TEK5020/9020 Mønstergjenkjenning
Høsten 2023

Forelesning 2 – Beslutningsteori (1)

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

UiO : Institutt for teknologisystemer

16. august 2023

Innhold i kurset

- Introduksjon til mønstergjenkjenning
- Beslutningsteori (desisjonsteori)
- Parametriske metoder
- Ikke-parametriske metoder
- Lineære og generaliserte diskriminantfunksjoner
- Evaluering av klassifikatorer
- Ikke-ledet læring
- Klyngeanalyse.

Beslutningsteori – grunnleggende begreper

Objekter skal tilordnes *klasser/tilstander*:

$$\omega_1, \omega_2, ..., \omega_c$$

der c er antall klasser i problemet. Til hver klasse hører en *a priori sannsynlighet*:

$$P(\omega_1), P(\omega_2), ..., P(\omega_c)$$

som er sannsynligheten for at hver klasse skal opptre (før målinger er foretatt).

Til hver klasse hører også *klassebetingede sannsynlighetstetthetsfunksjoner*:

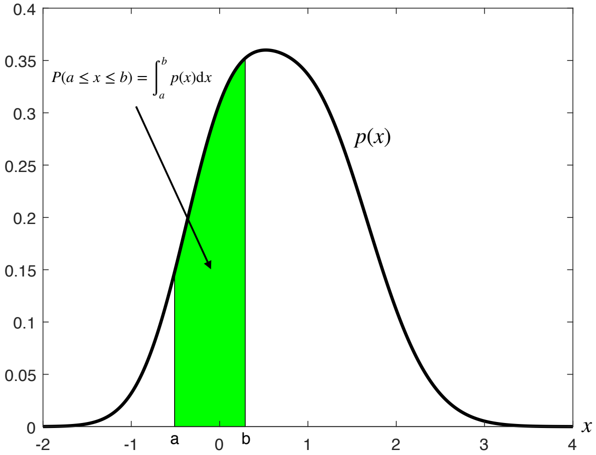
$$p(\boldsymbol{x}|\omega_i), \; i = 1, ..., c.$$

Her er vektoren:

$$\boldsymbol{x} = [x_1, x_2, ..., x_d]^t$$

en målt *egenskapsvektor* for det aktuelle objektet.

Sannsynlighetstetthetsfunksjon



Eksempel på tetthetsfunksjon. Det grønne arealet tilsvarer sannsynligheten for at et vilkårlig sample skal opptre med egenskapsverdi x i intervallet mellom a og b.

Bayes regel

Bayes regel for *a posteriori sannsynlighet*:

$$\boxed{P(\omega_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\omega_i)P(\omega_i)}{\sum_{j=1}^{c} p(\boldsymbol{x}|\omega_j)P(\omega_j)}, \; i = 1, \ldots, c}$$

knytter sammen a priori sannsynligheter og klassebetingede tetthetsfunksjoner.

$P(\omega_i|\boldsymbol{x})$ er sannsynligheten for at klasse ω_i skal opptre, gitt den målte egenskapsvektoren $\boldsymbol{x}$, dvs. sannsynligheten for klasse ω_i *etter* målingen $\boldsymbol{x}$.

Summen av a posteriorisannsynlighetene over alle klasser er én, dvs.

$$\sum_{i=1}^{c} P(\omega_i|\boldsymbol{x}) = 1.$$

Handlinger

Handlinger (actions):

$$\alpha_1, \alpha_2, \ldots, \alpha_a$$

er noe som utføres på bakgrunn av den målte egenskapsvektoren.

Hvor mange handlinger?

- Vanligvis er $a = c$, dvs. én-til-én sammenheng mellom klasser og handlinger (handlingen α_i består i å klassifisere til klasse ω_i),
- Generelt er $a \neq c$, f.eks. $a = c + 1$ der handling α_{c+1} tilsvarer forkasting (ingen klassifisering).

Desisjonsfunksjonen:

$$\alpha(\boldsymbol{x}) \rightarrow \alpha_1, \alpha_2, \ldots, \alpha_a$$

er en funksjon av egenskapsvektoren $\boldsymbol{x}$, som har én av de mulige handlingene som utfall.

Kostnader knyttet til handlinger

Kostfunksjonen:

$$\lambda(\alpha_i|\omega_j), \text{ der } i = 1, \dots, a \text{ og } j = 1, \dots, c,$$

angir kostnaden (tapet) ved å velge handlingen α_i når ω_j er sann klasse.

Det kan f.eks. være et større tap forbundet ved å klassifisere bjørk som ask enn omvendt, slik at kostnadene for disse tilfellene kan være

$$\lambda(\text{velg bjørk}|\text{ask}) = 1$$
$$\lambda(\text{velg ask}|\text{bjørk}) = 10$$

mens kostnadene for riktig valg av handling som oftest vil settes til null, dvs.

$$\lambda(\text{velg bjørk}|\text{bjørk}) = \lambda(\text{velg ask}|\text{ask}) = 0.$$

Risiko knyttet til handlinger

Betinget risk (forventet tap) er kostnaden forbundet ved en gitt handling, gitt en måling (dvs. egenskapsvektoren 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

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

for en gitt desisjonsfunksjon $\alpha(\boldsymbol{x})$ med utfallene $\alpha_1, \alpha_2, \dots, \alpha_a$. 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 $\boldsymbol{x}$.

