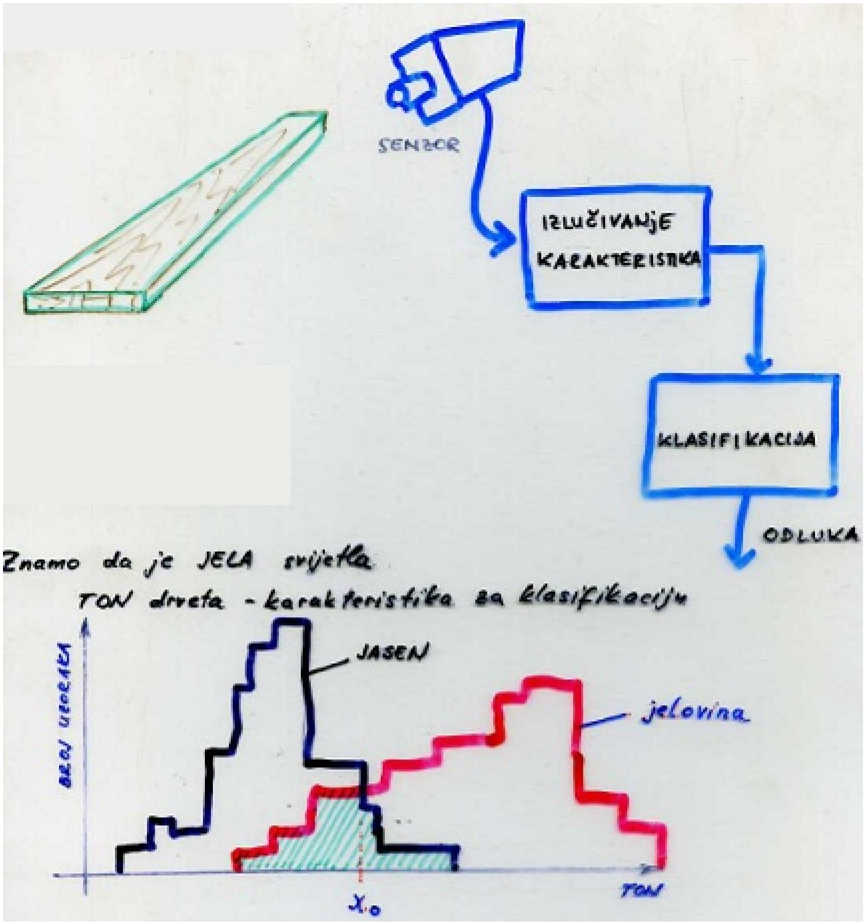
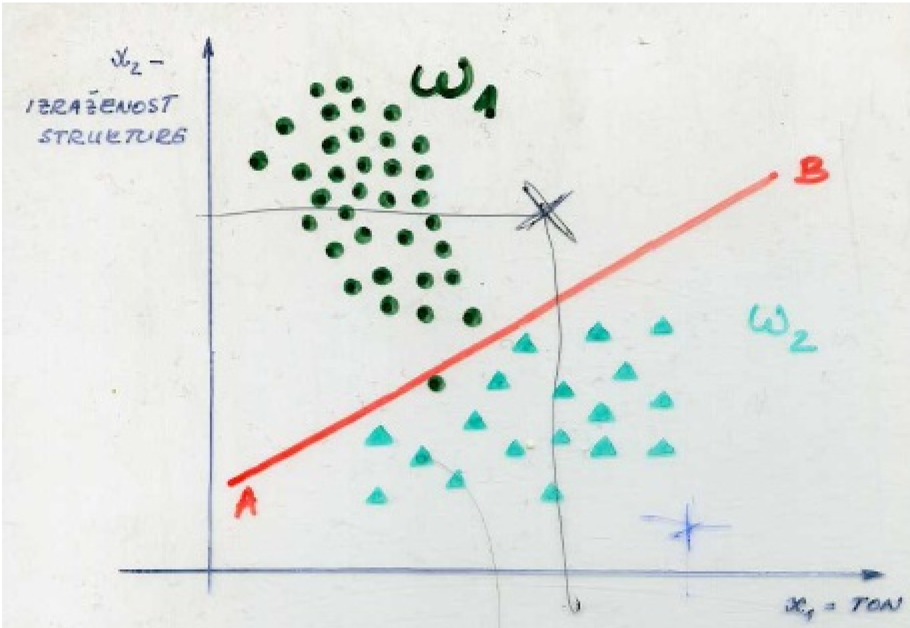
1 Klasifikacija temeljena na Bayesovoj decizijskoj teoriji /Statistički pristup raspoznavanju uzoraka/

PRIMJER: M=2
$\omega = \omega_1$ JASEN
$\omega = \omega_2$ JELA



Slika 1: Primjer

x_0 - nije dovoljan
DRUGA KARAKTERISTIKA: IZRAŽENOST STRUKTURE DRVETA!
JASEN ima izraženiju strukturu drveta
VEKTOR UZORKA
$\vec{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $\quad x_1$ - TON, x_2 - STRUKTURA



Slika 2: Prikaz klasa

PRETPOSTAVKA: PROMATRAMO POJAVLJIVANJE DASKE
- SLUČAJNO POJAVLJIVANJE VRSTE DASKE (JASEN ili JELA)
"IGRA PRIRODE" → STANJE PRIRODE (ili VANJSKOG SVIJETA) NEKA JE ω
$\omega = \omega_1$ jasen
$\omega = \omega_2$ jela

- Budući da se stanje prirode ne može predvidjeti ω se promatra kao slučajna (engl. random) varijabla

- Ako tvornica proizvodi jednak broj jelovih i jasenovih dasaka → slijedeći komad daske ima jednaku vjerojatnost da bude jasen ili pak jela

