Løsningsforslag eksamen TEK9020 - 2020H

Oppgave 1

Innledning

a) Her kan det lages en skisse i likhet med figur 3 i forelesningsnotatene, som beskriver sensor (kamera), egenskapsuttrekker og klassifikator, med litt forklaring av disse begrepene (se side 7 i forelesningsnotatene).

b) Et treningssett er en samling av egenskapsvektorer fra klassene som inngår i problemet. Vanligvis er klassetilhørigheten til samplene i treningssettet kjent (ledet læring, se avsnitt 1.3 i forelesningsnotatene), men den kan også være ukjent (ikke-ledet læring). Som eksempler på metoder kan nevnes parametriske metoder, ikke-parametriske metoder og trening av lineære diskriminantfunksjoner, med kort forklaring av prinsippene.

c) Kort forklaring, som i avsnitt 2.5 i forelesningsnotatene (side 19-20), med et par eksempler på mulige diskriminantfunksjoner (side 20).

d) Bruk av et uavhengig testsett er viktig for å få et realistisk estimat av ytelsen til en klassifikator. Dette gjør det mulig å avdekke eventuell overtrening av klassifikatoren (spesialisering til treningssettet) og ellers forsikre seg om at klassifikatoren kan generalisere til nye data. Mulige fremgangsmåter er beskrevet i avsnitt 6.1.2 i forelesningsnotatene.

Oppgave 2

Beslutningsteori

a) Se avsnitt 1.4.1 og 2.1 (til og med Bayes regel) i forelesningsnotatene.

b) Se avsnitt 2.1 i forelesningsnotatene.

c) Betinget risk (forventet tap) er kostnaden forbundet ved en gitt handling, gitt en måling (dvs. egenskapsvektoren $\boldsymbol{x}$ for et ukjent objekt):

$$R(\alpha_i|\boldsymbol{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\boldsymbol{x}), \text{ i=1,...,a.}$$

Total risk er gitt ved

$$R = \int_{\mathbb{R}^d} R(\alpha(\boldsymbol{x})|\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \tag{1}$$

for en gitt desisjonsfunksjon $\alpha(\boldsymbol{x})$ med utfallene $\alpha_1, \alpha_2, \ldots, \alpha_a$. Den totale risken skal minimaliseres ved å velge α_i slik at den betingede risken $R(\alpha(\boldsymbol{x})|\boldsymbol{x})$ er minimum for enhver $\boldsymbol{x}$. Dette leder til Bayes desisjonsregel, som kan skrives som

$$\text{Velg } \alpha_m \text{ hvis } R(\alpha_m|\boldsymbol{x}) \leq R(\alpha_j|\boldsymbol{x}),\ j = 1, \ldots, a. \tag{2}$$

Utfallet av desisjonsfunksjonen er da α_m, dvs. $\alpha(\boldsymbol{x}) = \alpha_m$. Dette er den handling som gir minimum betinget risk, og samtidig minimum total risk (minimum av integralet i likning 1) og kalles derfor *minimum risk klassifisering*.

d) Den betingede risken forbundet med hver handling er her

$$\begin{aligned} R(\alpha_1|x) &= \lambda_{11}P(\omega_1|x) + \lambda_{12}P(\omega_2|x) \\ R(\alpha_2|x) &= \lambda_{21}P(\omega_1|x) + \lambda_{22}P(\omega_2|x). \end{aligned}$$

Desisjonsgrensen er da gitt ved

$$\begin{gathered} R(\alpha_1|x) = R(\alpha_2|x) \\ \Downarrow \\ (\lambda_{11} - \lambda_{21})P(\omega_1|x) = (\lambda_{22} - \lambda_{12})P(\omega_2|x). \end{gathered}$$

Innsatt de univariat normalfordelte klasser i denne oppgaven blir dette

$$\frac{\lambda_{11} - \lambda_{21}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2\right] P(\omega_1) = \frac{\lambda_{22} - \lambda_{12}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2\right] P(\omega_2),$$

der nevneren i Bayes formel er strøket på begge sider av likhetstegnet. Ved å ta logaritmen på begge sider og multiplisere ut kvadratuttrykkene i eksponenten, får man da

