TEK5020/9020 – Mønstergjenkjenning
Høsten 2023

Løsningsforslag – Øvingsoppgaver 1

Idar Dyrdal (idar.dyrdal@its.uio.no)

UiO : Institutt for teknologisystemer

23. august 2023

Oppgave 1

Kostfunksjonene er her gitt ved

$$\lambda\left(\alpha_i \mid \omega_j\right) = \begin{cases} 0 & i = j \\ \lambda_r & i = c+1 \\ \lambda_s & \text{ellers (dvs. } i \neq j) \end{cases}$$

Betinget risk er derved gitt ved

$$R\left(\alpha_{i, i=1,\dots,c} \mid \mathbf{x}\right) = \sum_{j=1}^{c} \lambda\left(\alpha_i \mid \omega_j\right) P\left(\omega_j \mid \mathbf{x}\right) = \sum_{\substack{j=1 \\ j \neq i}}^{c} \lambda_s P\left(\omega_j \mid \mathbf{x}\right) = \lambda_s\left(1 - P\left(\omega_i \mid \mathbf{x}\right)\right)$$

og

$$R\left(\alpha_{c+1} \mid \mathbf{x}\right) = \sum_{j=1}^{c} \lambda\left(\alpha_{c+1} \mid \omega_j\right) P\left(\omega_j \mid \mathbf{x}\right) = \sum_{j=1}^{c} \lambda_r P\left(\omega_j \mid \mathbf{x}\right) = \lambda_r \sum_{j=1}^{c} P\left(\omega_j \mid \mathbf{x}\right) = \lambda_r .$$

Oppgave 1 (forts.)

Vi skal alltid velge den handling som gir minimum betinget risk. Her betyr dette at vi skal velge den klassen ω_k som gir størst a posteriori sannsynlighet, forutsatt at

$$R(\alpha_k \mid \mathbf{x}) \leq R(\alpha_{c+1} \mid \mathbf{x}).$$

I motsatt fall skal objektet forkastes, siden det i så fall vil være denne handlingen som gir minst betinget risk.

Dette gir

$$\lambda_s(1 - P(\omega_k \mid \mathbf{x})) \leq \lambda_r \Rightarrow P(\omega_k \mid \mathbf{x}) \geq 1 - \frac{\lambda_r}{\lambda_s}$$

som betingelse for å klassifisere til en av klassene (i motsatt fall skal objektet forkastes).

Oppgave 1 (forts.)

Beslutningsregelen blir derved

$$\boxed{\text{Velg } \omega_k \text{ dersom } P(\omega_k \mid \boldsymbol{x}) = \max_j P(\omega_j \mid \boldsymbol{x}) \text{ og } P(\omega_k \mid \boldsymbol{x}) \geq 1 - \frac{\lambda_r}{\lambda_s}, \text{ ellers forkast.}}$$

Svar på tilleggsspørsmål:

- Dersom $\lambda_r = 0$ vil vi alltid forkaste, siden terskelen for å klassifisere aldri blir overskredet.
- Dersom $\lambda_r \geq \lambda_s$ vil vi derimot aldri forkaste. Det er i dette tilfellet mer kostbart å forkaste enn å klassifisere feil.

Oppgave 2

Her er $\sigma_{ij,i\neq j} = 0$ og $\sigma_{ii} = \sigma_i^2$ slik at

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & . & . & 0 \\ 0 & \sigma_2^2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 0 & \sigma_d^2 \end{bmatrix} \text{ som gir } \Sigma^{-1} = \begin{bmatrix} \sigma_1^{-2} & 0 & . & . & 0 \\ 0 & \sigma_2^{-2} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 0 & \sigma_d^{-2} \end{bmatrix} \text{ og } |\Sigma| = \prod_{i=1}^d \sigma_i^2$$

Fordelingen blir da

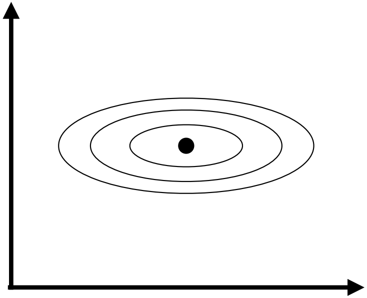
$$\begin{aligned} p(\boldsymbol{x}) &= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \\ &= \frac{1}{(2\pi)^{d/2}\prod_{i=1}^d \sigma_i} \exp\left\{-\frac{1}{2}\sum_{i=1}^d \left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right\}. \end{aligned}$$

