UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: TEK5020 - Mønstergjenkjenning
Eksamensdag: 6.12.2021
Tid for eksamen: 09:15 – 13:15
Oppgavesettet er på 4 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 begrepene *klassebetinget sannsynlighetstetthet*, *a priori sannsynlighet* og *a posteriori sannsynlighet*, og sett opp *Bayes regel* (Bayes formel) som knytter disse størrelsene sammen.

b) Forklar hva som menes med *minimum-feilrateprinsippet* for valg av klasse, og formulér dette som en generell beslutningsregel.

c) Forklar hva en *klassifikator* er og hva som typisk er inngangsdata til denne.

d) Lag en figur som viser trinnene i et typisk klassifiseringssystem, fra rådata (målinger) til klassifiseringsresultat.

Oppgave 2

Beslutningsteori

a) Den multivariate normalfordelingen er gitt ved

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right].$$

Forklar størrelsene som inngår. Hva er parametrene i denne fordelingen?

b) I et todimensjonalt problem med tre klasser er klassene multivariat normalfordelte med felles kovariansmatrise

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

Forventningsvektorene for hver klasse er

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \boldsymbol{\mu}_2 = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \text{ og } \boldsymbol{\mu}_3 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}.$$

Klassifisér egenskapsvektoren

$$\boldsymbol{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

i henhold til minimum-feilrateprinsippet, dersom klassenes a priori sannsynligheter er like.

c) Lag en figur som viser forventningsvektorene og punktet $\boldsymbol{x}_0$ i egenskapsrommet.

d) Hvilken form har beslutningsgrensene mellom klassene i dette tilfellet? Forklar hvorfor. Skissér beslutningsgrensene i figuren.

Oppgave 3

Parametriske metoder

a) Gjør rede for *maksimum-likelihoodmetoden* for estimering av parametervektoren $\boldsymbol{\theta}$ i en antatt fordelingsfunksjon $p(\boldsymbol{x}|\boldsymbol{\theta})$ ved ledet læring.

b) Sett opp *likelihoodfunksjonen* $p(\mathscr{X}|\boldsymbol{\theta})$ og utled et likningssystem for estimatet av $\boldsymbol{\theta}$ basert på treningssettet $\mathscr{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ trukket fra den aktuelle fordelingsfunksjonen. Hvilken forutsetning må man gjøre om disse samplene?

c) Finn maksimum-likelihood estimatet av parameteren θ i den univariate fordelingen gitt ved

$$p(x|\theta) = \frac{1}{6}\theta^4 x^3 \mathrm{e}^{-\theta x},$$

der $x \geq 0$ og $\theta > 0$. La treningssettet være gitt ved $\mathscr{X} = \{x_{1,\ldots,}x_n\}$.

Oppgave 4

Lineære diskriminantfunksjoner

a) Sett opp en lineær diskriminantfunksjon $g(\boldsymbol{x})$ for et toklasseproblem, og vis hvordan denne kan omskrives til *utvidet* form, som indreproduktet av en utvidet vektvektor $\boldsymbol{a}$ med en utvidet egenskapsvektor $\boldsymbol{y}$. Anta at det opprinnelige egenskapsrommet er av dimensjon d. Hvilken dimensjon har det utvidede egenskapsrommet?

b) Beskriv hvordan man kan trene opp vektvektoren $\boldsymbol{a}$ ved gradientsøk, basert på et treningssett av egenskapsvektorer. Hva menes med en *kriteriefunksjon*? Hva menes med en *løsningsvektor*?

c) Gjør rede for *fast-inkrementregelen* for trening av vektvektoren $\boldsymbol{a}$ ved hjelp av treningssettet $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$. Hva er forutsetningen for at denne algoritmen skal lede til en løsningsvektor etter et endelig antall iterasjoner?

d) Anta et univariat treningssett som består av samplene $1, 2, 6, 7$ der de to første er fra klasse ω_1 og de to siste fra ω_2. Skriv disse samplene på utvidet form (husk fortegnskonvensjonen) og bruk fast-inkrementregelen til å finne en løsningsvektor. Sett startvektoren til $\boldsymbol{a}_0 = [0, 0]^t$.

e) Sett løsningsvektoren inn i diskriminantfunksjonen, og finn beslutningsgrensen (terskelen) mellom klassene i det opprinnelige egenskapsrommet.

Oppgave 5

Ikke-parametriske metoder

a) Gjør rede for hva som skiller ikke-parametriske metoder fra parametriske metoder, og sett opp et uttrykk for tetthetsestimatet i et punkt $\boldsymbol{x}$ basert på et treningssett av egenskapsvektorer $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$.

b) Sett dette estimatet inn i Bayes formel for å utlede et estimat av a posteri sannsynlighet for hver av klassene i et problem med c klasser.

c) Forklar hvordan dette estimatet leder til *k-nærmeste-naboregelen*, og formulér denne beslutningsregelen.

d) Formulér spesialtilfellet *nærmeste-naboregelen* og angi en øvre grense for den asymtotiske feilraten som funksjon av den optimale feilraten.