Bayes desisjonsregel

Minimalisering av total risk R leder til *Bayes desisjonsregel*, som kan skrives som

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

Utfallet av desisjonsfunksjonen er da α_m, dvs.

$$\alpha(\mathbf{x}) = \alpha_m.$$

Dette er det valg av handling som gir minimum betinget risk og sikrer minimum total risk, dvs. *minimum-risk klassifisering*.

Bayes desisjonsregel for to klasser

La klassene være $\omega_1$og ω_2 og de tilhørende handlingene henholdsvis α_1 og α_2, slik at handlingene her består i å velge den ene eller den andre klassen, dvs.

$$\alpha_i : \boldsymbol{x} \rightarrow \omega_i$$

der $a = c = 2$ og α_1 og α_2 er de mulige utfallene. La kostfunksjonen være gitt ved

$$\lambda(\alpha_i|\omega_j) = \lambda_{ij} \text{ der } \lambda_{ij} > \lambda_{ii}, i \neq j.$$

Den betingede risken forbundet med hver handling blir da

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

Bayes desisjonsregel for to klasser (forts.)

Beslutningsregelen blir derfor

$$\text{Velg } \alpha_1 \text{ hvis } R(\alpha_1|\boldsymbol{x}) \leq R(\alpha_2|\boldsymbol{x}), \text{ ellers } \alpha_2$$
$$\Downarrow$$
$$\text{Velg } \omega_1 \text{ hvis } (\lambda_{11} - \lambda_{21})P(\omega_1|\boldsymbol{x}) \leq (\lambda_{22} - \lambda_{12})P(\omega_2|\boldsymbol{x}), \ \omega_2 \text{ ellers}$$
$$\Downarrow$$
$$\text{Velg } \omega_1 \text{ hvis } \frac{P(\omega_1|\boldsymbol{x})}{P(\omega_2|\boldsymbol{x})} \geq \underbrace{\frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}}}_{>0}, \ \omega_2 \text{ ellers.}$$

Her er ulikhetstegnet snudd fordi $\lambda_{11} - \lambda_{21} < 0$.

Bayes desisjonsregel for to klasser (forts.)

Ved bruk av Bayes regel kan den foregående ulikheten skrives som

$$\frac{p(\boldsymbol{x}|\omega_1)P(\omega_1)}{p(\boldsymbol{x}|\omega_2)P(\omega_2)} \geq \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}$$

Dette gir beslutningsregelen

$$\text{Velg } \omega_1 \text{ hvis } \frac{p(\boldsymbol{x}|\omega_1)}{p(\boldsymbol{x}|\omega_2)} \geq \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}, \ \omega_2 \text{ ellers.}$$

Terskelen på høyre side av ulikhetstegnet er uavhengig av egenskapsvektoren $\boldsymbol{x}$ og er forøvrig > 0.

Bayes desisjonsregel for to klasser (forts.)

Anta spesialtilfellet (egentlig et ganske vanlig valg)

$$\lambda_{ii} = 0, \lambda_{ij,i\neq j} > 0,$$

slik at riktig klassifisering er uten kostnad og alle feilklassifiseringer har en endelig kostnad. Desisjonsregelen reduseres derved til

$$\text{Velg } \omega_1 \text{ hvis } \frac{P(\omega_1|\boldsymbol{x})}{P(\omega_2|\boldsymbol{x})} \geq \frac{\lambda_{12}}{\lambda_{21}}, \ \omega_2 \text{ ellers.}$$

Hvis det er mest kostbart å velge ω_1 hvis ω_2 er sann (dvs. $\lambda_{12} \geq \lambda_{21}$) innebære dette at $P(\omega_1|\boldsymbol{x})$ må overskride $P(\omega_2|\boldsymbol{x})$ med faktoren $\lambda_{12}/\lambda_{21} > 1$ før ω_1 kan velges. Omskriving ved hjelp av Bayes regel gir her beslutningsregelen

$$\text{Velg } \omega_1 \text{ hvis } \frac{p(\boldsymbol{x}|\omega_1)}{p(\boldsymbol{x}|\omega_2)} \geq \frac{\lambda_{12}}{\lambda_{21}} \frac{P(\omega_2)}{P(\omega_1)}, \ \omega_2 \text{ ellers.}$$

Terskelen på høyre side er også her > 0 og uavhengig av $\boldsymbol{x}$.

Minimum-feilrate klassifisering

Ved å velge en null-én kostfunksjon (og sette a=c), dvs.

$$\lambda(\alpha_i|\omega_j) = \lambda_{ij} = \begin{cases} 0 & i = j \quad \text{(ingen kost for riktig klassifisering)} \\ 1 & i \neq j \quad \text{(lik kost for alle typer feilklassifiseringer)}, \end{cases}$$