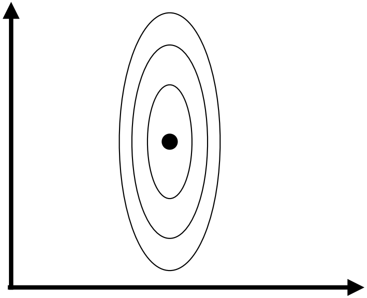
Oppgave 2 (forts.)

Mahalanobisavstanden er gitt ved

$$r^2 = (\boldsymbol{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) = \sum_{i=1}^{d} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2.$$

Kurvene gjennom punkter med konstant tetthet blir generelt hyperellipsoider med hovedaksene parallelt med aksene i egenskapsrommet.





Oppgave 3a

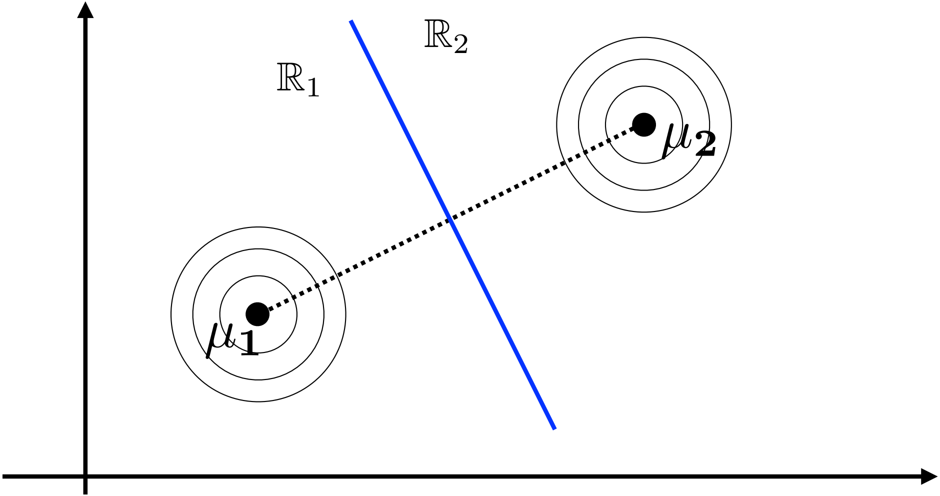
I denne oppgaven er

$$p(\mathbf{x} \mid \omega_i) \sim N\left(\boldsymbol{\mu}_i, \sigma^2 I\right), \text{ der } c = 2, \text{ og}$$
$$P(\omega_1) = P(\omega_2) = \frac{1}{2}.$$

Feilraten blir da

$$\begin{aligned} P_e &= P(\omega_1)\int_{\mathbb{R}_2} p(\mathbf{x}|\omega_1)\mathrm{d}\mathbf{x} + P(\omega_2)\int_{\mathbb{R}_1} p(\mathbf{x} \mid \omega_2)\mathrm{d}\mathbf{x} \\ &= \int_{\mathbb{R}_2} p(\mathbf{x} \mid \omega_1)\mathrm{d}\mathbf{x} \text{ (dette fordi } \Sigma_1 = \Sigma_2 = \Sigma = \sigma^2 I \text{ og } P(\omega_1) = P(\omega_2)) \\ &= \int_{\mathbb{R}_2} \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^t\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)}\mathrm{d}\mathbf{x}, \text{ der } \Sigma^{-1} = \frac{1}{\sigma^2} I. \end{aligned}$$

Oppgave 3a (forts.)

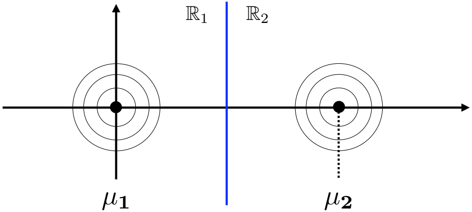


Oppgave 3a (forts.)

