Klasifikacija temeljena na Bayesovoj decizijskoj teoriji

/ statistički pristup raspoznavanju uzoraka /

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

á priori vjerojatnost $P(\omega_1)$ /da daska bude od jasena)/
á priori vjerojatnost $P(\omega_2)$ / jelova daska/

→ NAŠE "ZNANJE" O PROCESU

$$P(\omega_1) \geq 0 \quad P(\omega_2) \geq 0 \qquad \text{i} \qquad P(\omega_1) + P(\omega_2) = 1$$

Pretpostavimo da odluku moramo donijeti samo na temelju vrijednosti á priori vjerojatnosti :

DECIZIJSKO PRAVILO

Odlučujem se za ω_1 ako je

$$P(\omega_1) > P(\omega_2),\ \text{inače}$$

odlučujem ω_2

"Čudno" pravilo → uvijek donosim istu odluku iako znamo da se oba tipa dasaka pojavljuju

Ako je $P(\omega_1) >> P(\omega_2)$ O.K.

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.

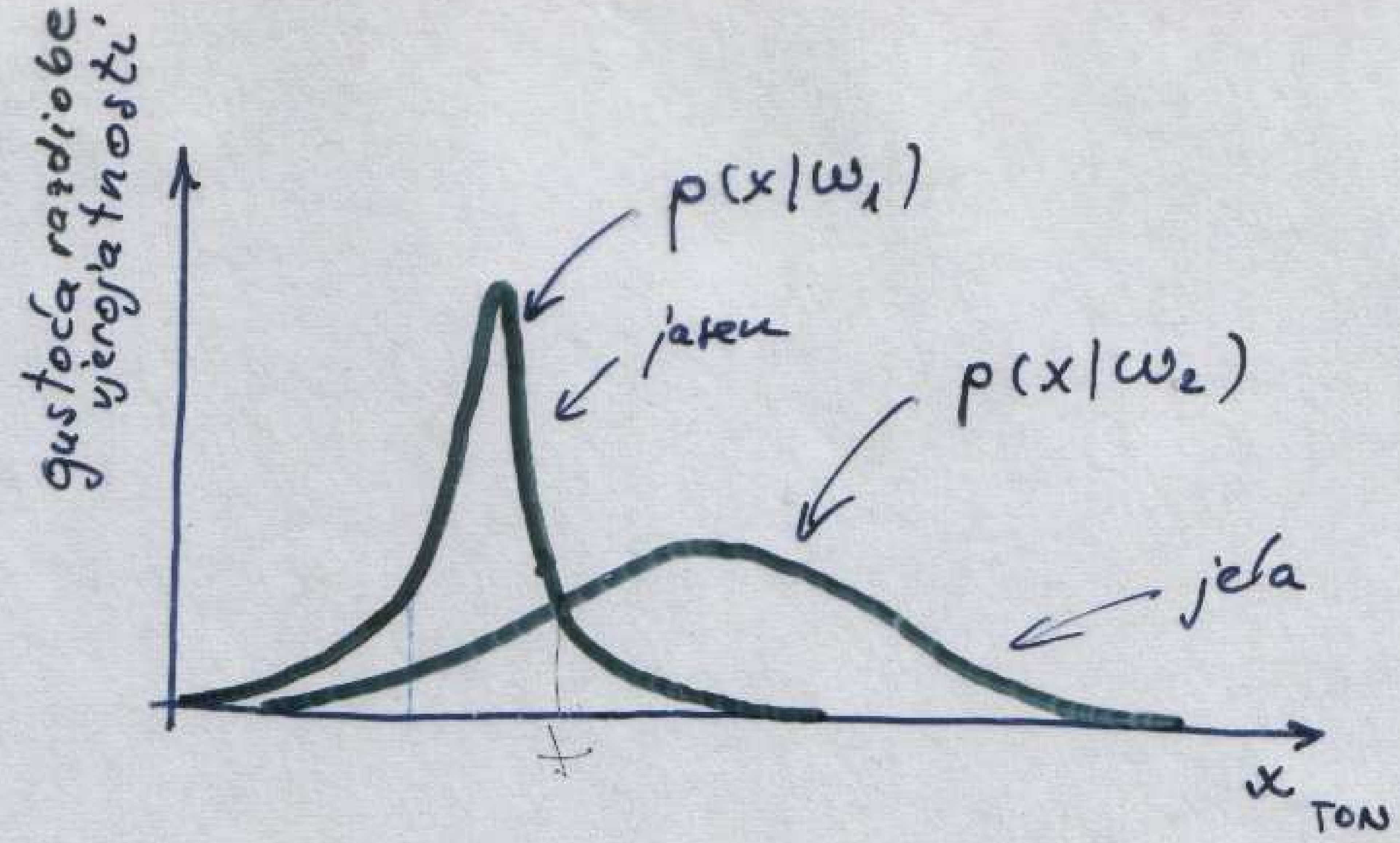
$p(x \mid \omega_j)$ - uvjetna gustoća razdiobe vjerojatnosti (state-conditional probability density function)

→ **gustoća razdiobe vjerojatnosti**

/ funkcija gustoće 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. $\int_{-\infty}^{+\infty} f(x)\,dx = 1$
3. $\int_{x_1}^{x_2} f(x)\,dx = P\{x_1 < x < x_2\}$



PRETPOSTAVKA:

- znamo á priori vjerojatnosti $P(\omega_j)$

  i uvjetne gustoće razdiobe vjerojatnosti

$$p(x|\omega_j)$$

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:
(1764. godina)