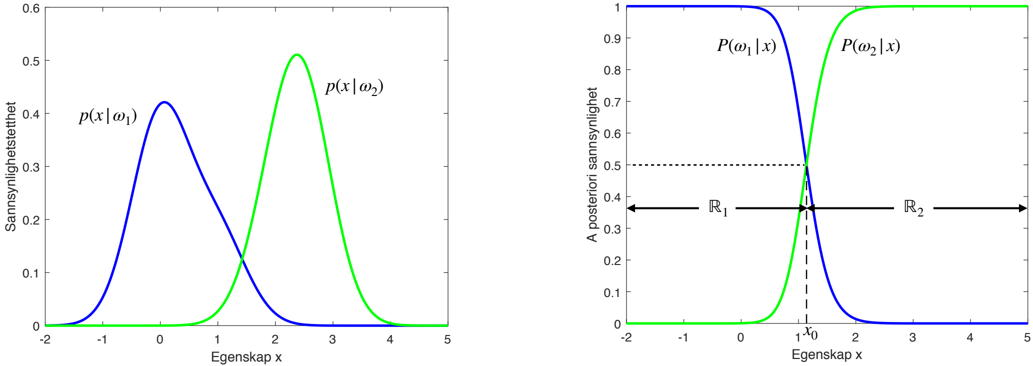
forenkles den betingede risken til

$$\begin{aligned} R(\alpha_i|\mathbf{x}) &= \sum_{j=1}^{c} \lambda_{ij} P(\omega_j|\mathbf{x}) = \sum_{j=1, j\neq i}^{c} P(\omega_j|\mathbf{x}) \\ &= 1 - P(\omega_i|\mathbf{x}), \ \text{ siden } \sum_{j=1}^{c} P(\omega_j|\mathbf{x}) = 1, \end{aligned}$$

som gir *minimum feilrate* desisjonsregelen

$$\boxed{\text{Velg } \omega_i \text{ hvis } P(\omega_i|\mathbf{x}) \geq P(\omega_j|\mathbf{x}), j = 1, \ldots, c.}$$

A posteriori sannsynlighet – minimum feilrate klassifisering



Plott av tetthetsfunksjoner (til venstre) og a posteriori sannsynligheter (til høyre) for to klasser. Optimal desisjonsgrense (x_0) der de a posteriori sannsynlighetene for klassene er like ($P(\omega_1|x) = P(\omega_2|x) = 0,5$).

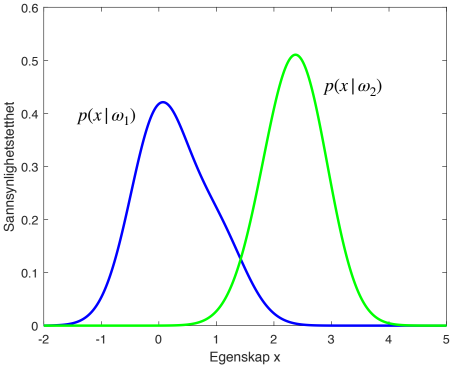
Maksimum-tiltro klassifisering

La oss se på spesialtilfellet med like a priori sannsynligheter, dvs.

$$P(\omega_i) = 1/c.$$

Da forenkles minimum-feilrate regelen til *maksimum tiltro* (maximum likelihood) regelen

$$\boxed{\text{Velg } \omega_i \text{ hvis } p(\mathbf{x}|\omega_i) \geq p(\mathbf{x}|\omega_j),\ j = 1, \ldots, c.}$$



I dette tilfellet inngår altså kun tetthetsfunksjonene i beslutningsregelen.

Feilrate og desisjonsregioner

Feilsannsynligheten (feilraten) kan skrives som

$$P(feil) = 1 - P(rett), \text{ der sannsynligheten for riktig valg er}$$

$$P(rett) = \sum_{j=1}^{c} P(\mathbf{x} \in \mathbb{R}_j, \omega_j) = \sum_{j=1}^{c} P(\mathbf{x} \in \mathbb{R}_j | \omega_j) P(\omega_j)$$
$$= \sum_{j=1}^{c} \left\{ \int_{\mathbb{R}_j} p(\mathbf{x}|\omega_j) \mathrm{d}\mathbf{x} \right\} P(\omega_j) = \sum_{j=1}^{c} \int_{\mathbb{R}_j} p(\mathbf{x}|\omega_j) P(\omega_j) \mathrm{d}\mathbf{x}.$$

$P(rett)$ maksimaliseres (og feilraten $P(feil)$ minimaliseres) ved å velge $\mathbb{R}_i$, $i = 1, \dots, c$ slik at

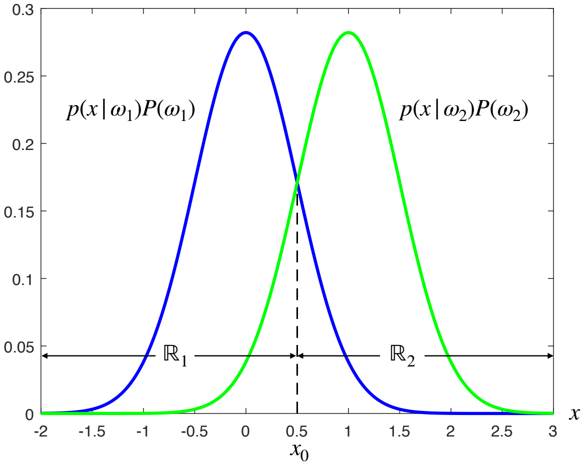
$$p(\mathbf{x}|\omega_i)P(\omega_i) \geq p(\mathbf{x}|\omega_j)P(\omega_j), \ j = 1, \dots, c \text{ for alle } \mathbf{x} \in \mathbb{R}_i.$$

De optimale desisjonsgrensene går da gjennom punkter $\mathbf{x}$ der

$$p(\mathbf{x}|\omega_i)P(\omega_i) = p(\mathbf{x}|\omega_j)P(\omega_j) \text{ for } i \neq j$$