Roterer og translerer rommet slik at

$$\boldsymbol{\mu}_1 = (0, 0, \ldots, 0)^t$$
$$\boldsymbol{\mu}_2 = (\mu_2, 0, \ldots, 0)^t$$

(forenkling uten tap av generalitet).



Kan da skrive feilraten som

$$P_e = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}_2} e^{-\frac{(x_i - \mu_{1i})^2}{2\sigma^2}} \mathrm{d}x_i$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{\frac{\mu_2}{2}}^{\infty} e^{-\frac{x_1^2}{2\sigma^2}} dx_1 \underbrace{\prod_{i=2}^{d} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{x_i^2}{2\sigma^2}} \mathrm{d}x_i}_{=1} .$$

Oppgave 3a (forts.)

Her kan man skifte variabel til u, der

$$u = \frac{x_1}{\sigma} \quad \text{slik at } du = \frac{dx_1}{\sigma},$$

og definere

$$a = \frac{||\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1||}{2\sigma} = \frac{\mu_2}{2\sigma}.$$

Dette gir feilraten

$$P_e = \frac{1}{\sqrt{2\pi}} \int_a^{\infty} e^{-\frac{1}{2}u^2} \mathrm{d}u,$$

hvilket skulle bevises.

Oppgave 3b

Velger nå

$$\boldsymbol{\mu}_1 = (0,0,\ldots,0)^t$$
$$\boldsymbol{\mu}_2 = (\mu,\mu,\cdots,\mu)^t,$$

som gir

$$a = \frac{\|\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\|}{2\sigma} = \frac{\left(\sum_{i=1}^{d} \mu^2\right)^{1/2}}{2\sigma} = \frac{\mu\sqrt{d}}{2\sigma}.$$

Benytter ulikheten

$$P_e = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{1}{2}u^2} \mathrm{d}u \leqslant \frac{1}{\sqrt{2\pi}a} e^{-\frac{1}{2}a^2}.$$

Herav følger da

$$P_e = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{1}{2}u^2} \mathrm{d}u \leqslant \frac{2\sigma}{\sqrt{2\pi}\mu\sqrt{d}} e^{-\frac{1}{2}\cdot\frac{\mu^2 d}{4\sigma^2}} = \frac{2\sigma}{\sqrt{2\pi d}\mu} e^{-\frac{\mu^2 d}{8\sigma^2}} \longrightarrow 0 \text{ når } d \to \infty.$$

Dette viser at felraten avtar når antall uavhengige egenskaper øker.

Oppgave 4a

I denne oppgave er fordelingene til de to klassene

$$p(\mathbf{x} \mid w_i) \sim N(\boldsymbol{\mu}_i, \Sigma), \quad i = 1, 2, \quad \text{(like kovariansmatriser)}$$

der den kvadrerte Mahalanobisavstanden fra hvert klassemiddel blir

$$r_i^2 = (\mathbf{x} - \boldsymbol{\mu}_i)^t \, \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \, .$$

For å finne gradienten til r_i^2 kan vi bruke relasjonen:

$$\nabla \left(\mathbf{a}^t A \mathbf{a} \right) = 2A\mathbf{a} \quad \text{(kan vises)}$$
$$\Downarrow$$
$$\nabla r_i^2 = \nabla \{ \underbrace{(\mathbf{x} - \boldsymbol{\mu}_i)^t}_{\mathbf{a}^t} \underbrace{\Sigma^{-1}}_{A} \underbrace{(\mathbf{x} - \boldsymbol{\mu}_i)}_{\mathbf{a}} \}$$
$$= \underline{\underline{2\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)}}.$$

Oppgave 4b

La nå

$$\boldsymbol{x} = \boldsymbol{\mu}_i + \lambda \boldsymbol{a},$$

der λ er en skalar. Dette beskriver en linje gjennom $\boldsymbol{\mu}_i$ med retning gitt av $\boldsymbol{a}$.

Ved å sette inn for $\boldsymbol{x}$ i uttrykket for gradienten til den kvadrerte Mahalanobisavstanden i deloppgave 4a, får vi

$$\nabla r_i^2 = 2\Sigma^{-1}(\boldsymbol{\mu}_i + \lambda \boldsymbol{a} - \boldsymbol{\mu}_i) = 2\lambda\Sigma^{-1}\boldsymbol{a},$$

der $\Sigma^{-1}\boldsymbol{a}$ er en konstant vektor.