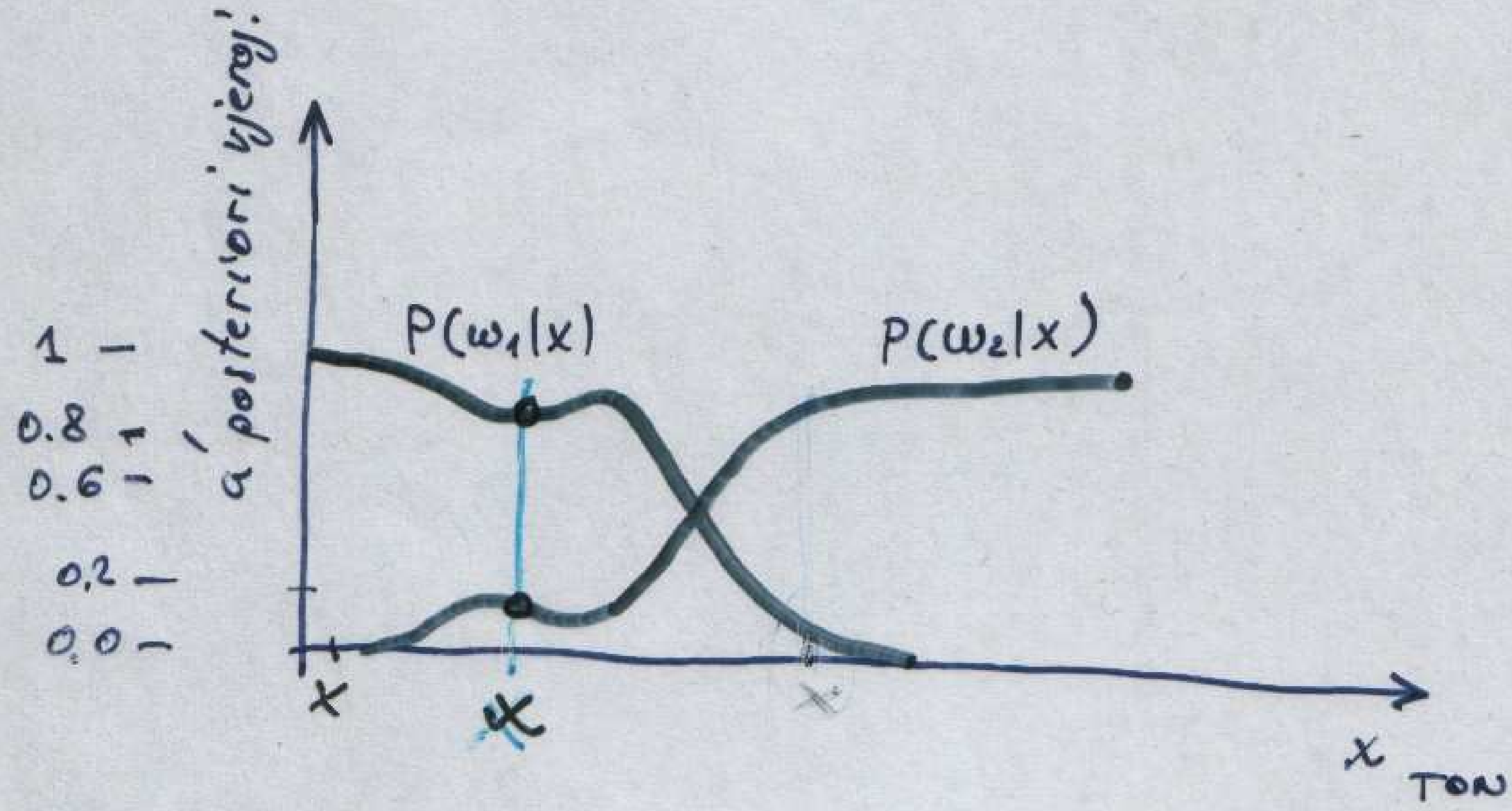
$$P(\omega_j|x) = \frac{p(x|\omega_j)\,P(\omega_j)}{p(x)}$$

→ engl. evidence

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

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

Npr. ako je $P(\omega_1) = \frac{2}{3}$ i $P(\omega_2) = \frac{1}{3}$



- imamo promatranje x za koje je

$$P(\omega_1|x) > P(\omega_2|x)$$

odluka: stanje prirode je ω_1!

VJEROJATNOST POGREŠKE ?

(vjerojatnost pogreške pri svakoj odluci : $P(\text{error}|x)$)

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

MINIMIZIRAMO 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 →

$$P(error) = \int_{-\infty}^{+\infty} P(error, x)\,dx$$

$$= \int_{-\infty}^{+\infty} P(error|x)\,p(x)\,dx$$

jer ako je sa svaki x $P(error|x)$ najmanja moguća vrijednost tada i integral mora biti najmanji.

Bayesovo decizijsko pravilo za minimizaciju vjerojatnosti pogreške:

$$\boxed{\begin{array}{l}\textbf{ODLUČI SE ZA } \omega_1 \ \underline{\textbf{ako}} \textbf{ je} \\ \quad P(\omega_1|x) > P(\omega_2|x) \ \textbf{inače} \\ \quad \textbf{odluči se za } \omega_2\end{array}}$$

$$P(\omega_j|x) = \frac{p(x|\omega_j)\,P(\omega_j)}{p(x)}$$

← nije bitno!!! (samo skalarni faktor koji osigurava da $P(\omega_1|x) + P(\omega_2|x) = 1$

MODIFICIRANO PRAVILO:

ODLUČI SE ZA ω_1 <u>ako</u> je $p(x|\omega_1) \cdot P(\omega_1) >$

$> p(x|\omega_2) \cdot P(\omega_2)$

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 $P(\omega_1) = P(\omega_2)$