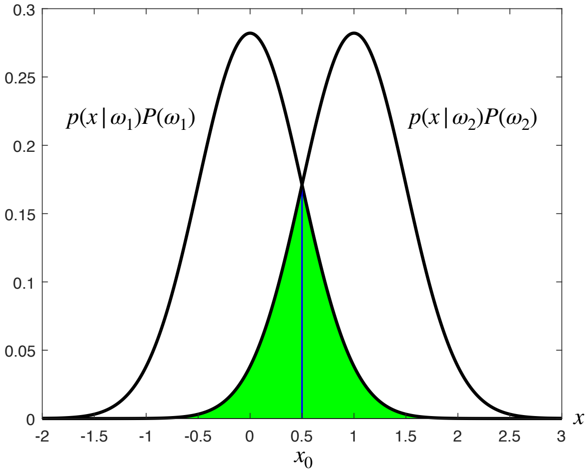
forutsatt at $p(\mathbf{x}|\omega_i)P(\omega_i)$ er maksimum over alle $\omega_1, \dots, \omega_c$.

Univariat toklasseproblem – optimale desisjonsregioner



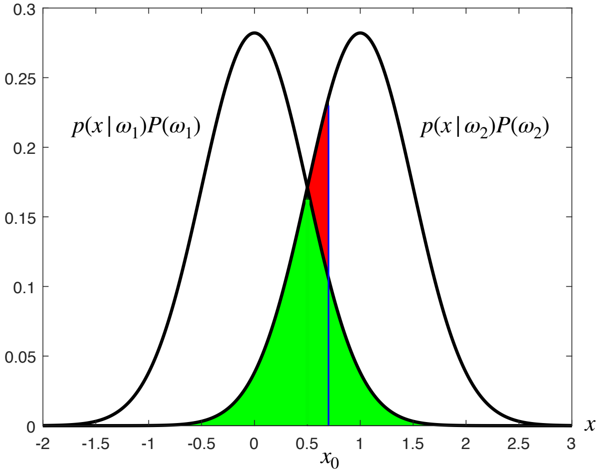
Tetthetsfunksjoner for to klasser, veiet med a priori sannsynlighet. Den stiplede linjen markerer terskelen der de veiede tetthetene er like.

Univariat toklasseproblem – minimum feilrate



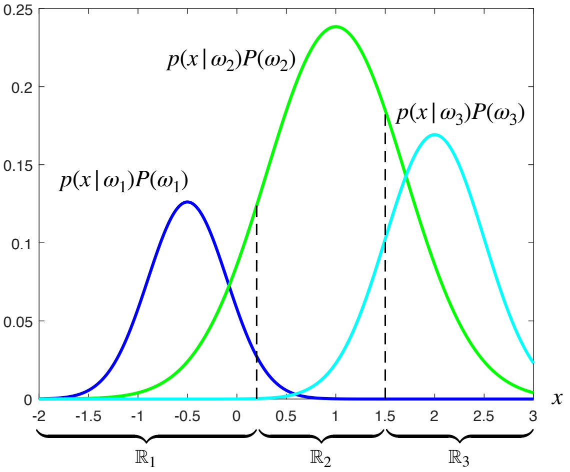
Tetthetsfunksjoner for to klasser, veiet med a priori sannsynlighet. Det grønne arealet viser feilraten med den optimale desisjonsgrensen (stiplet linje).

Feilrate med suboptimal desisjonsgrense



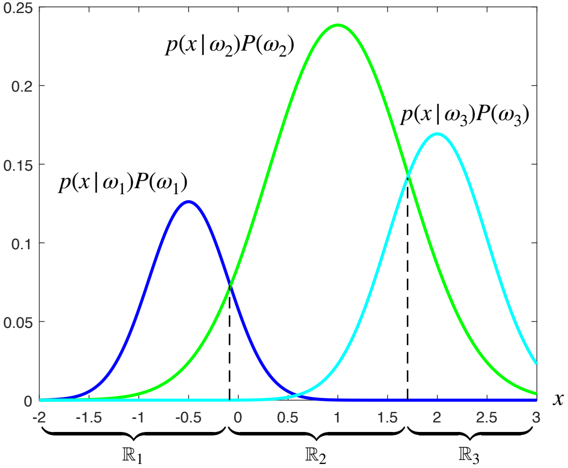
Tetthetsfunksjoner for to klasser, veiet med a priori sannsynlighet. Det røde arealet tilsvarer den ekstra feilraten ved et suboptimalt valg av desisjonsgrense.

Inndeling i desisjonsregioner – univariat problem med tre klasser



Veiede tetthetsfunksjoner for problem med tre klasser. Desisjonsgrensene deler det éndimensjonale egenskapsrommet inn i tre desisjonsregioner, men er *ikke* optimale.

Optimale desisjonsgrenser – minimum feilrate



Veiede tetthetsfunksjoner for problem med tre klasser. Desisjonsgrensene gir en optimal inndeling av egenskapsrommet i tre desisjonsregioner.