$\left.\begin{array}{l}\text{á priori vjerojatnost P}(\omega_1)\text{/ da daska bude jasena/} \\ \text{á priori vjerojatnost P}(\omega_2)\text{ / jelova daska/}\end{array}\right\}$ NAŠE "ZNANJE" O PRO-
CESU: $P(\omega_1) \geq 0$; $P(\omega_2) \geq 0$ i $P(\omega_1)+P(\omega_2)=1$
Pretpostavimo da odluku moramo donijeti samo na temelju vrijednosti *á* priori vjerojatnosti:
DECIZIJSKO PRAVILO: Odlučujem se za ω_1 ako je $P(\omega_1)>P(\omega_2)$, inače odlučujem ω_2
"Čudno pravilo" → uvijek donosim istu odluku iako znamo da se oba tipa dasaka pojavljuju
Ako je $P(\omega_1) \gg P(\omega_2)$ OK
Ako je $P(\omega_1)=P(\omega_2)$ 50%-50%
UZMIMO U OBZIR SVJETLINU odn. TON DRVETA!

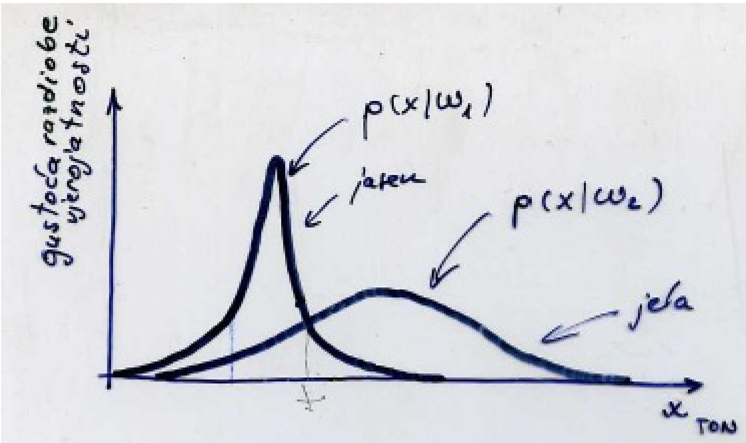
- različiti uzorci dasaka daju različitu svjetlinu: Neka je x kontinuirana slučajna varijabla čija distribucija zavisi od stanja prirode.

- uvjetna gustoća razdiobe vjerojatnosti (state-conditional probability density function)

p(x|ω_j) - gustoća razdiobe vjerojatnosti

Funkcija gustoće razdiobe vjerojatnosti (probability density function) za kontinuiranu varijablu x jeste takva funkcija $f(x)$ koja ima svojstva:

1. $f(x) \geq 0$ za svaki x
2. $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$
3. $\int\limits_{x_1}^{x_2} f(x)dx = P\{x_1 < x < x_2\}$



Slika 3: Gustoća razdiobe vjerojatnosti

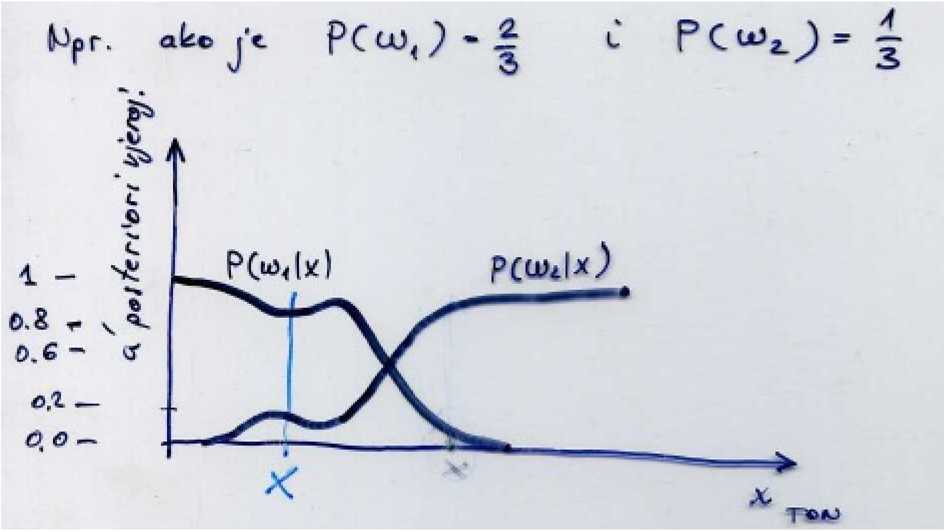
PRETPOSTAVKA:

- znamo á priori vjerojatnosti P(ω_j) i uvjetne gustoće razdiobe vjerojatnosti p(x|ω_j) (engl. likelihood of ω_j, vjerodostojnost)

Pretpostavimo da smo izmjerili svjetlinu daske i dobili vrijednost x
KAKO TO MJERENJE UTJEČE NA NAŠU ODLUKU O PRAVOM STANJU PRIRODE?
BAYESOVO PRAVILO:
$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{\underbrace{p(x)}_{\text{engl. evidence}}}$

$p(x) = \sum\limits_{j=1}^{2} p(x|\omega_j)P(\omega_j)$

BAYESOVO PRAVILO$\rightarrow$ pokazuje kako promatrana vrijednost x mijenja *á* priori vjerojatnost $P(\omega_j)$ u *á* posteriori vjerojatnost $P(\omega_j|x)$



Slika 4: Aposteriori vjerojatnosti

- imamo promatranje x za koje je $P(\omega_1|x)>P(\omega_2|x)$ odluka: stanje prirode je ω_2! VJEROJATNOST POGREŠKE?