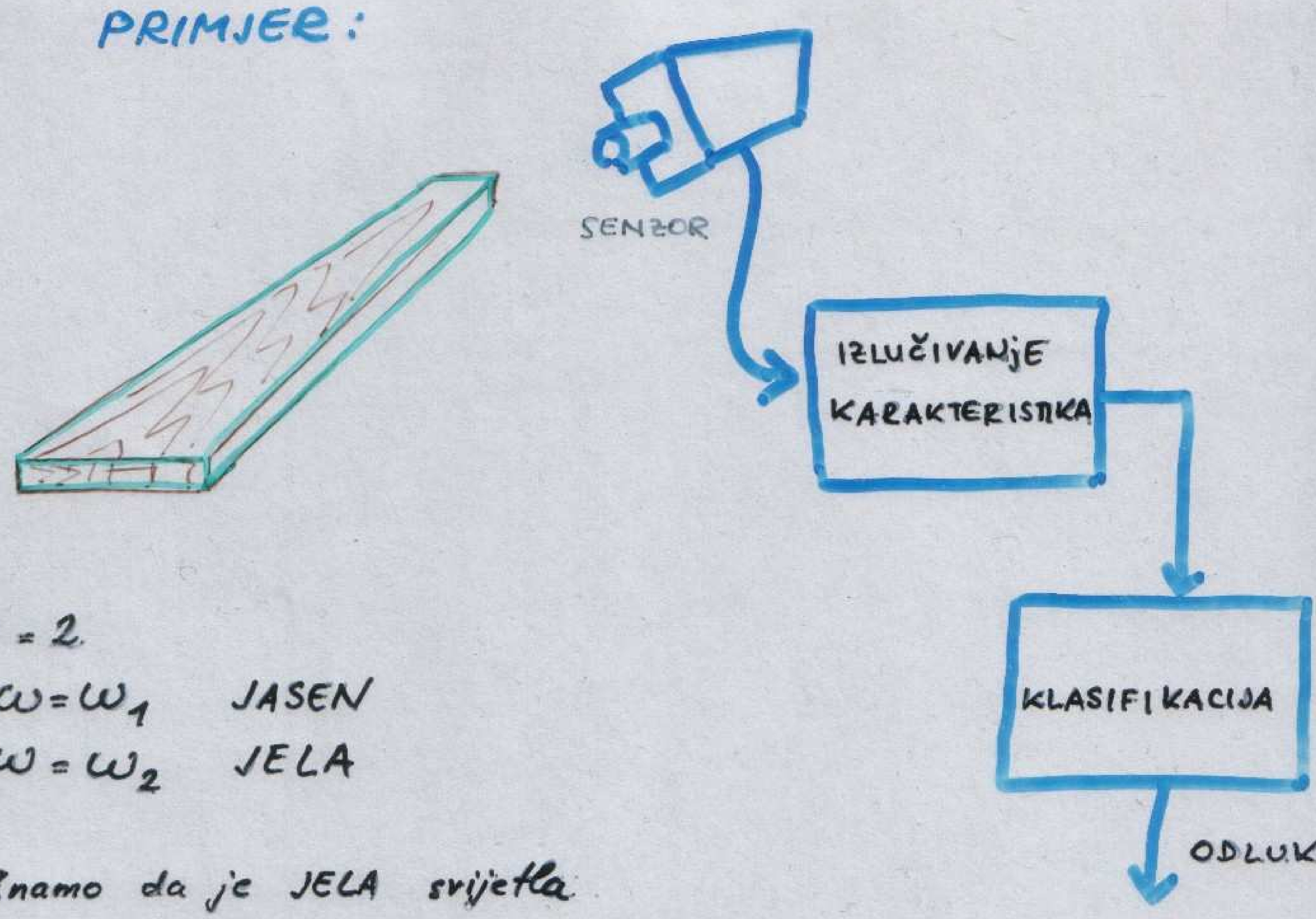
odluka se temelji na $p(x|\omega_j)$

$\Rightarrow$ 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

PRIMJER:



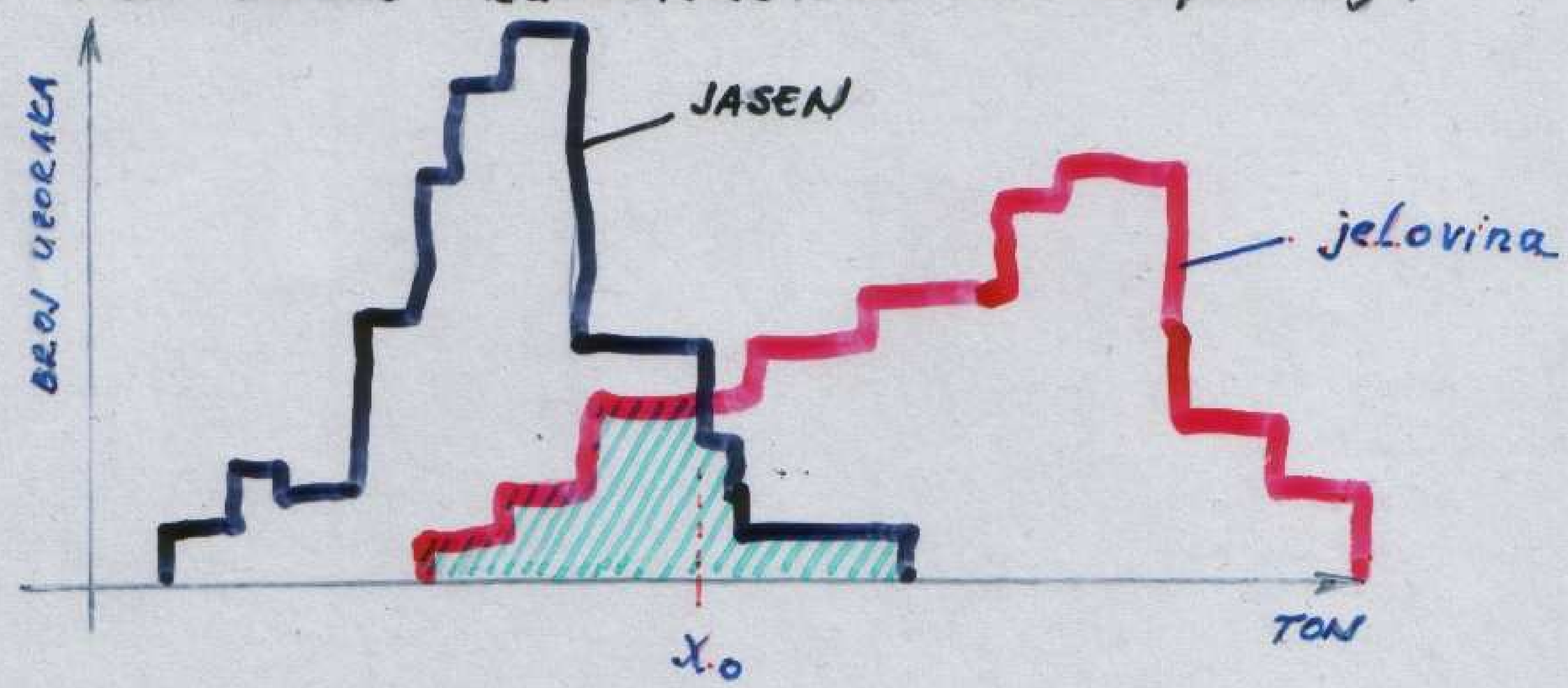
$M = 2$

$\omega = \omega_1$ JASEN

$\omega = \omega_2$ JELA

Znamo da je JELA svijetla

TON drveta - karakteristika za klasifikaciju



x_0 - nije dovoljan

DRUGA KARAKTERISTIKA : IZRAŽENOST STRUKTURE DRVETA!