Eksempel – desisjonsgrense mellom to klasser

I et endimensjonalt (univariat) problem med to klasser ω_1 og ω_2 og tilhørende handlinger α_1 og α_2 (antall handlinger lik antall klasser), er fordelingsfunksjonene på formen

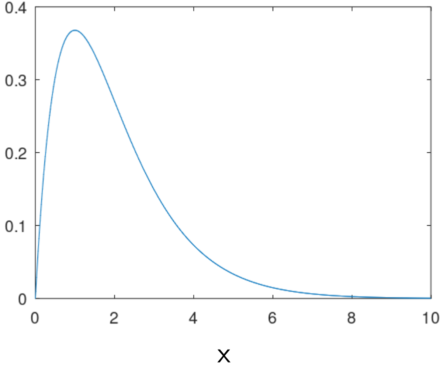
$$p(x|\theta) = \theta^2 x e^{-\theta x},$$

der $\theta > 0$ og $x \geq 0$.

Parameteren θ bestemmer fordelingen.

La de to klassene ha parametrene θ_1 og θ_2 og apriorisannsynlighetene $P(\omega_1)$ og $P(\omega_2)$.

Erlang-2 fordeling med $\theta = 1$



Eksempel – desisjonsgrense mellom to klasser (forts.)

Anta kostnadene

- $\lambda_{11} = \lambda_{22} = 0$ (null kostnad for feilfri klassifisering),
- $\lambda_{12} > 0$ og $\lambda_{21} > 0$ (kostnad for feilklassifiseringer større enn null).

Terskelen (desisjonsgrensen) x_0 som minimaliserer den totale risken finner vi der den betingede risken for hver av handlingene er like, dvs.

$$R(\alpha_1|x_0) = R(\alpha_2|x_0).$$

Den betingede risken forbundet med handlingene blir, etter innsetting for kostnadene og aposteriorisannsynlighetene, i dette tilfellet

$$\begin{aligned} R(\alpha_1|x) &= \lambda_{11}P(\omega_1|x) + \lambda_{12}P(\omega_2|x) = \lambda_{12}P(\omega_2|x) = \lambda_{12}\theta_2^2 x e^{-\theta_2 x}P(\omega_2)/p(x), \\ R(\alpha_2|x) &= \lambda_{21}P(\omega_1|x) + \lambda_{22}P(\omega_2|x) = \lambda_{21}P(\omega_1|x) = \lambda_{21}\theta_1^2 x e^{-\theta_1 x}P(\omega_1)/p(x). \end{aligned}$$

Eksempel – desisjonsgrense mellom to klasser (forts.)

Terskelen kan da bestemmes ved å sette disse størrelsene like:

$$R(\alpha_1|x) = R(\alpha_2|x)$$
$$\Downarrow$$
$$\lambda_{12}\theta_2^2 x e^{-\theta_2 x} P(\omega_2) = \lambda_{21}\theta_1^2 x e^{-\theta_1 x} P(\omega_1)$$
$$\Downarrow$$
$$e^{(\theta_1 - \theta_2)x} = \frac{\lambda_{21}\theta_1^2 P(\omega_1)}{\lambda_{12}\theta_2^2 P(\omega_2)}.$$

Ved å ta logaritmen på begge sider av likhetstegnet kan det løses ut for x, slik at terskelen blir

$$x_0 = \frac{1}{\theta_1 - \theta_2} \ln\left[\frac{\lambda_{21}\theta_1^2 P(\omega_1)}{\lambda_{12}\theta_2^2 P(\omega_2)}\right].$$

Eksempel – desisjonsgrense mellom to klasser (forts.)

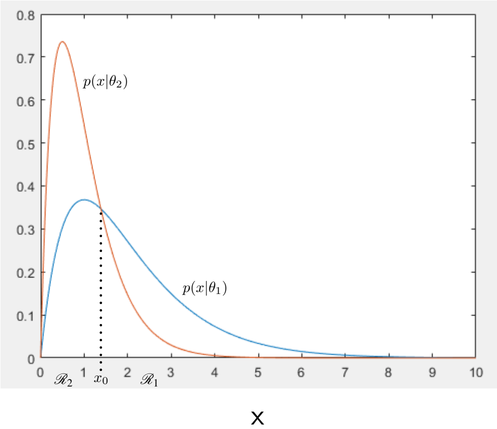
Anta nå

- $\theta_1 = 1$ og $\theta_2 = 2$,
- $\lambda_{12} = \lambda_{21}$,
- $P(\omega_1) = P(\omega_2)$.

Innsetting i uttrykket for x_0 gir da resultatet

$$x_0 = -\ln[1/2^2] = \ln(4) \approx 1,3863.$$

Figuren t.h. viser fordelingene, terskelen og desisjonsregionene i dette tilfellet.



Diskriminantfunksjoner

Diskriminantfunksjoner er et sett av funksjoner av egenskapsvektoren $\boldsymbol{x}$:

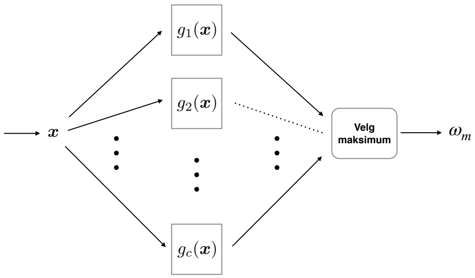
$$g_i(\boldsymbol{x}),\ i = 1, \ldots, c \quad \text{(én funksjon for hver klasse)}$$

slik at beslutningsregeler kan skrives på generell (kanonisk) form:

$$\boxed{\textit{Velg } \omega_i \text{ hvis } g_i(\boldsymbol{x}) = \max_j \{g_j(\boldsymbol{x})\}}$$

Desisjonsgrenser går gjennom punkter der to funksjoner er like, og større enn verdien til de øvrige diskriminantfunksjoner, dvs. dersom

$$g_i(\boldsymbol{x}) = g_j(\boldsymbol{x}) \text{ når } \mathbb{R}_i \text{ og } \mathbb{R}_j \text{ er naboer.}$$



Klassifiseringsmaskin.

