprinsippet.

d) Klassifisér objektet representert ved egenskapsvektoren: $\bar{x}_0=[2.0, 1.0]^t$.

e) Hvilken form har desisjonsflaten? Begrunn svaret.

f) Hva skjer med desisjonsflaten dersom $P(\omega_1)>P(\omega_2)$?

g) Hva er vinkelen mellom desisjonsflaten og linjen mellom forventningsvektorene?

Oppgave 4: Parametriske metoder

a) Beskriv maksimum likelihood metoden for estimering av parametre i en antatt tetthetsfunksjon.

b) Finn maksimum likelihood estimatet av parameteren θ i den éndimensjonale Rayleigh-fordelingen gitt ved:

$$p(x|\theta)=\begin{cases} 2\theta x \mathrm{e}^{-\theta x^2} & x \geqslant 0 \\ 0 & ellers \end{cases}$$

der $\theta>0$ og treningssettet er gitt ved $\chi=\{x_{1,} \ldots, x_n\}$.

Oppgave 5: Ikke-parametriske metoder

a) Beskriv kort prinsippene for tetthetsestimering, og sett opp et estimat for tettheten i et vilkårlig punkt $\bar{x}$.

b) Beskriv nærmeste-nabo regelen (NNR) og k-nærmeste-nabo regelen (k-NNR).

c) Angi en øvre grense for den asymptotiske feilraten til NNR som funksjon av den optimale (Bayesiske) feilraten. Lag en figur. Hva skjer med den øvre grensen for k-NNR når $k \rightarrow \infty$.

d) Treningssettet for et éndimensjonalt toklasseproblem består av 5 sampler fra hver klasse, der:

$\chi_1=\{0.1, 0.9, 1.3, 1.5, 1.8\}$ er samplesettet fra klasse ω_1 og
$\chi_2=\{1.6, 1.9, 2.1, 2.3, 2.5\}$ er samplesettet fra klasse ω_2.

Klassifisér det ukjente sampelet x=1.65 med NNR. Hva blir resultatet med k-NNR med k=3?