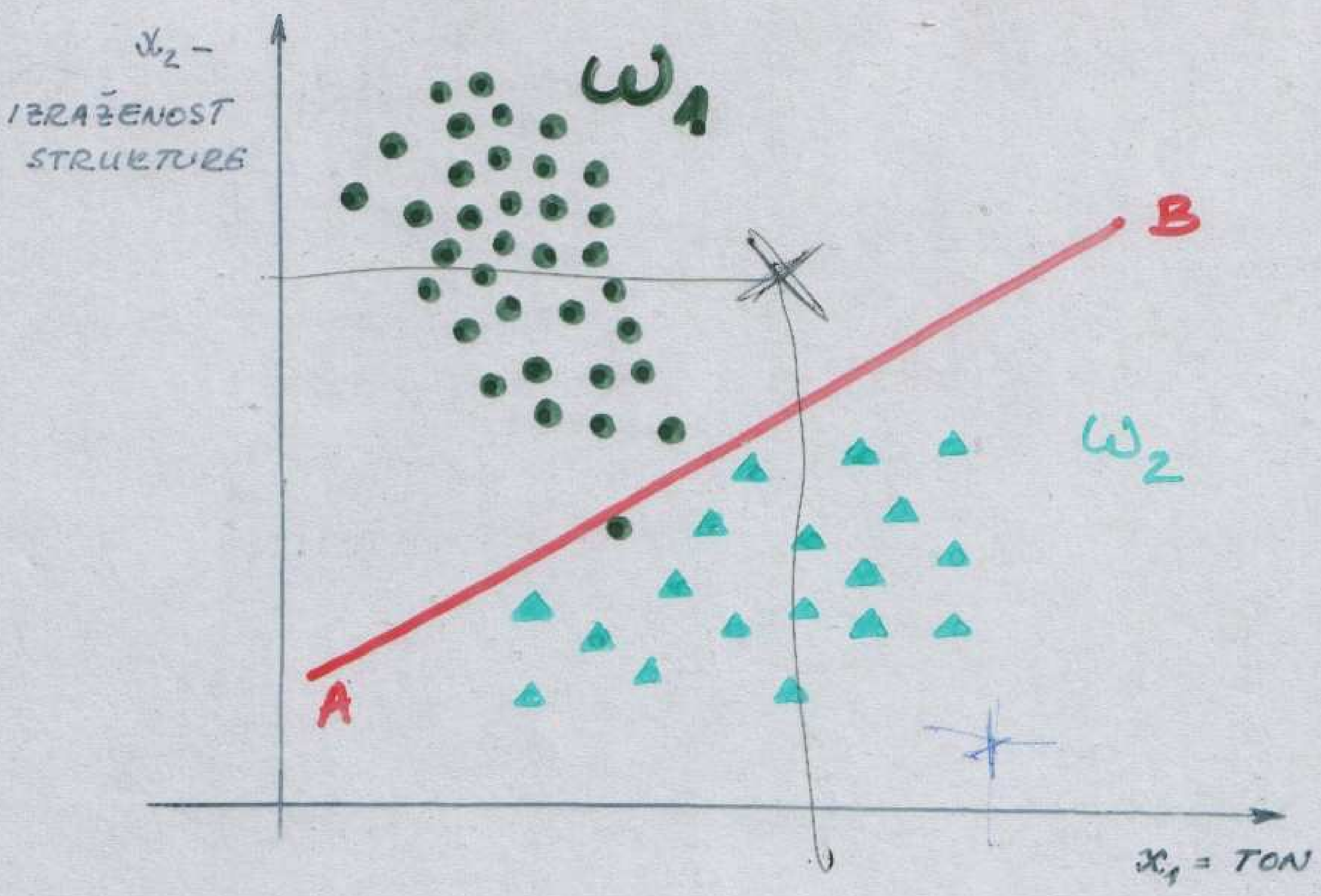
JASEN ima izraženiju strukturu drveta

VEKTOR UZORKA

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

x_1 - TON

x_2 - STRUKTURA



x_2 -
IZRAŽENOST
STRUKTURE
ω_1
B
ω_2
A
x_1 = TON

Decizijsko pravilo temeljeno na statističkom pristupu može se formulirati na dva glavna načina:

(i) Pretvorbom apriorne vjerojatnosti $P(\omega_i)$ u mjerenjem uvjetovanu (aposteriornu) vjerojatnost

$$P(\omega_i | \vec{x});$$

(ii) Definiranjem mjere očekivane klasifikacijske pogreške ili rizika te izborom decizijskog pravila koje MINIMIZIRA tu mjeru;

Tri jednostavna primjera:

Primjer 1: Nema mjerenja; broj razreda $c=2$

Pretpostavimo da je $P(\omega_1) = 0.7$ i $P(\omega_2) = 0.3$.

POZOR: nismo obavili mjerenje i nemamo vektor značajki!

Kakvo je klasifikacijsko pravilo?

- Označimo vjerojatnost klasifikacijske pogreške sa $P(\text{error})$;

$$P(\text{error}) = P(\text{izabrali } \omega_2 | \omega_1) \cdot P(\omega_1) + P(\text{izabrali } \omega_1 | \omega_2) \cdot P(\omega_2)$$

Decizijsko pravilo: Uvijek izaberi ω_1 jer je $P(\omega_1) > P(\omega_2)$.

Taj pristup vodi do pogreške

$$P(\text{error}) = P(\text{izaberi } \omega_1 | \omega_2) \cdot P(\omega_2) = (1) \cdot (0.3) = 0.3$$