$$\begin{gathered} -\frac{1}{2\sigma^2}(x^2 - 2\mu_1 x + \mu_1^2) + \ln[(\lambda_{11} - \lambda_{21})P(\omega_1)] = -\frac{1}{2\sigma^2}(x^2 - 2\mu_2 x + \mu_2^2) + \ln[(\lambda_{22} - \lambda_{12})P(\omega_2)] \\ \Downarrow \\ \frac{1}{2\sigma^2}(2\mu_1 x - \mu_1^2) + \ln[(\lambda_{11} - \lambda_{21})P(\omega_1)] = \frac{1}{2\sigma^2}(2\mu_2 x - \mu_2^2) + \ln[(\lambda_{22} - \lambda_{12})P(\omega_2)] \\ \Downarrow \\ \frac{1}{2\sigma^2}(2\mu_1 x - \mu_1^2) - \frac{1}{2\sigma^2}(2\mu_2 x - \mu_2^2) = \ln[(\lambda_{22} - \lambda_{12})P(\omega_2)] - \ln[(\lambda_{11} - \lambda_{21})P(\omega_1)] \\ \Downarrow \\ \frac{1}{2\sigma^2}(2\mu_1 x - \mu_1^2) - \frac{1}{2\sigma^2}(2\mu_2 x - \mu_2^2) = \ln\frac{(\lambda_{22} - \lambda_{12})P(\omega_2)}{(\lambda_{11} - \lambda_{21})P(\omega_1)} \\ \Downarrow \\ 2(\mu_1 - \mu_2)x - (\mu_1^2 - \mu_2^2) = 2\sigma^2 \ln\frac{(\lambda_{12} - \lambda_{22})P(\omega_2)}{(\lambda_{21} - \lambda_{11})P(\omega_1)} \end{gathered}$$

som kan løses mht. x, slik at terskelen blir

$$x_0 = \underline{\underline{\frac{\mu_1 + \mu_2}{2} + \frac{\sigma^2}{\mu_1 - \mu_2} \ln\left[\frac{(\lambda_{12} - \lambda_{22})P(\omega_2)}{(\lambda_{21} - \lambda_{11})P(\omega_1)}\right]}}.$$

e) Se avsnitt 2.3 i forelesningsnotatene.

f) I dette tilfellet er $\lambda_{11} = \lambda_{22} = 0$ og $\lambda_{12} = \lambda_{21} = 1$. Terskelen forenkles da til

$$\underline{\underline{x_0 = \frac{\mu_1 + \mu_2}{2} + \frac{\sigma^2}{\mu_1 - \mu_2} \ln \frac{P(\omega_2)}{P(\omega_1)}}}.$$

Med like a priori sannsynligheter faller det siste leddet bort, og x_0 ligger da midt imellom μ_1 og μ_2.

Skissen av fordelingene, med feilrate og desisjonsgrense blir tilsvarende figur 12 (på side 15) i forelesningsnotatene.

Oppgave 3

Parametriske metoder

a) Se avsnitt 3.1 på side 34 i forelesningsnotatene. Det forutsettes at samplene i treningssettet er innbyrdes uavhengige.

b) Med fordelingen $p(x|\theta)$ i oppgaveteksten blir likelihoodfunksjonen

$$P(\mathscr{X}|\theta) = \prod_{k=1}^{n} \left[\frac{1}{2} \theta^3 x_k^2 \, \mathrm{e}^{-\theta x_k} \right].$$

Log-likelihoodfunksjonen blir da

$$\mathscr{L}(\theta) = \sum_{k=1}^{n} \ln \left[\frac{1}{2} \theta^3 x_k^2 \, \mathrm{e}^{-\theta x_k} \right] = \sum_{k=1}^{n} \left[3 \ln \theta - \ln 2 + 2 \ln x_k - \theta x_k \right].$$

Maksimum-likelihood estimatet av θ finnes ved å sette den deriverte av log-likelihoodfunksjonen til null, dvs.

$$\frac{d\mathscr{L}(\theta)}{d\theta} = \sum_{k=1}^{n} \left[\frac{3}{\theta} - x_k \right] = 0,$$

som medfører at

$$\frac{3n}{\theta} = \sum_{k=1}^{n} x_k.$$

Estimatet blir da

$$\underline{\underline{\hat{\theta} = \frac{3n}{\sum_{k=1}^{n} x_k}}}.$$

Oppgave 4

Lineære diskriminantfunksjoner

a) Se avsnitt 5.1.1 på side 65 i forelesningsnotatene.

b) Se avsnitt 5.3 nederst på side 69 i forelesningsnotatene. Det utvidede egenskapsrommet har dimensjonen $d+1$.

c) Se avsnitt 5.7 og underavsnitt 5.7.1 i forelesningsnotatene.

d) Her blir datamatrisen

$$Y = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ -1 & -5 \\ -1 & -6 \\ -1 & -7 \end{bmatrix}.$$

