UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: TEK5020 - Mønstergjenkjenning
Eksamensdag: 14.12.2018
Tid for eksamen: 09:15 – 13:15
Oppgavesettet er på 3 sider
Vedlegg: Ingen
Tillatte hjelpemidler: Ingen trykte eller håndskrevne hjelpemidler tillatt. Bestemt, enkel kalkulator tillatt.

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

Oppgave 1 - Innledning

a) Gjør rede for gangen i et typisk mønstergjenkjenningssystem, og forklar hva som menes med en *klassifikator*.

b) Nevn noen eksempler på hvordan man kan trene opp en klassifikator.

c) Forklar hvordan man kan estimere feilraten til en ferdig opptrent klassifikator.

d) Hva menes med at en klassifikator er "overtrent"? Hvordan kan man gå frem for å unngå dette?

Oppgave 2 - Bayesisk beslutningsteori

a) Forklar hva som menes med klassebetinget tetthetsfunksjon, a priori sannsynlighet og a posteriori sannsynlighet, og sett opp Bayes regel som knytter disse størrelsene sammen.

b) Gjør rede for minimum feilrateprinsippet, og formulér en generell beslutningsregel basert på dette. Lag en figur som viser univariate (éndimensjonale) tetthetsfunksjoner for to klasser, og markér de optimale (minimum feilrate) desisjonsregionene dersom klassenes a priori sannsynligheter er like.

c) I et klassifiseringsproblem med to klasser er de klassebetingede tetthetsfunksjonene gitt ved univariate normalfordelinger

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

der forventningsverdiene for klassene er henholdsvis $\mu_1 = 1$ og $\mu_2 = 2$, og standardavvikene er $\sigma_1 = 0,5$ og $\sigma_2 = 1$. Anta videre at klassenes a priori sannsynligheter er like, dvs. $P(\omega_1) = P(\omega_2) = 0,5$. Finn de optimale desisjonsgrensene (tersklene) mellom klassene, og lag en figur som viser desisjonsregionene i dette tilfellet.

d) Forklar kort hva som menes med handlinger (actions) og kostnader (tap) knyttet til ulike handlinger $\alpha_1, \ldots, \alpha_a$, og sett opp et uttrykk for den betingede risken $R(\alpha_i|\boldsymbol{x})$ knyttet til handling α_i for et objekt med målt egenskapsvektor $\boldsymbol{x}$. Hviken handling skal man velge for å minimalisere den totale risken?

Oppgave 3 - Parametriske metoder

a) Beskriv maksimum likelihoodmetoden (ML-metoden) for estimering av parametrene i en gitt tetthetsfunksjon, ved ledet læring.

b) Bruk ML-metoden til å finne den ukjente parameteren θ i Rayleigh-fordelingen

$$p(x|\theta) = \begin{cases} 2\theta x e^{-\theta x^2} & x \geq 0 \\ 0 & \text{ellers} \end{cases}, \quad \text{der } \theta > 0.$$

La estimatet være basert på et treningssett $\mathscr{X} = \{x_1, \ldots, x_n\}$ med n univariate sampler trukket fra fordelingen.

c) Anta et univariat toklasseproblem med Rayleighfordelte klasser (som i punkt b), der parameterestimatene for de to klassene er henholdsvis $\hat{\theta}_1$ og $\hat{\theta}_2$. Anta videre at klassene har like a priori sannsynligheter. Utled en beslutningsregel for dette klassifiseringsproblemet, basert på minimum feilrateprinsippet. Lag en figur som viser fordelingene og desisjonsgrensen (terskelen) mellom klassene.

Oppgave 4 - Ikke-parametriske metoder

a) Beskriv prinsippet for ikke-parametrisk tetthetsestimering, og sett opp et estimat for tettheten i et punkt $\boldsymbol{x}$ i egenskapsrommet, basert på et treningssett med n sampler fra den klassen man ser på. Lag en figur som illustrerer prinsippet.

b) Utled et estimat for a posteriori sannsynlighet for en vilkårlig klasse i et problem med c klasser, med utgangspunkt i tetthetsestimatet fra punkt a.

c) Hvike enkle beslutningsregler leder dette til? Formulér disse reglene.

d) Gi en kort redegjørelse for hvorfor nærmeste naboregelen (NNR) kan være nyttig å bruke til utprøving av egenskapskombinasjoner.

e) Hvorfor kan det være fornuftig å bruke f.eks. parametriske metoder til å trene opp en klassifikator for praktisk bruk?

Oppgave 5 - Diskriminantfunksjoner

a) Forklar hva som menes med diskriminantfunksjoner, og gjør rede for hvordan de brukes til klassifisering av objekter.

b) Sett opp en lineær diskriminantfunksjon for to klasser, forklar størrelsene som inngår og formulér beslutningsregelen for dette tilfellet.

c) Vis hvordan diskriminantfunksjonen for toklasseproblemet (fra punkt b) kan skrives om til et produkt av en *utvidet vektvektor* $\boldsymbol{a}$ og en *utvidet egenskapsvektor* $\boldsymbol{y}$.

d) Gjør rede for minste kvadraters metode til trening av den utvidede vektvektoren, og vis hvordan man kan komme frem til en minste kvadraters løsning ved hjelp av *Pseudoinvers*-metoden.

e) La treningssettet bestå av de univariate samplene -1 og -2 fra ω_1 og 1 og 2 fra ω_2. Finn vektvektoren ved å beregne den pseudoinverse til datamatrisen Y. La marginvektoren $\boldsymbol{b}$ ha kun enere som komponenter. Hvor finner man desisjonsgrensen i dette tilfellet?