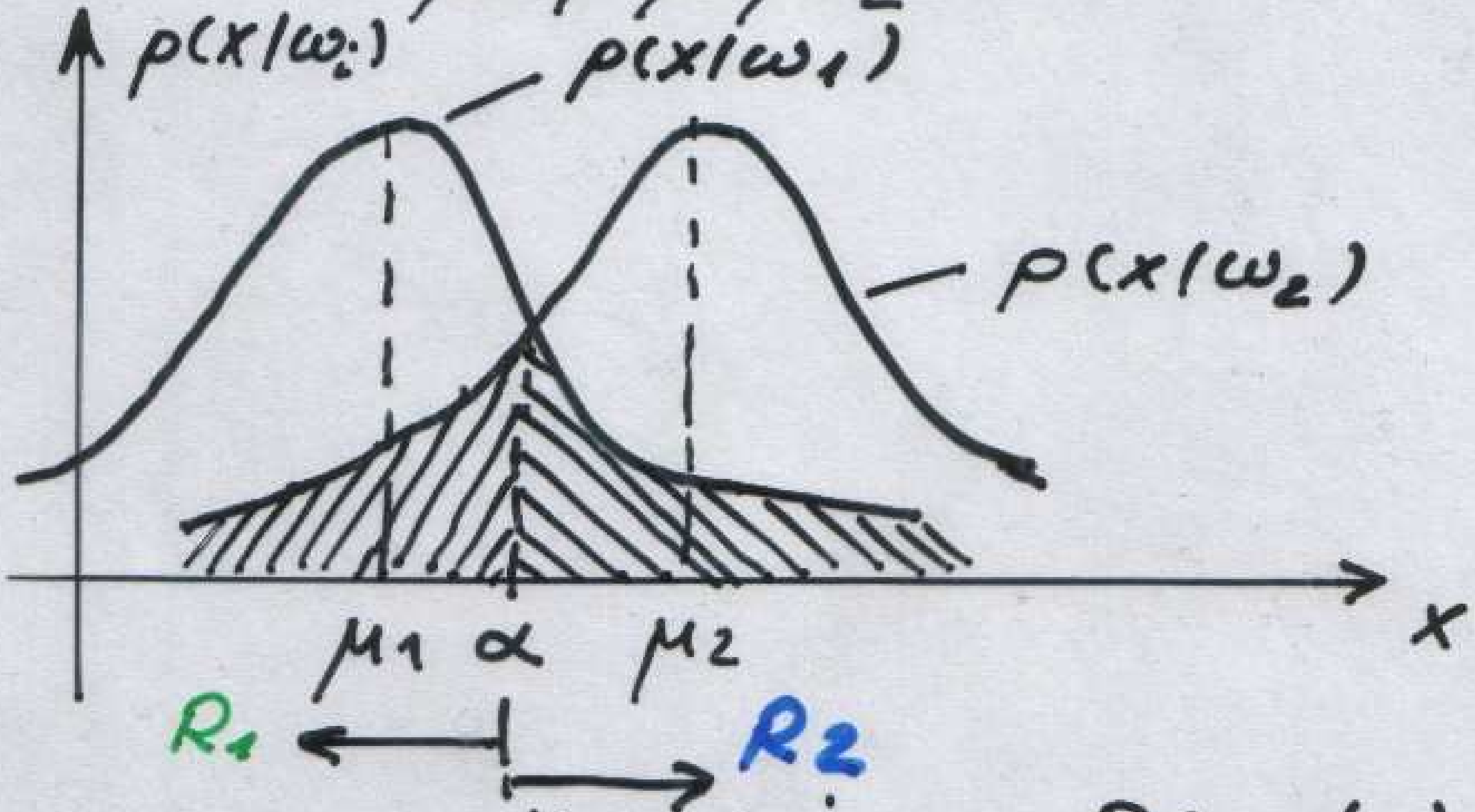
Primjer 2: Jedno mjerenje; $c=2$,
Gaussova gustoća razdiobe vjerojatnosti

- dimenzionalnost vektora značajki d (ili ℓ) $=1$
  $c=2$
  - Pretpostavimo da su apriorne vjerojatnosti jednake $P(\omega_1)=P(\omega_2)$.

$$p(x|\omega_i)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma}\right)^2\right\}$$

-poseban slučaj: varijanca i apriorne vjerojatnosti jednake $\rightarrow$ samo su srednje vrijednosti različite

$\mu_1 \neq \mu_2$!



Bayesovo pravilo daje: $P(\omega_i|x)=\dfrac{p(x|\omega_i)\,P(\omega_i)}{p(x)}$

Decizijsko pravilo:

Izaberi ω_1 ako je $p(x|\omega_1) > p(x|\omega_2)$

-Ovaj pristup vodi definiranju područja R_1 i R_2 (vidi sliku na V6a) za izabranu klasifikacijsku strategiju možemo razmotriti klasifikacijsku pogrešku:

$$P(error) = P(\vec{x} \text{ je dodijeljen pogrešnom razredu})$$

za $c=2$ imamo:

$$P(error) = P(\text{izaberi } \omega_1, \text{ a stvarno je } \vec{x} \text{ iz razreda } \omega_2) + P(\text{izaberi } \omega_2, \text{ a stvarno je } \vec{x} \text{ iz razreda } \omega_1)$$

$$= P(error|\omega_1)\cdot P(\omega_1) + P(error|\omega_2)\cdot P(\omega_2)$$

$$P(error) = P(\vec{x} \in R_2 | \omega_1)\cdot P(\omega_1) + P(\vec{x} \in R_1 | \omega_2)\cdot P(\omega_2)$$

$$= P(x < \alpha | \omega_2)\cdot P(\omega_2) + P(x > \alpha | \omega_1)\cdot P(\omega_1)$$

$$P(error) = P(\omega_2)\int_{-\infty}^{\alpha} p(\xi|\omega_2)\,d\xi + P(\omega_1)\int_{\alpha}^{+\infty} p(\xi|\omega_1)\,d\xi$$

Integrali $\int_{-\infty}^{\alpha} \cdots$ i $\int_{\alpha}^{+\infty}$ određuju šrafiranu površinu (vidi sliku na V5a)

(Re)interpretacija Primjera 2:

"Klasifikator na temelju minimalne udaljenosti" (engl. 'Minimum-Distance' Classifier)

- izbor praga α vodi dijeljenju prostora R^1 na potprostore (područja) R_1 i R_2 $\rightarrow$ za slučaj $P(\omega_1) = P(\omega_2)$ možemo reći:
"Dodijeli x u R_i pri čemu je x najbliži μ_i-u."