(vjerojatnost pogreške pri svakoj odluci: $P(error|x)$)

$P(error|x)) = \begin{cases} P(\omega_1|x) & \text{ako se odlučimo za } \omega_2 \\ P(\omega_2|x) & \text{ako se odlučimo za } \omega_1 \end{cases}$

MINIMIZIRAJMO VJEROJATNOST POGREŠKE DONOŠENJEM SLIJEDEĆIH ODLUKA:
ω_1 ako je $P(\omega_1|x) > P(\omega_2|x)$ ili ω_2 ako je $P(\omega_2|x) > P(\omega_1|x)$
Da li ovo pravilo minimizira srednju vjerojatnost pogreške?
Da $\rightarrow P(error) = \int\limits_{-\infty}^{+\infty} P(error, x)dx = \int\limits_{-\infty}^{+\infty} P(error|x)p(x)dx$
jer ako je za svaki x $P(error|x)$ najmanja moguća vrijednost tada i integral mora biti najmanji.

Bayesovo decizijsko pravilo za minimizaciju vjerojatnosti pogreške:
ODLUČI SE ZA ω_1 ako je $P(\omega_1|x) > P(\omega_2|x)$ inače odluči se za ω_2

$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{\underbrace{p(x)}_{\text{(nije bitno!!!)}}}$

p(x) samo skalarni faktor koji osigurava da $P(\omega_1|x) + P(\omega_2|x) = 1$
MODIFICIRANO PRAVILO:
ODLUČI SE ZA ω_1 ako je $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$ inače ω_2

Pretpostavimo neke situacije:

- ako je za neki x $p(x|\omega_1) = p(x|\omega_2)$ odluka se temelji na á priori vjerojatnosti
- ako je $P(\omega_1) = P(\omega_2)$ odluka se temelji na $p(x|\omega_j) \rightarrow$ vjerojatnost ω_j u odnosu na x.

UOPĆENJE:

1. **upotreba više od jedne značajke,**
2. **više od dva stanja prirode,**
3. **dopustit ćemo i akciju umjesto same odluke o stanju prirode,**
4. **uvest ćemo funkciju gubitka kao općenitiju mjeru u odnosu na vjerojatnost pogreške**

1. upotreba više od jedne značajke
    - umjesto skalara x rabimo vektor značajki $\vec{X} \in R^d$; $\vec{X}$ je d-dimenzionalni vektor iz Euklidskog prostora R^d
2. više od dva stanja prirode
    - Neka je $\{\omega_1, \ldots, \omega_c$ konačan skup od c stanja (razreda, kategorija)
3. uvođenje akcije umjesto samo odluke o stanju prirode
    - skup $\{\alpha_1, \ldots, \alpha_a$ konačan skup od a mogućih akcija
4. funkcija gubitka (engl. loss function) $\lambda(\alpha_i|\omega_j)$
   $\lambda(\alpha_i|\omega_j)$ opisuje gubitak nastao poduzimanjem akcije α_i kada je stanje prirode ω_j
   $\vec{X}$ - d-dimenzionalni vektor značajki
   $p(\vec{X}|\omega_j)$ - funkcija uvjetne razdiobe gustoće vjerojatnosti
   $p(\omega_j)$ - apriorna vjerojatnost da je stanje prirode ω_j

Posteriorna vjerojatnost $P(\omega_j|\vec{X})$ (Bayesova formula):
$P(\omega_j|\vec{X}) = \frac{p(x|\omega_j)P(\omega_j)}{p(\vec{X})}$
$p(\vec{X}) = \sum_{j=1}^{c} p(x|\omega_j)P(\omega_j)$

- Pretpostavimo da smo promotrili $\vec{X}$ i da smo na temelju tog promatranja poduzeli akciju α_i
    - ako je stanje prirode ω_j, po definiciji, pretrpili smo gubitak $\lambda(\alpha_i|\omega_j)$
- Budući da je $P(\omega_j|\vec{X})$ vjerojatnost da je stanje prirode zaista ω_j očekujemo gubitak koji nastaje poduzimanjem akcije α_i
  $R(\alpha_i|\vec{X}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\vec{X}) \rightarrow$ Očekivani gubitak (rizik)
  $R(\alpha_i|\vec{X}) \rightarrow$ uvjetni rizik

Kadgod se susretnemo s određenim $\vec{X}$, možemo MINIMIZIRATI očekivani gubitak tako da izaberemo AKCIJU koja minimizira uvjetni rizik - Bayesova decizijska procedura daje optimalnu performansu
Decizijsko pravilo je funkcija $\alpha_(\vec{X})$ koja nam kaže koju akciju poduzeti za svako promatranje $\vec{X}$:
Za svaki $\vec{X}$ decizijska funkcija $\alpha(\vec{)}$ podrazumijeva izbor jedne od a akcija $\alpha_1, \ldots, \alpha_a$.
Ukupan rizik R je očekivani gubitak pridruen zadanom decizijskom pravilu.
$R(\alpha_i|\vec{X})$ je uvjetni rizik pridružen akciji α_i a budući da decizijsko pravilo specificira akciju ukupan rizik je :
$R = \int R(\alpha(\vec{X}), \vec{X})p(\vec{X})d\vec{X}$,
$d\vec{X}$ d-volumenski element a $\int$ se "proteže" nad cijelim prostorom značajki.
Ako $\alpha(\vec{X})$ je izabran tako da je $R(\alpha_i(\vec{X}))$ što je moguće manji za svaki $\vec{X}$, tako da se ukupan rizik minimizira.
Bayesovo decizijsko pravilo:
Da bi minimizirali ukupan rizik, izračunajmo uvjetni rizik
$R(\alpha_i|\vec{X}) = \sum\limits_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\vec{X})$ za $i = 1, 2, \ldots, a$ i tada IZABERIMO AKCIJU α_i za koju je $R(\alpha_i|\vec{X})$ MINIMUM.
Rezultirajući minimalni ukupni rizik naziva se Bayesov rizik.