Legg merke til at fortegnet på samplene fra klasse ω_2 er snudd (alternativt kan man snu fortegnet på de tre siste elementene i marginvektoren). Den pseudoinverse til Y kan beregnes f.eks. ved hjelp av Matlab, og blir

$$Y^\dagger = (Y^tY)^{-1}Y^t = \begin{bmatrix} 0.5952 & 0.4524 & 0.3095 & -0.0238 & 0.1190 & 0.2619 \\ -0.1071 & -0.0714 & -0.0357 & -0.0357 & -0.0714 & -0.1071 \end{bmatrix}.$$

Med marginvektoren $\boldsymbol{b} = [1,1,1,1,1,1]^t$ blir vektvektoren

$$\boldsymbol{a} = Y^\dagger \boldsymbol{b} = \begin{bmatrix} 1.7143 \\ -0.4286 \end{bmatrix}.$$

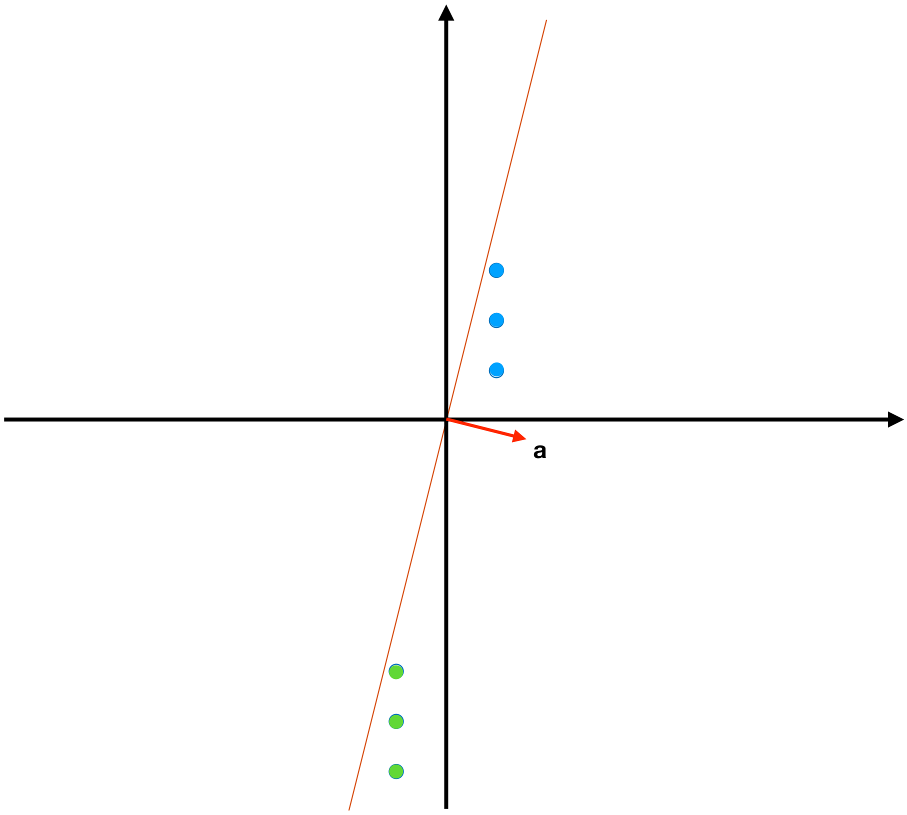
To-klasse diskriminantfunksjonen kan da skrives som $g(x) = 1.7143 - 0.4284x$. Terskelen mellom klassene er gitt ved nullpunktet for denne funksjonen, dvs.

$$x_0 = \frac{1.7143}{0.4284} = \underline{\underline{4}}.$$

e) Figuren på neste side viser treningssamplene, desisjonsgrensen og vektvektoren i det utvidede egenskapsrommet.

Samplene fra ω_1 ligger langs linjen $y_1 = 1$, mens fra samplene fra ω_2 ligger langs $y_1 = -1$ (fortegnet er snudd i henhold til fortegnskonvensjonen).

Alle sampler ligger på den positive siden av hyperplanet (linjen gjennom origo) og vektvektoren $\boldsymbol{a}$ peker inn i desisjonsregionen for ω_1, siden vi med fortegnskonvensjonen ønsker å få alle sampler på den positive siden av hyperplanet. Vektvektoren står normalt på desisjonsgrensen.