Primjer 3: $c=2$ $d(lišin) = 2$

- primjer iz uvoda

ω_1 = jasen

ω_2 = jela

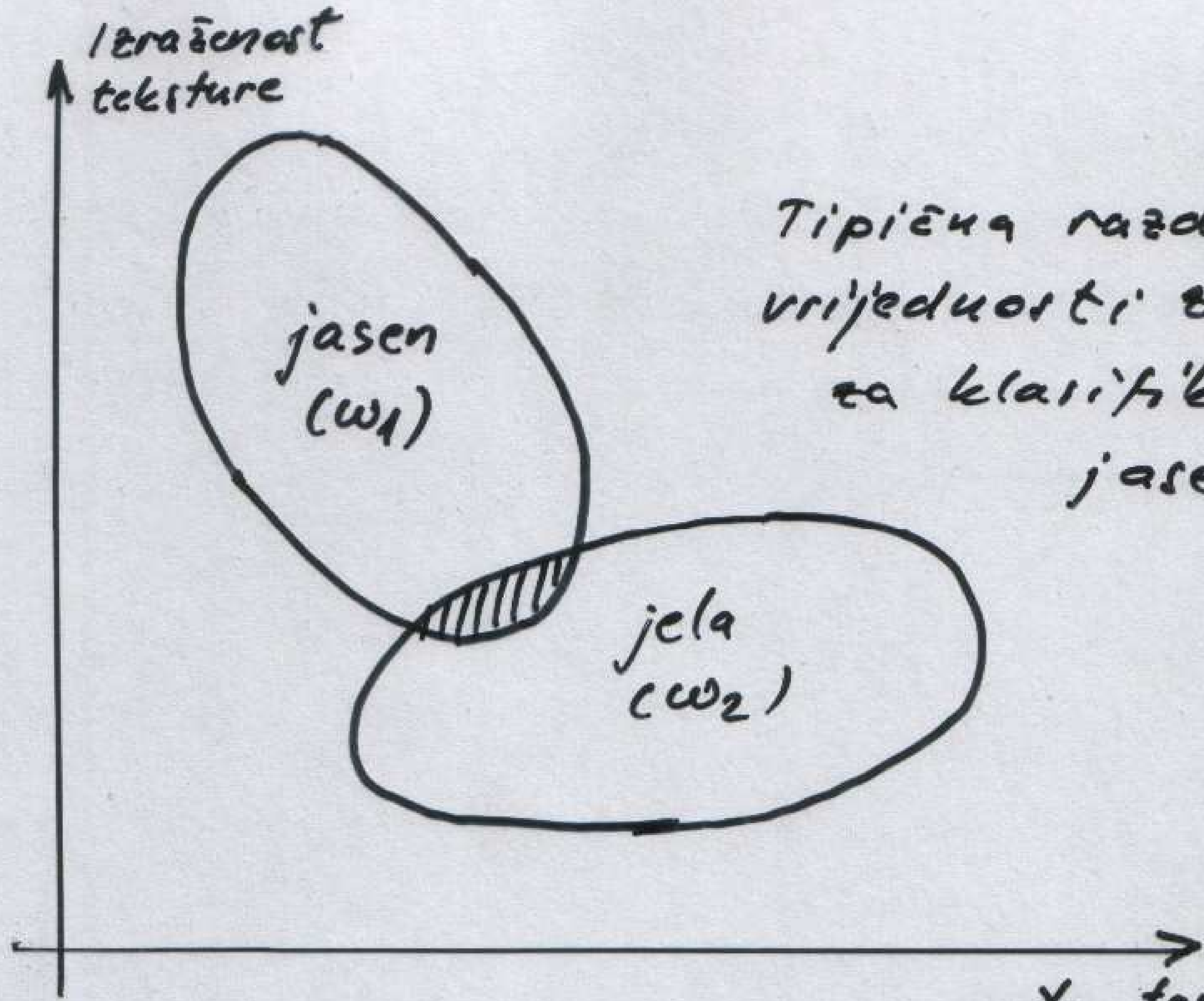
vektor značajki $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$; $\vec{x}^T = (x_1, x_2)$

x_1 - ton drveta (svjetlina)

x_2 - izraženost teksture

$p(x_1; x_2 | \omega_i) = p(\vec{x} | \omega_i)$; $i = 1,2$

$P(\omega_i)$; $i = 1,2$



Tipična razdioba vrijednosti značajki za klasifikaciju jasen - jela

Pretpostavimo da se četiri puta više proizvede jasenovih dasaka negoli jelovih

$$P(\omega_1) = 0.8$$
$$P(\omega_2) = 0.2$$

- Odrediti strategiju za klasifikaciju dasaka na temelju mjerenja $\vec{x}$ i apriornih vjerojatnosti $P(\omega_1)$ i $P(\omega_2)$ uz pretpostavku da su poznate razdiobe gustoća vjerojatnosti $p(\vec{x}/\omega_1)$ i $p(\vec{x}/\omega_2)$.

Strategija se temelji na Bayesovom pravilu:

$$P(\omega_i/\vec{x}) = \frac{p(\vec{x}/\omega_i) \cdot P(\omega_i)}{p(\vec{x})}$$

$$p(\vec{x}) = \sum_i p(\vec{x}/\omega_i) \cdot P(\omega_i)$$

PRAVILO:

$$\text{izaberi} \begin{cases} \omega_1 \text{ ako } p(\vec{x}|\omega_1)\cdot P(\omega_1) > p(\vec{x}|\omega_2)\cdot P(\omega_2) \\ \omega_2 \text{ ako } p(\vec{x}|\omega_2)\cdot P(\omega_2) > p(\vec{x}|\omega_1)\cdot P(\omega_1) \end{cases}$$

! POZOR: Monotono nepadajuća funkcija od $P(\omega_i|\vec{x})$ MOŽE SE RABITI ZA GORNJI TEST!

VAŽNO: U gornjem pravilu klasifikacije vidimo da su apriorna vjerojatnost $P(\omega_i)$ i mjerno-zavisna informacija $p(\vec{x}|\omega_i)$ kombinirane u decizijskoj proceduri!

Primjer 4:

Uniformna gustoća; $c=2$; $d=1$

$$p(x|\omega_1) = \begin{cases} \frac{1}{a_2-a_1} & \text{za } x \in [a_1, a_2] \\ 0 & \text{drugdje,} \end{cases}$$

$$p(x|\omega_2) = \begin{cases} \frac{1}{b_2-b_1} & \text{za } x \in [b_1, b_2] \\ 0 & \text{drugdje,} \end{cases}$$

Pretpostavljamo $P(\omega_1) = P(\omega_2)$!

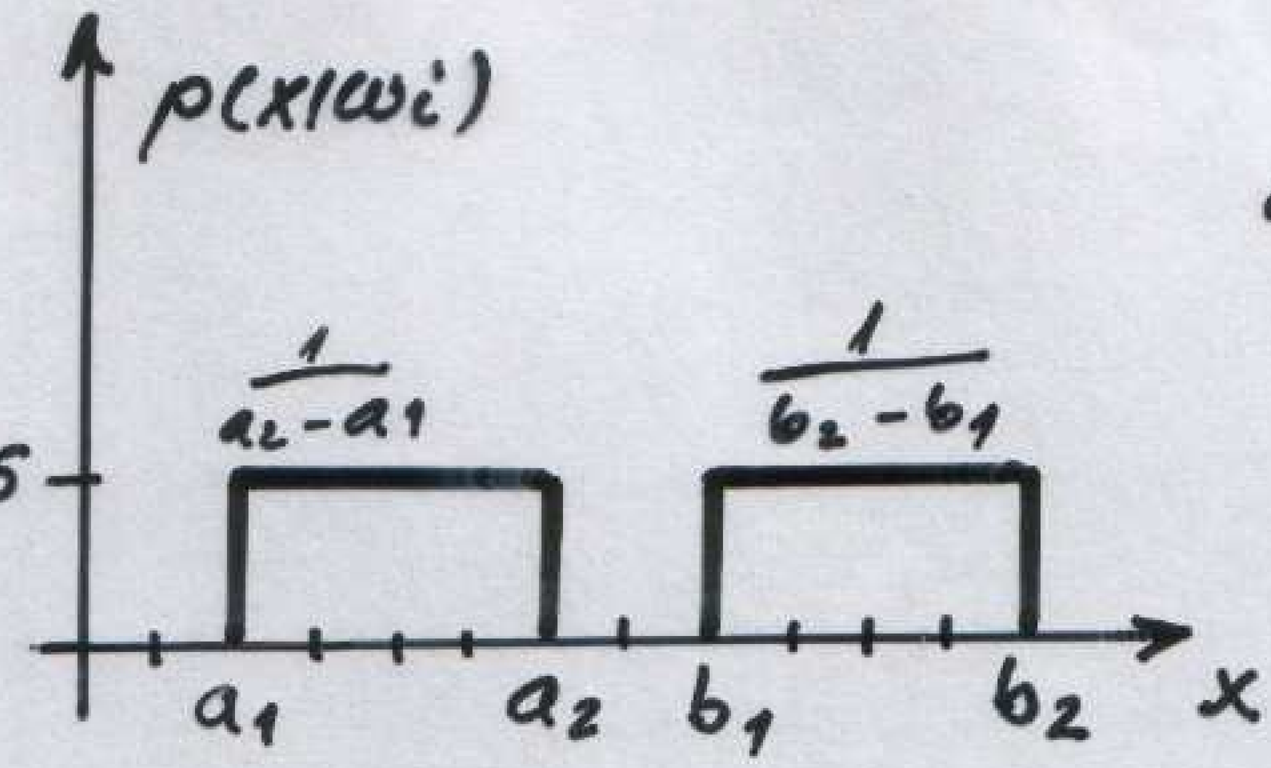
(i) $a_1 = 2$ $a_2 = 6$
$b_1 = 8$ $b_2 = 12$