Primjer klasifikacije za M=2

α_1 - akcija koja odgovara odluci da je pravo stanje prirode ω_1
α_2 - akcija koja odgovara odluci ω_2
Ispišimo uvjetni rizik:
$R(\alpha_1|\vec{X}) = \lambda(\alpha_1|\omega_1)P(\omega_1|\vec{X}) + \lambda(\alpha_1|\omega_2)P(\omega_2|\vec{X})$
$R(\alpha_2|\vec{X}) = \lambda(\alpha_2|\omega_1)P(\omega_1|\vec{X}) + \lambda(\alpha_2|\omega_2)P(\omega_2|\vec{X})$
Bayesovo decizijsko pravilo:
Odluči se za ω_1 ako je $R(\alpha_1|\vec{X}) < R(\alpha_2|\vec{X})$
Kraće pišemo $\lambda_{ij} = \lambda(\alpha_i|\omega_j)$!
$\lambda_{11}P(\omega_1|\vec{X}) + \lambda_{12}P(\omega_2|\vec{X}) < \lambda_{21}P(\omega_1|\vec{X}) + \lambda_{22}P(\omega_2|\vec{X})$
$(\lambda_{11} - \lambda_{21})P(\omega_1|\vec{X}) < (\lambda_{22} - \lambda_{12})P(\omega_2|\vec{X})$
$-(\lambda_{21} - \lambda_{11})P(\omega_1|\vec{X}) < -(\lambda_{12} - \lambda_{22})P(\omega_2|\vec{X}) \Big/ (-1)$
$(\lambda_{21} - \lambda_{11})P(\omega_1|\vec{X}) > (\lambda_{12} - \lambda_{22})P(\omega_2|\vec{X})$
Odlučujemo se za ω_1 ako je $(\lambda_{21} - \lambda_{11})P(\omega_1|\vec{X}) > (\lambda_{12} - \lambda_{22})P(\omega_2|\vec{X})$
Općenito vrijedi da je $\lambda_{11} < \lambda_{21}$ i $\lambda_{22} < \lambda_{12}$
(gubitak za ispravnu klasifikaciju je manji negoli za neispravnu)
$\lambda_{21} - \lambda_{11} > 0$
$\lambda_{12} - \lambda_{22} > 0$
Odlučimo se za ω_1 ako je:
$(\lambda_{21} - \lambda_{11})P(\omega_1|\vec{X}) > (\lambda_{12} - \lambda_{22})P(\omega_2|\vec{X})$
$\frac{p(\vec{X}|\omega_1)}{p(\vec{X}|\omega_2)} > \frac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})} \cdot \frac{P(\omega_2)}{P(\omega_1)} \rightarrow$ omjer vjerodostojnosti (engl. likelihood ratio)
Odlučujemo se za ω_1, ako je omjer vjerodostojnosti prešao prag koji je nezavisan od $\vec{X}$!
Prag: $\frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}$

- Klasifikacija s najmanjom pogreškom
  U problemu klasifikacije → svakom se stanju prirode obično pridružuje jedan od c razreda, a akcija α_i se tumači kao odluka da je pravo stanje prirode ω_i:
  - Ako je α_i akcija poduzeta i ako je pravo stanje prirode ω_j tada je odluka ispravna ako je i=j, a pogrešna ako je $i \neq j$.
  - Ako se pogreška želi izbjeći prirodno je tražiti decizijsko pravilo koje minimizira vjerojatnost pogreške (error rate).

Funkcija gubitka za ovaj slučaj (klasifikacija) se obično izabire kao SIMETRINČNA ili NULA-JEDAN funckija gubitka:

$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \; i, j = 1, \ldots, c$

- funkcija gubitka dodjeljuje gubitak 0 za ispravnu odluku, 1 za bilo koju pogrešku (sve pogreške jednako "koštaju").
Rizik koji odgovara toj funkciji je:

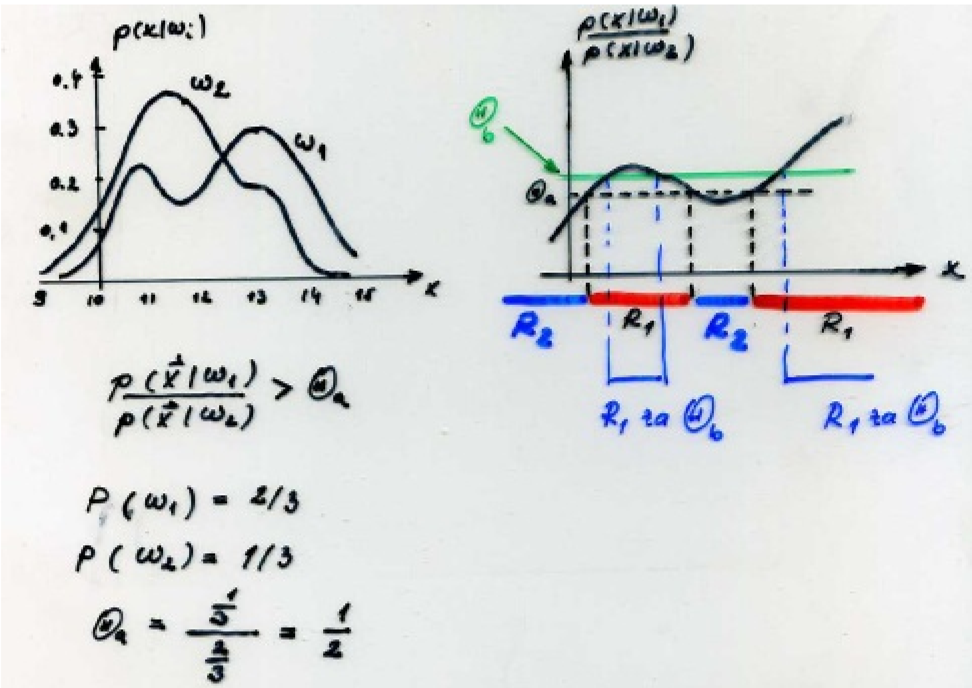
$R(\alpha_i|\vec{X}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\vec{X}) = \sum_{j\neq i} P(\omega_j|\vec{X}) = 1 - P(\omega_i|\vec{X})$

$P(\omega_i|\vec{X})$ je uvjetna vjerojatnost da je akcija $\underline{\alpha_i}$.