Dette viser at gradienten har samme *orientering* i alle punkt på linjen.

Oppgave 4c

Uttrykket

$$\boldsymbol{x} = \boldsymbol{\mu}_1 + \lambda\left(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\right), 0 < \lambda < 1,$$

beskriver linjen mellom $\boldsymbol{\mu}_1$ og $\boldsymbol{\mu}_2$.

Innsetting i resultatet fra deloppgave 4b gir gradientene

$$\begin{aligned}
\nabla r_1^2 &= 2\Sigma^{-1}\left(\lambda\left(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\right)\right) = 2\lambda\Sigma^{-1}\left(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\right) \text{ og} \\
\nabla r_2^2 &= 2\Sigma^{-1}\left(\boldsymbol{\mu}_1 + \lambda\boldsymbol{\mu}_2 - \lambda\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right) = 2(\lambda - 1)\Sigma^{-1}\left(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\right).
\end{aligned}$$

Vi ser av dette at ∇r_1^2 og ∇r_2^2 peker i motsatt retning siden λ og $\lambda - 1$ har motsatt fortegn.

Oppgave 4d

Her skal vi finne gradientene i skjæringspunktet $\boldsymbol{x}_0$ mellom det optimale hyperplanet og linjen mellom $\boldsymbol{\mu}_1$ og $\boldsymbol{\mu}_2$:

$$\boldsymbol{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) - \frac{\ln\{P(\omega_1)/P(\omega_2)\}}{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

$$\Downarrow$$

$$\nabla r_i^2 = \Sigma^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) - \frac{2\ln\{P(\omega_1)/P(\omega_2)\}}{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - 2\Sigma^{-1}\boldsymbol{\mu}_i$$

$$\Downarrow$$

$$\nabla r_1^2 = \Sigma^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) + \alpha\Sigma^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \sim \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \sim \boldsymbol{w}$$
$$\nabla r_2^2 = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \alpha\Sigma^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \sim \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \sim \boldsymbol{w}$$

$$\Downarrow$$

Hyperplanet er tangent til hyperellipsoidene.

Oppgave 5

Her er

$$\boldsymbol{x} = [x_1, \ldots, x_d]^t \text{ der } x_i = \begin{cases} 0 \\ 1 \end{cases} \text{ og innbyrdes uavhengige.}$$

Diskriminantfunksjonene kan da skrives som

$$\begin{aligned} g_j(\boldsymbol{x}) &= \ln P(\boldsymbol{x} \mid \omega_j) P(\omega_j), \quad j = 1, \cdots, c \\ &= \ln \left\{ \left[\prod_{i=1}^{d} p_{ij}^{x_i} (1 - p_{ij})^{1-x_i} \right] P(\omega_j) \right\} \\ &= \sum_{i=1}^{d} \ln \left[p_{ij}^{x_i} (1 - p_{ij})^{1-x_i} \right] + \ln P(\omega_j) \\ &= \sum_{i=1}^{d} \left[\ln \left(\frac{p_{ij}}{1 - p_{ij}} \right)^{x_i} + \ln(1 - p_{ij}) \right] + \ln P(\omega_j) \\ &= \underline{\underline{\sum_{i=1}^{d} x_i \ln \frac{p_{ij}}{1 - p_{ij}} + \sum_{i=1}^{d} \ln(1 - p_{ij}) + \ln P(\omega_j).}} \end{aligned}$$

Oppgave 5 (forts.)

Dette er lineære diskriminantfunksjoner (lineære i $\boldsymbol{x}$) som kan skrives på formen

$$g_j(\boldsymbol{x}) = \boldsymbol{w}_j^t \boldsymbol{x} + w_{0j} = \sum_{i=1}^{d} x_i w_{ij} + w_{0j}$$

der

$$w_{ij} = \ln \frac{p_{ij}}{1 - p_{ij}}$$

$$w_{0j} = \sum_{i=1}^{d} \ln(1 - p_{ij}) + \ln P(\omega_j)$$

Desisjonsregelen blir da

$$\text{Velg } \omega_k \text{ hvis } g_k(\boldsymbol{x}) \geqslant g_j(\boldsymbol{x}), \quad j \neq k,$$

som oppgaven spurte om.