(ii) $a_1 = 2$ $a_2 = 8$
$b_1 = 6$ $b_2 = 12$

(iii) $a_1 = 2$ $a_2 = 6$
$b_1 = 5$ $b_2 = 7$

(iv) $a_1 = 2$ $a_2 = 12$
$b_1 = 5$ $b_2 = 7$

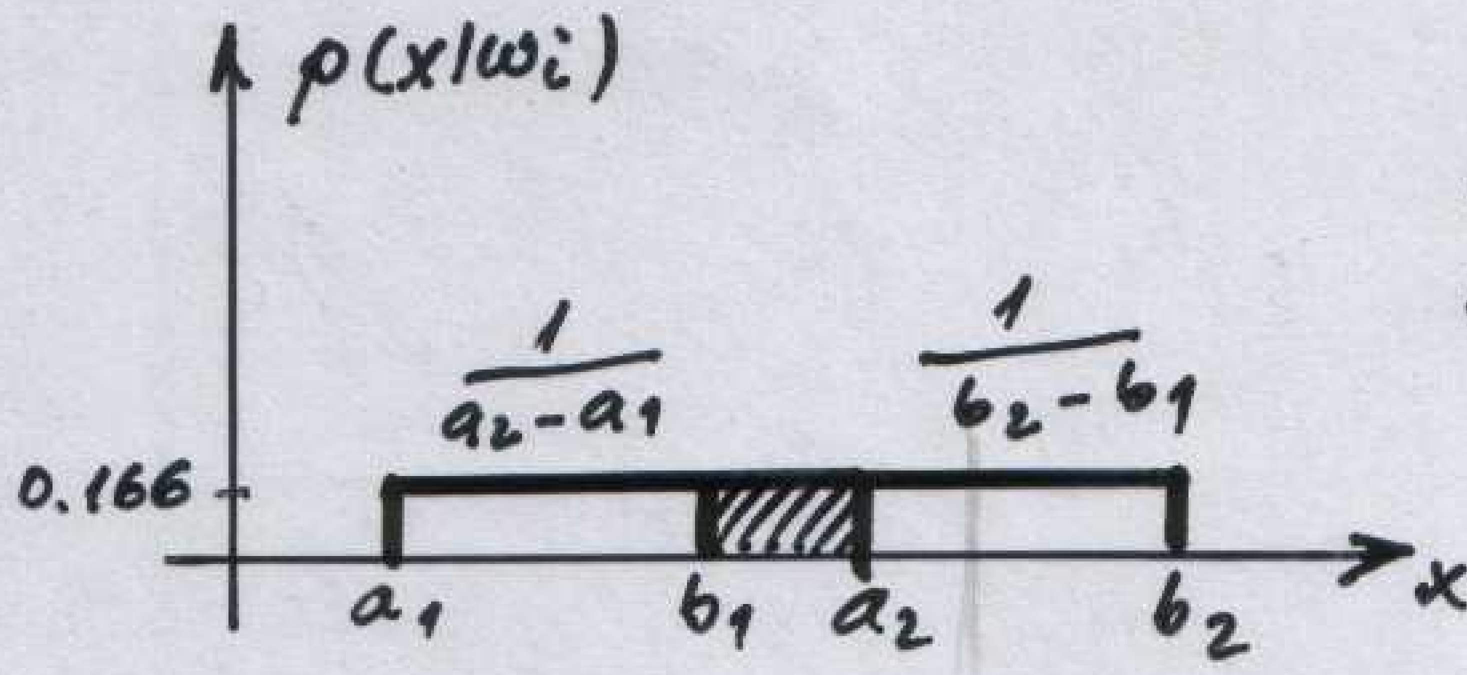
(i)