Bayesovo decizijsko pravilo koje minimizira rizik traži izbor akcije koja minimizira uvjetni rizik.
Da minimiziramo srednju vjerojatnost pogreške, moramo izabrati i koji maksimizira posteriori vjerojatnost $P(\omega_i|\vec{X})$.
Minimalna pogreška se dobiva na ovaj način:
Odluči se za ω_i ako je $P(\omega_i|\vec{X}) > P(\omega_j|\vec{X})$ za sve $j \neq i$
Bayesovo decizijsko pravilo za minimiziranje vjerojatnosti pogreške: Odluči se za ω_1 ako je $P(\omega_1|\vec{X}) > P(\omega_2|\vec{X})$ inače odluči se za ω_2

Primjer:
$\frac{p(\vec{X}|\omega_1)}{p(\vec{X}|\omega_2)} > \frac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})} \cdot \frac{P(\omega_2)}{P(\omega_1)}$
"0-1" funckija gubitka:
$\lambda_{22} = 0;\ \lambda_{11} = 0;\ \lambda_{12} = 1;\ \lambda_{21} = 1;$
$\frac{p(\vec{X}|\omega_1)}{p(\vec{X}|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$

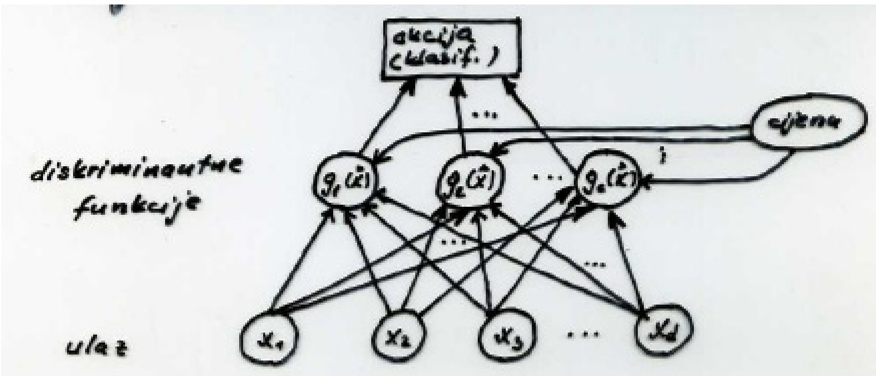
- Ako funkcija gubitka "kažnjava" više pogrešnu klasifikaciju uzorka $\vec{X}$ koji stvarno pripada ω_2 onda je $\lambda_{12} = \lambda(\alpha_1|\omega_2) > \lambda(\alpha_2|\omega_1)$ i pri $\lambda_{11} = \lambda_{22}$ $\Theta_a \rightarrow \Theta_b$ (područje R_1 se smanjuje!)



Slika 5: Primjer

Primjer klasifikacije za M=c (c>2) razreda

- skup diskriminantnih funkcija $g_i(\vec{X}), i = 1, \ldots, c$
- klasifikator razvrstava nepoznati vektor značajki $\vec{X}$ u razred ω_i ako $g_i\vec{X}) > g_j\vec{X})$ za sve $j \neq i$
  Klasifikator kao "stroj" koji računa c diskriminantnih funkcija i izabire razred koji ima najveću vrijednost diskriminante funkcije



Slika 6: Shema klasifikatora

Bayesov klasifikator se također može prikazati strukturom na slici

- Za opći slučaj s rizikom možemo izabrati $g_i(\vec{X}) = -R(\alpha_i|\vec{X})$ (maksimalna vrijednost diskriminantne funkcije odgovara minimalnoj vrijednosti uvjetnog rizika)
- za slučaj minimalne pogreške možemo uzeti $g_i(\vec{X}) = P(\omega_i|\vec{X})$ tako da maksimum diskriminantne funkcije odgovara maksimumu posteriorne vjerojatnosti
- VAŽNO: izbor diskriminantne funkcije nije jedinstven
  Npr.
  - diskriminantne fuckije možemo pomnožiti s istom pozitivnom konstantom
  - diskriminantnim funkcijama moemo pribrojiti istu pozitivnu konstantu
  - svaku $_i(\vec{X})$ možemo zamijeniti s $f(g_i(\vec{X}))$, gdje je f(.) monotono rastuća funkcija

$g_i(\vec{X}) = P(\omega_i|\vec{X}) = \frac{p(\vec{X}|\omega_i) \cdot P(\omega_i)}{\sum\limits_{j=1}^{c} p(\vec{X}|\omega_j)(\omega_j)}$
$g_i(\vec{X}) = p(\vec{X}|\omega_i) \cdot P(\omega_i)$
$g_i(\vec{X}) = \ln p(\vec{X}|\omega_i) + \ln P(\omega_i)$, gdje je ln prirodni logaritam.