Diskriminantfunksjoner (forts.)

Eksempler på diskriminantfunksjoner:

$$
\begin{aligned}
g_i(\mathbf{x}) &= P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{\sum_{j=1}^{c} p(\mathbf{x}|\omega_j)P(\omega_j)} \\
g_i(\mathbf{x}) &= p(\mathbf{x}|\omega_i)P(\omega_i) \\
g_i(\mathbf{x}) &= \ln\left[p(\mathbf{x}|\omega_i)P(\omega_i)\right] = \ln p(\mathbf{x}|\omega_i) + \ln P(\omega_i) \\
g_i(\mathbf{x}) &= -R(\alpha_i|\mathbf{x}).
\end{aligned}
$$

Legg merke til at bruk av logaritmen (som i det tredje eksempelet) ofte gir en enklere beslutningsregel, spesielt dersom tetthetsfunksjonene er på eksponensiell form.

Bruk av logaritmen (eller en annen monotont voksende funksjon) på f.eks. forholdet mellom a posteriori sannsynlighetene, forandrer heller ikke rangeringen av diskriminantfunksjonen og derved ikke valg av klasse.

Diskriminantfunksjoner for toklasseproblemet

For to klasser kan man for enkelhets skyld innføre en felles diskriminantfunksjon

$$g(\boldsymbol{x}) = g_1(\boldsymbol{x}) - g_2(\boldsymbol{x})$$

slik at beslutningsregelen for toklasseproblemet kan skrives som

$$\boxed{\text{Velg } \omega_1 \text{ hvis } g(\boldsymbol{x}) > 0 \text{ og } \omega_2 \text{ ellers.}}$$

Mulige diskriminantfunksjoner for to klasser er

$$g(\boldsymbol{x}) = P(\omega_1|\boldsymbol{x}) - P(\omega_2|\boldsymbol{x})$$

$$g(\boldsymbol{x}) = \ln \frac{P(\omega_1|\boldsymbol{x})}{P(\omega_2|\boldsymbol{x})} = \ln \frac{p(\boldsymbol{x}|\omega_1)}{p(\boldsymbol{x}|\omega_2)} + \ln \frac{P(\omega_1)}{P(\omega_2)}.$$

Normalfordelingen

Univariat normalfordeling (Gaussfordelingen):

$$p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = N(\mu,\sigma^2)$$

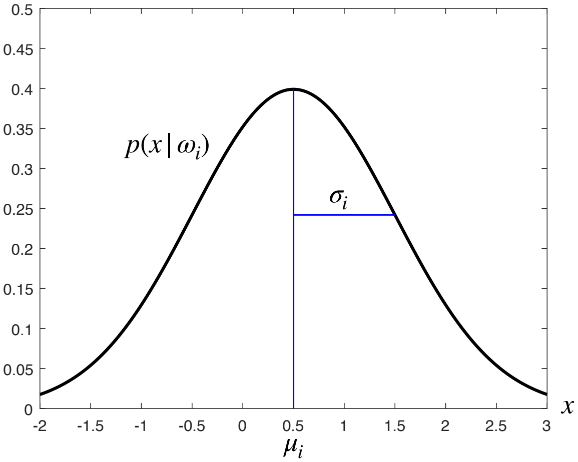
der μ er forventningsverdien og σ^2 variansen.

Multivariat normalfordeling:

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

der $\boldsymbol{\mu}$ er forventningsvektoren og Σ er kovariansmatrisen.

Univariat normalfordeling



Klassebetinget normalfordeling for klasse ω_i, der forventningsverdien er μ_i og standardavviket er σ_i.

Diskriminantfunksjoner for univariate normalfordelinger

Anta c univariat normalfordelte klasser, slik at

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

Et mulig valg for diskriminantfunksjonene er da

$$\begin{aligned} g_i(x) &= \ln P(\omega_i|x) \\ &= \ln p(x|\omega_i) + \ln P(\omega_i) - \ln p(x) \\ &= -\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2 - \frac{1}{2}\ln 2\pi - \ln \sigma_i + \ln P(\omega_i) - \ln p(x). \end{aligned}$$

Diskriminantfunksjoner for univariate normalfordelinger (forts.)

Etter fjerning av ledd som er like for alle klasser, reduseres dette til

$$g_i'(x) = \underbrace{-\frac{1}{2\sigma_i^2}}_{a_i} x^2 + \underbrace{\frac{\mu_i}{\sigma_i^2}}_{b_i} x \underbrace{- \left(\frac{\mu_i^2}{2\sigma_i^2} + \ln \sigma_i - \ln P(\omega_i) \right)}_{c_i}$$

$$= \underline{\underline{a_i x^2 + b_i x + c_i}}\,,\ i = 1, \ldots, c$$

Dette er *kvadratiske* diskriminantfunksjoner (kvadratiske med hensyn til egenskapen x).

Diskriminantfunksjoner for univariate normalfordelinger (forts.)