Oppgave 5

Ikke-ledet læring

a) Ikke-ledet læring går ut på å trene en klassifikator ved hjelp av et treningssett med *umerkede* sampler, dvs. klassetilhørigheten til treningssamplene er ukjent. En blandingstetthet er en vektet sum av de klassebetingede tetthetsfunksjonene, der vektene er klassenes a priori sannsynligheter.

b) Blandingstettheten for et problem med to klasser kan skrives som

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{i=1}^{2} p(\boldsymbol{x}|\omega_i, \boldsymbol{\theta}_i)P(\omega_i) = p(\boldsymbol{x}|\omega_1, \boldsymbol{\theta}_1)P(\omega_1) + p(\boldsymbol{x}|\omega_2, \boldsymbol{\theta}_2)P(\omega_2).$$

Her er $p(\boldsymbol{x}|\omega_i, \boldsymbol{\theta}_i)$, $i = 1,2$ komponenttetthetene, mens $P(\omega_i)$ er blandingsparametrene.

c) Se utledningen av likningssystemet for c klasser i avsnitt 7.1 (side 97) i forelesningsnotatene.

d) Innsetting av normalfordelingene

$$p(x_k|\omega_i, \mu_i) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x - \mu_1)^2\right], \; i = 1,2$$

i likningssystemet

$$\sum_{k=1}^{n} P(\omega_i|x_k, \boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}_i} \ln p(x_k|\omega_i, \boldsymbol{\theta}_i) = 0, \; i = 1,2,$$

gir da

$$\sum_{k=1}^{n} P(\omega_i|x_k,\mu_1,\mu_2)\frac{\mathrm{d}}{\mathrm{d}\mu_i}\left[-\frac{1}{2}(x-\mu_i)^2 - \ln\sqrt{2\pi}\right] = 0,\ i=1,2,$$

som gir

$$\sum_{k=1}^{n} P(\omega_i|x_k,\mu_1,\mu_2)(x-\mu_i) = 0,\ i=1,2.$$

Forventningsestimatene kan da uttrykkes ved

$$\hat{\mu}_i = \frac{\sum_{k=1}^{n} P(\omega_i|x_k,\hat{\mu}_1,\hat{\mu}_2)x_k}{\sum_{k=1}^{n} P(\omega_i|x_k,\hat{\mu}_1,\hat{\mu}_2)},\ i=1,2.$$

der

$$\begin{aligned} P(\omega_i|x_k,\hat{\mu}_1,\hat{\mu}_2) &= \frac{p(x_k|\omega_i,\hat{\mu}_i)P(\omega_i)}{\sum_{j=1}^{2} p(x_k|\omega_j,\hat{\mu}_j)P(\omega_j)} \\ &= \frac{\exp\{-\frac{1}{2}(x_k-\hat{\mu}_i)^2\}}{\exp\{-\frac{1}{2}(x_k-\hat{\mu}_1)^2\}+\exp\{-\frac{1}{2}(x_k-\hat{\mu}_2)^2\}},\ i=1,2, \end{aligned}$$

etter innsetting av normalfordelingene og forkorting av like faktorer over og under brøkstreken. Her er $x_{k,k=1\ldots n}$ de umerkede samplene i treningssettet.

Dette er implisitte uttrykk som kan løses ved iterasjon, som beskrevet på side 98 i forelesningsnotatene. Man velger startverdier for μ_1 og μ_2 og oppdaterer deretter aposteriorisannsynlighetene og forventningsestimatene rekursivt.

Oppgave 6

Klyngeanalyse

a) Klyngeanalyse består i å dele et datasett inn i grupper (klynger), slik at sampler innen hver klynge er mest mulig like, mens det er størst mulig ulikhet mellom sampler i forskjellige klynger. Klyngeinndelingen er altså datadrevet, ved at samplenes (objektenes) egenskapsvektorer brukes direkte, og ikke basert på a priori kunnskap i form av f.eks. klassetilhørighet.

Klyngeanalyse brukes ofte til å kartlegge strukturen til ukjente data, f.eks. finne ut hvorvidt samplene i datasettet kan deles inn i et antall kompakte og godt adskilte klynger, om det kanskje består av langstrakte klynger eller har en mer komplisert struktur.

Hovedtyper av metoder som kan nevnes er

- Optimalisering av kriteriefunksjon,
- Hierarkiske metoder (agglomerative og divisive).

b) Den agglomerative (samlende) metoden består av følgende hovedtrinn:

1. Start med n klynger, dvs. $\mathscr{X}_i = \{\boldsymbol{x}_i\}, i = 1, \ldots, c; \hat{c} = n$
2. Finn nærmeste par av klynger, f.eks. $\mathscr{X}_i, \mathscr{X}_j$
3. Slå sammen disse klyngene; $\hat{c} \to \hat{c} - 1$
4. Gjenta trinn 2 og 3 inntil ønsket antall klynger $\hat{c} = c$ er funnet.