- Bez obzira na oblik diskriminante funkcije decizijsko pravilo ostaje isto
  Efekt decizijskog pravila je dijeljenje prostora značajki na decizijska područja $R_1, R_2, \ldots, R_c$.
  Ako je $g_i(\vec{X}) > g_j(\vec{X})$ za sve $i \neq j$ tada je $\vec{X}$ iz R_i te decizijsko pravilo razvrstava $\vec{X}$ u razred ω_i. Područja su odjeljena decizijskim granicama

Diskrimnantne funkcije i decizijske plohe

- minimizacija vjerojatnosti pogreške je ekvivalentna dijeljenju prostora značajki u M područja (za M razreda);

- Ako su područja R_i i R_j susjedna tada ih dijeli decizijska ploha (u višedimenzionalnom prostoru značajki)

- decizijska ploha je u tom slučaju definirana jednadžbom:
  $P(\omega_i|\vec{X}) - P(\omega_j|\vec{X}) = 0$ /s jedne strane plohe razlika je pozitivna, a s druge strane plohe negativna/

- $g_i(\vec{X}) = f(P(\omega_i|\vec{X}))$ gdje je f(.) monotono rastuća funkcija
  $_i(\vec{X})$ - diskriminantna funkcija
  Pravilo klasirikacije:
  Klasificiraj $\vec{X}$ u ω_i ako $g_i(\vec{X}) > g_j(\vec{X})$; $j \neq i$

- decizijska ploha koja odvaja dvije susjedne regije (područja) je opisana s:
  $g_{ij}(\vec{X}) = g_i(\vec{X}) - g_j(\vec{X}) = 0$; $i, j = 1, 2, \ldots, M$; $i \neq j$

Bayesova klasifikacija za normalne distribucije

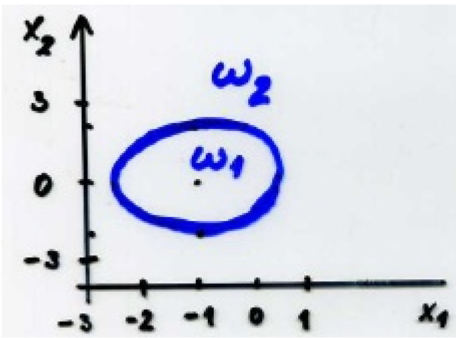
- Jedna od najčešće korištenih funkcija gustoće razdiobe vjerojatnosti (pdf-probability density function) u praksi je Gaussova ili normalna gustoća razdiobe vjerojatnosti.
  Razlog : modelira razdiobu u velikom broju slučajeva

- Pretpostavimo da funkcija vjerodostojnosti (engl. likelihood function) od ω_i u odnosu na $\vec{X}$ su u l-dimenzionalnom prostoru značajki predočene s:
  $p(\vec{X}|\omega_i) = \frac{1}{(2\pi)^{l/2}|\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(\vec{X} - \vec{\mu_i})^T \Sigma_i^{-1}(\vec{X} - \vec{\mu_i})\right)$, gdje je $\vec{\mu_i} = E(\vec{X})$ srednja vrijednost razreda ω_i; Σ_i je $l \times l$ kovarijantna matrica definirana kao:
  $\Sigma_i = E[(\vec{X} - \vec{\mu_i})(\vec{X} - \vec{\mu_i})^T]$
  $|\Sigma_i|$ - označava determinantu od Σ_i, a E[.] je srednja vrijednost (ili matematička nada; očekivanje) slučajne varijable.

Gaussova distribucija (pdf):
$\mathcal{N}(\vec{\mu}, \Sigma)$
Zadatak: Oblikovati Bayesov klasifikator

- eksponencijalni oblik nije podoban!
- koristit ćemo diskriminantnu funkciju koja rabi <u>monotonu</u> logaritamsku funkciju ln(.):
  $g_i(\vec{X}) = \underbrace{f}_{\ln(.)}(P(\omega_i|\vec{X}));$
  $g_i(\vec{X}) = \ln(p(\vec{X}|\omega_i)P(\omega_i)) = \ln p(\vec{X}|\omega_i) + \ln P(\omega_i))$
  $g_i(\vec{X}) = -\frac{1}{2}\vec{X}^T\Sigma_i^{-1}\vec{X} +$
  $\frac{1}{2}\vec{X}^T\Sigma_i^{-1}\vec{\mu_i} -$
  $\frac{1}{2}\vec{\mu_i}^T\Sigma_i^{-1}\vec{\mu_i} +$
  $\frac{1}{2}\vec{\mu_i}^T\Sigma_i^{-1}\vec{X} + \ln P(\omega_i) + c_i,$
  gdje je c_i konstanta jednaka $-\frac{l}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i|$
  Općenito, $g_i(\vec{X})$ je nelinearna kvadratna forma; Na primjer, za l=2 i za $\Sigma_i = \begin{pmatrix} \sigma_i^2 & 0 \\ 0 & \sigma_i^2 \end{pmatrix}$
  decizijska funkcija $g_i(\vec{X})$ ima oblik:
  $g_i(\vec{X}) = -\frac{1}{2\sigma_i^2}(x_1^2+x_2^2) + \frac{1}{\sigma_i^2}(\mu_{i1}x_1 + \mu_{i2}x_2) - \frac{1}{2\sigma_i^2}(\mu_{i1}^2+\mu_{i2}^2) + \ln P(\omega_i) + c_i$

Decizijske krivulje:
$g_i(\vec{X}) - g_j(\vec{X}) = 0$ su kvadratnog oblika (elipsoidi, parabole, hiperbole)
za l>2 decizijske granice → hiperkvad. forme.
<u>Primjer:</u>
$P(\omega_1) = P(\omega_2)$
$\vec{\mu_1} = [0,0]^T$ i $\vec{\mu_2} = [1,0]^T$
$\Sigma_1 = \begin{bmatrix} 0.1 & 0.0 \\ 0.0 & 0.15 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0.2 & 0.0 \\ 0.0 & 0.25 \end{bmatrix}$