For to klasser kan diskriminantfunksjonene slås sammen til én felles funksjon $g(x)$ for begge klassene, slik at

$$
\begin{aligned}
g(x) &= g_1(x) - g_2(x) \\
&= \underbrace{\frac{1}{2}\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)}_{a} x^2 + \underbrace{\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right)}_{b} x + \underbrace{\frac{1}{2}\left(\frac{\mu_2^2}{\sigma_2^2} - \frac{\mu_1^2}{\sigma_1^2}\right) + \ln\frac{\sigma_2 P(\omega_1)}{\sigma_1 P(\omega_2)}}_{c} \\
&= \underline{\underline{ax^2 + bx + c}},
\end{aligned}
$$

som også er en kvadratisk diskriminantfunksjon.

Diskriminantfunksjoner for univariate normalfordelinger (forts.)

Koeffisientene a, b og c er her gitt ved

$$a = \frac{1}{2}\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)$$

$$b = \left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right)$$

$$c = \frac{1}{2}\left(\frac{\mu_2^2}{\sigma_2^2} - \frac{\mu_1^2}{\sigma_1^2}\right) + \ln\frac{\sigma_2 P(\omega_1)}{\sigma_1 P(\omega_2)}.$$

Diskriminantfunksjoner for univariate normalfordelinger (forts.)

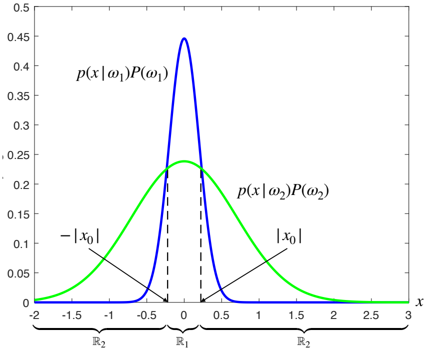
La oss se på spesialtilfellet $\mu_1 = \mu_2 = 0$. Dette gir

$$g(x) = ax^2 + c,$$

og likningen $g(x) = 0$ for desisjonsgrensene gir løsningen

$$x_0 = \pm\sqrt{-\frac{c}{a}},$$

dvs. to terskler på tallinjen.



Tersklene deler egenskapsrommet opp i to desisjonsregioner. Klassen med minst varians får en enkeltsammenhengende region omkring origo, mens den andre klassen får en todelt region for $x < -|x_0|$ til venstre og $x > |x_0|$ til høyre.

Diskriminantfunksjoner for univariate normalfordelinger (forts.)

La oss nå se på spesialtilfellet $\sigma_1 = \sigma_2 = \sigma$.

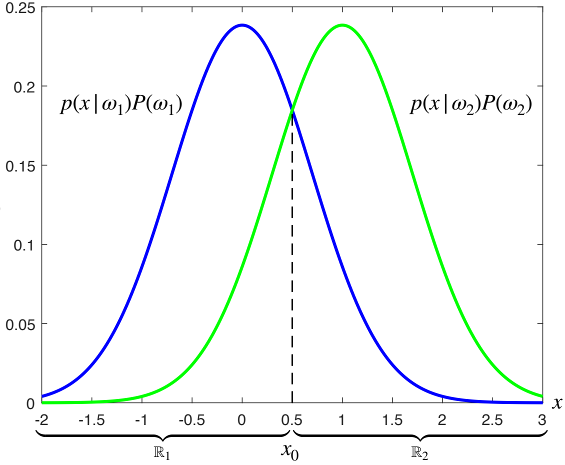
Diskriminantfunksjonen blir da

$$g(x) = \underbrace{\frac{1}{\sigma^2}(\mu_1 - \mu_2)}_{b} x + \underbrace{\frac{\mu_2^2 - \mu_1^2}{2\sigma^2} + \ln \frac{P(\omega_1)}{P(\omega_2)}}_{c} = \underline{\underline{bx + c}}.$$

Dette er en lineær diskriminantfunksjon. Desisjonsgrensen finnes ved å løse likningen $g(x) = 0$, som i dette tilfellet gir:

$$x_0 = -\frac{c}{b} = -\frac{\dfrac{\mu_2^2 - \mu_1^2}{2\sigma^2} + \ln \dfrac{P(\omega_1)}{P(\omega_2)}}{\dfrac{1}{\sigma^2}(\mu_1 - \mu_2)} = -\frac{\mu_2^2 - \mu_1^2 + 2\sigma^2 \ln \dfrac{P(\omega_1)}{P(\omega_2)}}{2(\mu_1 - \mu_2)}.$$

Diskriminantfunksjoner for univariate normalfordelinger (forts.)



To klasser med like varianser og forskjellige forventningsverdier.

Multivariat normalfordeling

Den multivariate normalfordelingen (normalfordeling i et rom av vilkårlig dimensjon) er gitt ved

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} = N(\boldsymbol{\mu}, \Sigma)$$

Her er $\boldsymbol{\mu}$ *forventningsvektoren*

$$\boldsymbol{\mu} = E\{\boldsymbol{x}\} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_d \end{bmatrix} \text{ (d komponenter)}$$

og Σ *kovariansmatrisen*

$$\Sigma = E\{(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^t\} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1d} \\ \vdots & \vdots & & \vdots \\ \sigma_{d1} & \sigma_{d2} & \dots & \sigma_{dd} \end{bmatrix} \text{ } (d \times d \text{ komponenter}).$$

Multivariat normalfordeling (forts.)