Et *dendrogram* gir en grafisk fremstilling av innbyrdes avstander mellom klyngedannelser på ulike nivåer.

c) Det univariate datasettet

$$\mathscr{C} = \{1.50, 1.70, 2.00, 2.10, 2.85, 3.20, 3.85, 4.00\}.$$

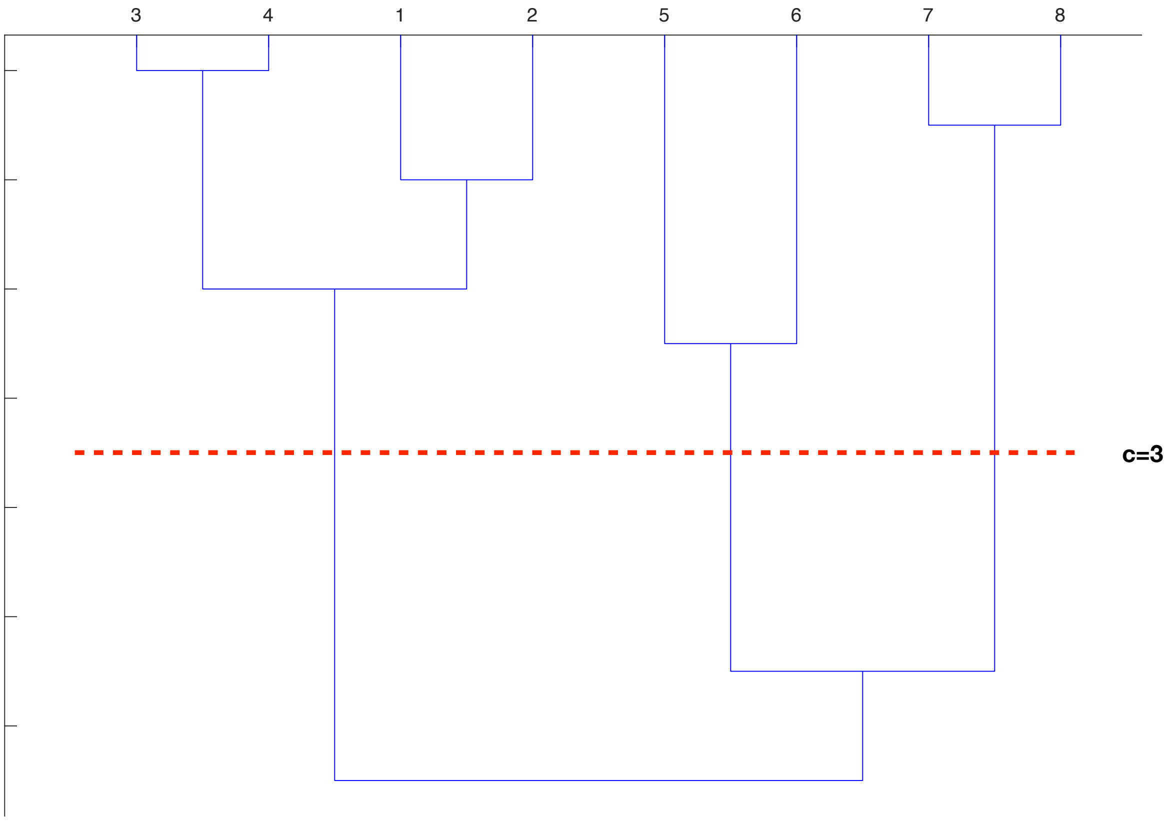
skal deles i *tre* klynger ved å bruke den agglomerative metoden med avstandsmålet d_{min}. Dette gir følgende sekvens av sammenslåinger:

- sample 3 og 4,
- sample 7 og 8,
- sample 1 og 2,
- samplene 1, 2, 3 og 4,
- sample 5 og 6

Resultatet blir de tre klyngene

- $\mathscr{C}_1 = \{1.50, 1.70, 2.00, 2.10\}$,
- $\mathscr{C}_2 = \{2.85, 3.20\}$,
- $\mathscr{C}_3 = \{3.85, 4.00\}$.

Løsningen er illustrert ved dendrogrammet i figuren på neste side. Samplene ligger på den horisontale aksen (tallverdiene er indeksene til samplene i datasettet), mens den vertikale aksen viser avstandsmålet mellom sampler/klynger.


3
4
1
2
5
6
7
8
c=3