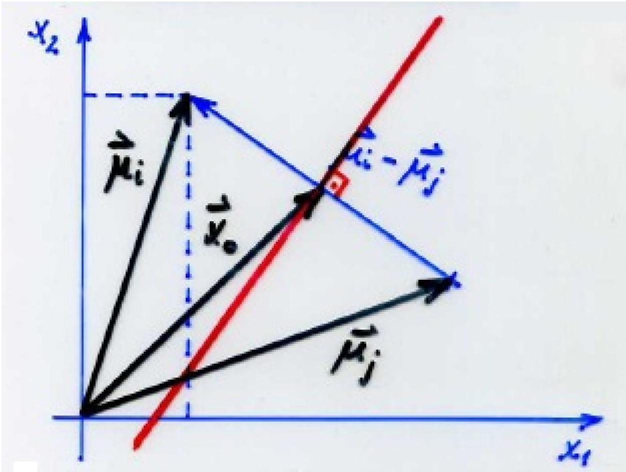


Slika 7: Primjer Gauss

<u>Decizijske hiperravnine</u>(plohe)
Doprinos kvadratnom obliku u izrazu:

$g_i(\vec{X}) = -\frac{1}{2}\vec{X}^T\Sigma_i^{-1}\vec{X} + \frac{1}{2}\vec{X}^T\Sigma_i^{-1}\vec{\mu_i} - \frac{1}{2}\vec{\mu_i}^T\Sigma_i^{-1}\vec{\mu_i} + \frac{1}{2}\vec{\mu_i}^T\Sigma_i^{-1}\vec{X} + \ln P(\omega_i) + c_i$
je prvi član $\vec{X}^T\Sigma_i^{-1}\vec{X}$
Ako pretpostavimo da je kovarijantna matrica ista za sve razrede:
$\Sigma_i = \Sigma$
kvadratni izraz $\vec{X}^T\Sigma_i^{-1}\vec{X}$ bit će jednak za sve diskriminantne funkcije!
- $\vec{X}^T\Sigma_i^{-1}\vec{X}$ i c_i možemo ispustiti!
Decizijska funkcija poprima oblik:
$g_i(\vec{X}) = \vec{W_i}^T\vec{X} + w_{i0}$, gdje je :
$\vec{W_i} = \Sigma^{-1}\vec{\mu_i}$ i $w_{i0} = \ln P(w_i) - \frac{1}{2}\vec{\mu_i}^T\Sigma^{-1}\vec{\mu_i}$
imamo $g_i(\vec{X}) = (\Sigma^{-1}\vec{\mu_i})^T\vec{X} + (\ln P(\omega_i) - \frac{1}{2}\vec{\mu_i}^T\Sigma^{-1}\vec{\mu_i})$
$g_i(\vec{X})$ je LINEARNA FUNKCIJA od $\vec{X}$ i decizijske plohe su hiperravnine!
Istražimo malo dalje:

- Kovarijantna matrica neka je dijagonalna s jednakim elementima
  - pojedine značajke koje oblikuju vektor značajki su međusobno nekorelirane i imaju jednaku varijancu:
    $E[(x_i - \mu_i)(x_j - \mu_j)] = \sigma^2\delta_{ij}$
    $\Sigma = \sigma^2 I$, gdje je I $l \times l$ jedinična matrica $g_i(\vec{X}) = \vec{W_i}^T\vec{X} + w_{i0}$, gdje je $\vec{W_i} = \Sigma^{-1}\vec{\mu_i}$ postaje:
    $$g_i(\vec{X}) = \frac{1}{\sigma^2}\vec{\mu_i}\vec{X} + w_{i0} \text{ jer je } \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 & 0 & \ldots & 0 \\ 0 & \frac{1}{\sigma^2} & 0 & \ldots & 0 \\ \vdots & & & & \\ 0 & 0 & 0\ldots & \frac{1}{\sigma^2} & \end{bmatrix}$$
  - Decizijska hiperravnina je:
    $g_{ij}(\vec{X}) = g_i(\vec{X}) - g_j(\vec{X}) = \vec{W}^T(\vec{X} - \vec{X_0}) = 0$ gdje je
    $\vec{W}\vec{\mu_i} - \vec{\mu_j}$ i $\vec{X_0} = \frac{1}{2}(\vec{\mu_i} + \vec{\mu_j}) - \sigma^2 \ln(\frac{P(\omega_i)}{P(\omega_j)})\frac{\vec{\mu_i}-\vec{\mu_j}}{\|\vec{\mu_i}-\vec{\mu_j}\|^2}$
    gdje je $\left\|\vec{X}\right\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_l^2}$ Euklidska norma od $\vec{X}$.
  - hiperravnina prolazi kroz točku $\vec{X_0}$; ako je $P(\omega_i) = P(\omega_j)$ onda je $\vec{X_0} = \frac{1}{2}(\vec{\mu_i} + \vec{\mu_j})$
    tj. hiperravnina prolazi srednjom vrijednosti od $\vec{\mu_i}$ i $\vec{\mu_j}$
  - decizijska hiperravnina (u našem slučaju: pravac) je ortogonalan na $\vec{\mu_i} - \vec{\mu_j}$
  - za svaku točku $\vec{X}$ koja leži na decizijskoj hiperravnini, vektor $\vec{X} - \vec{X_0}$ također leži na hiperravnini i vrijedi
    $g_{ij}(\vec{X}) = 0 \Rightarrow \vec{W}^T(\vec{X} - \vec{X_0}) = (\vec{\mu_i} - \vec{\mu_j})^T \cdot (\vec{X} - \vec{X_0}) = 0.$
    Iz toga slijedi da je $\vec{\mu_i} - \vec{\mu_j}$ ortogonalan na decizijsku hiperravninu.
  - hiperravnina je bliža $\vec{\mu_i}$ ako je $P(\omega_i < P(\omega_j)$, odnosno bliža $\vec{\mu_j}$ ako je $P(\omega_i) > P(\omega_j)$
    $\vec{X_0} = \frac{1}{2}(\vec{\mu_i} + \vec{\mu_j}) - \sigma^2 ln(\frac{P(\omega_i)}{P(\omega_j)})\frac{\vec{\mu_i}-\vec{\mu_j}}{\|\vec{\mu_i}-\vec{\mu_j}\|^2}$
  - ako je σ^2 mala vrijednost u odnosu na $\|\vec{\mu_i} - \vec{\mu_j}\|$ položaj hiperravnine je prilično "neostjetljiv" na vrijednosti $P(\omega_i)$ i $P(\omega_j)$.