Matrisen Σ er symmetrisk (dvs. $\sigma_{ij} = \sigma_{ji}$) og positiv semidefinitt slik at $|\Sigma| \geq 0$. I eksemplene som følger antar vi at $|\Sigma| > 0$.

Dersom $|\Sigma| = 0$ er fordelingen avgrenset til et underrom i det d-dimensjonale egenskapsrommet, noe som kan forekomme dersom egenskapene inneholder redundant informasjon (funksjonell kobling mellom egenskaper).

Komponenter i Σ:

$$\sigma_{ij,i\neq j} = E\{(x_i - \mu_i)(x_j - \mu_j)\} = \text{ kovarians (ikke-diagonale komponenter)},$$

$$\sigma_{ii} = E\{(x_i - \mu_i)^2\} = \sigma_i^2 = \text{ varians (diagonale komponenter)}.$$

Multivariat normalfordeling (forts.)

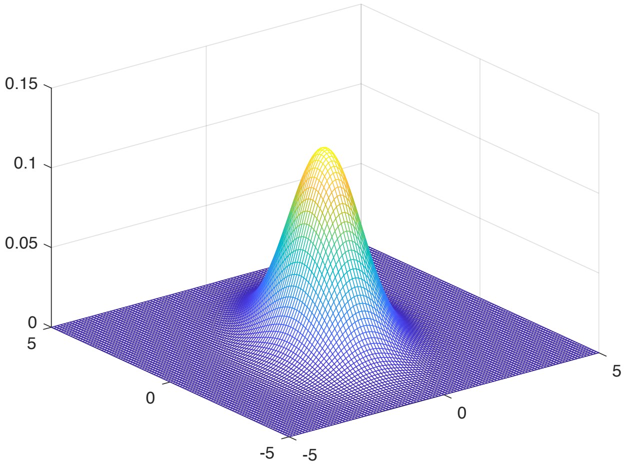
Hvis egenskapene x_i, x_j, $i \neq j$ er uavhengige medfører dette at

$$\sigma_{ij} = E\{(x_i - \mu_i)(x_j - \mu_j)\} = E\{x_i - \mu_i\}E\{x_j - \mu_j\} = 0 \cdot 0 = 0,$$

dvs. Σ er diagonal. Dette gir da

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

Multivariat normalfordeling (forts.)



Multivariat normalfordeling i to dimensjoner, dvs. *bivariat* normalfordeling.

Mahalanobis avstand

Det kvadratiske uttrykket som inngår i eksponenten i den multivariate normalfordelingen, gitt ved størrelsen

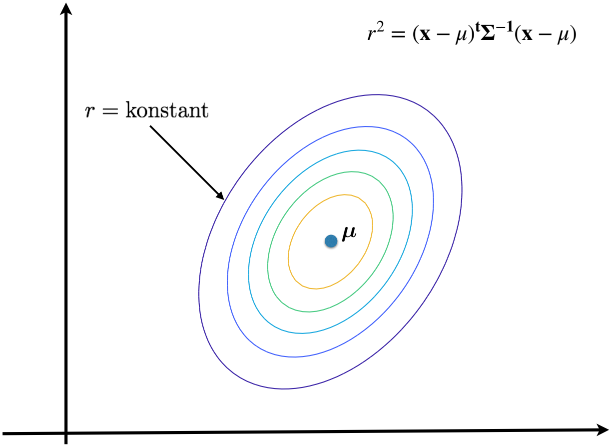
$$r^2 = (\boldsymbol{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$$

kalles gjerne den kvadrerte *Mahalanobis avstand*, dvs. et mål på avstand mellom forventningsverdien og et punkt $\boldsymbol{x}$ i egenskapsrommet.

Konturer gjennom punkter med konstant sannsynlighetstetthet, og derved konstant r, danner hyperellipsoider omkring forventningsvektoren.

Volumet av disse hyperellipsoidene er et mål på spredningen av fordelingen.

Mahalanobis avstand (forts.)



Bivariat normalfordeling representert ved ellipser med konstant tetthet (og konstant Mahalanobisavstand fra forventningsvektoren).

Mahalanobis avstand (forts.)

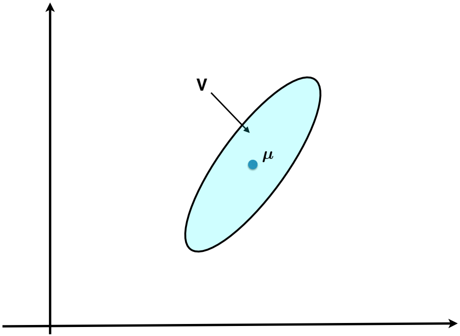
Volumet for gitt Mahalanobisavstand r er gitt ved

$$V = V_d |\Sigma|^{1/2} r^d$$

der $V_d =$ er volumet av en d-dimensjonal hyperkule med radius=1 (kan vises).

Spredningen av fordelingen er proporsjonal med $|\Sigma|^{1/2}$ for gitt verdi av d.

Stort volum indikerer stor spredning, lite volum indikerer at fordelingen er tett konsentrert omkring forventningsvektoren.



Innhold i kurset

- Introduksjon til mønstergjenkjenning
- Beslutningsteori (fortsetter neste gang)
- Parametriske metoder
- Ikke-parametriske metoder
- Lineære og generaliserte diskriminantfunksjoner
- Evaluering av klassifikatorer
- Ikke-ledet læring
- Klyngeanalyse.