Slika 8: Slika za dvodimenzionalni slučaj

/ mala varijanca pokazuje da su slučajni vektori grupirani u malom radijusu oko njihovih srednjih vrijednosti, tako da mali pomak hiperravnine ima mali utjecaj na rezultat klasifikacije/

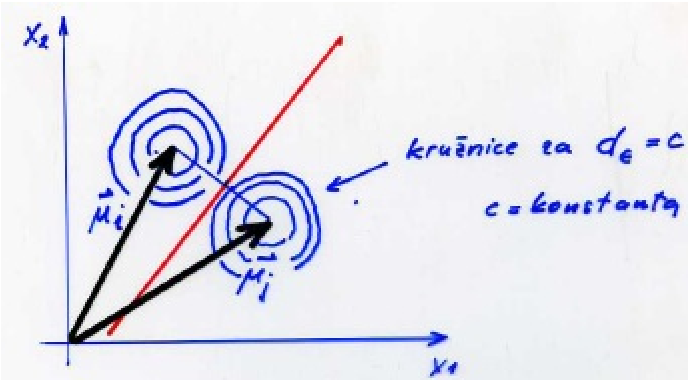
- Konvarijantna matrica nije dijagonalna
  $g_{ij}(\vec{X}) = \vec{W}^T(\vec{X} - \vec{X}_o) = 0$
  $\vec{W} = \Sigma^{-1}(\vec{\mu_i} - \vec{\mu_j})$ i $\vec{X}_0 = \frac{1}{2}(\vec{\mu_i} + \vec{\mu_j}) - ln(\frac{P(\omega_i)}{P(\omega_j)})\frac{\vec{\mu_i}-\vec{\mu_j}}{\|\vec{\mu_i}-\vec{\mu_j}\|^2}$
  gdje je $\left\|\vec{X}\right\|_{\Sigma^{-1}} \equiv (\vec{X}^T\Sigma^{-1}\vec{X})^{1/2}$, Σ^{-1} norma od $\vec{X}$
  Komentari izneseni za slučaj prije vrijede i sada (hiperravnina prolazi kroz točku $\vec{X}_0$, odnos hiperravnine naprema $\vec{\mu_i}$ i $\vec{\mu_j}$ u zavisnosti od $P(\omega_i)$ i $P(\omega_j)$) ali uz razliku da decizijska hiperravnina nije više ortogonalna na vektor $\vec{\mu_i}$ i $\vec{\mu_j}$ već na njegovu linearnu transformaciju $\Sigma^{-1}(\vec{\mu_i} - \vec{\mu_j})$.

Klasifikator na temelju minimalne udaljenosti

- pojednostavimo izraz:
$g_i(\vec{X}) = -\frac{1}{2}(\vec{X} - \vec{\mu_i})^T\Sigma_i^{-1}(\vec{X} - \vec{\mu_i}) + \ln P(\omega_i) + c_i$
pretpostavkom da su svi razredi jednako vjerojatni i da imaju istu kovarijantnu matricu:
$g_i(\vec{X}) = -\frac{1}{2}(\vec{X} - \vec{\mu_i})^T\Sigma_i^{-1}(\vec{X} - \vec{\mu_i})$
/konstantu c_i isto zanemarimo/

- za slučaj $\Sigma = \sigma^2 I$ maksimum $g_i(\vec{X})$ podrazumijeva minimum $d_e = \left\|\vec{X} - \vec{\mu_i}\right\|$, gdje je d_e Euklidska udaljenost

  - Vektor značajki klasificira se u razred u skladu s minimalnom vrijednosti Euklidske udaljenosti (između točke koja predstavlja taj vektor i srednju vrijednost $\vec{\mu_i}$)



Slika 9: Klasifikacija najmanjom udaljenosti

- za slučaj kada Σ nije dijagonalna matrica;
  $g_i(\vec{X})$ postiže maksimum kada je Σ^{-1} norma minimalna:
  $d_m = (\vec{X} - \vec{\mu_i})^T \Sigma^{-1} (\vec{X} - \vec{\mu_i})^{1/2}$ d_m= Mahalanobisova udaljenost

  - u tom slučaju za $d_m = c$ krivulje su elipse (hiperelipse)
  - Kovarijantna matrica je simetrična i može biti dijagonizirana
    $\Sigma = \phi \Lambda \phi^T$, gdje je $\phi^T = \phi^{-1}$ i Λ je dijagonalna matrica čiji elementi su svojstvene vrijednosti od Σ.
    ϕ ima stupce koji predstavljaju odgovarajuće (ortonormalne) svojstvene vektore od Σ
    $\phi = [\vec{v_1}, \vec{v_2}, \ldots, \vec{v_l}]$
    Kombinacijom $d_m = ((\vec{X} - \vec{\mu_i})^T \Sigma^{-1} (\vec{X} - \vec{\mu_i}))^{1/2}$ i $\Sigma = \phi \Lambda \phi^T$ dobivamo:
    $(\vec{X} - \vec{\mu_i})^T \phi \Lambda^{-1} \phi^T (\vec{X} - \vec{\mu_i}) = c^2$
    te $\vec{X}' = \phi^T \vec{X}$ dobivamo da su koordinate od $\vec{X}'$ jednake $\vec{v_k} \vec{X}$, $k = 1, 2, \ldots, l$, tj. projekcije $\vec{X}$ na svojstvene vektore.