$a_1 = 2 \quad a_2 = 6$

$b_1 = 8 \quad b_2 = 12$

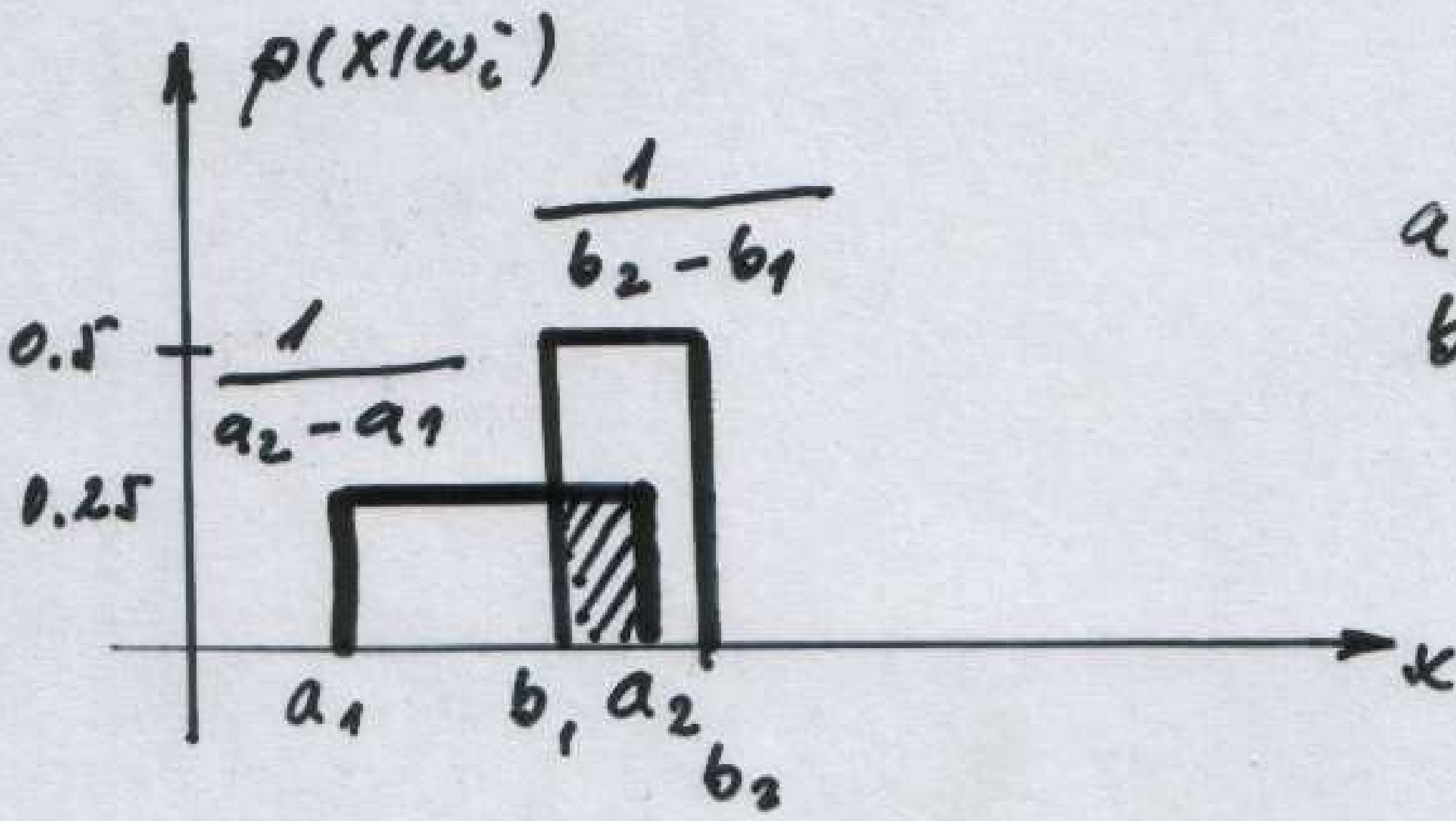
(ii)



$a_1 = 2 \quad a_2 = 8$

$b_1 = 6 \quad b_2 = 12$

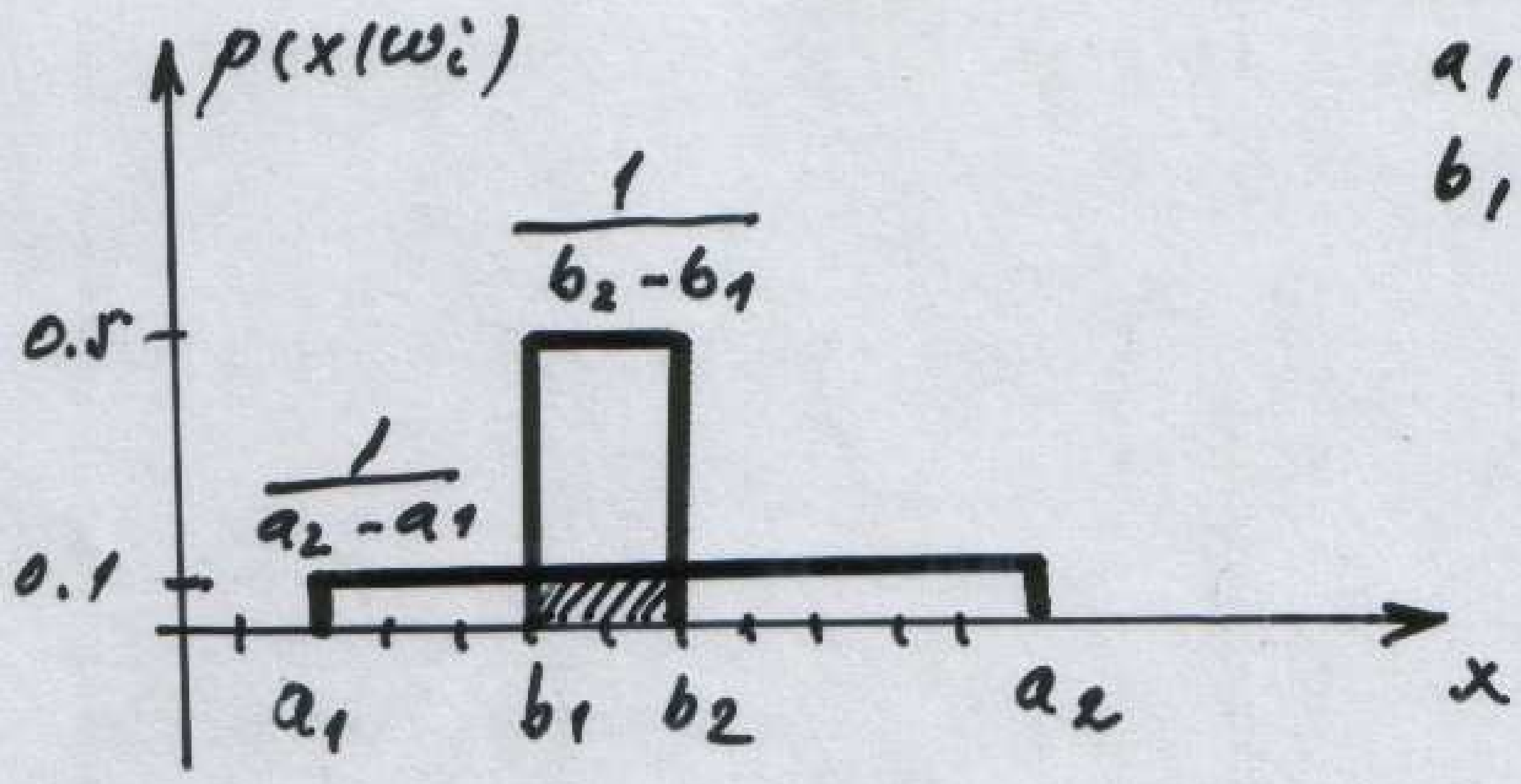
(iii)



$a_1 = 2 \quad a_2 = 6$

$b_1 = 5 \quad b_2 = 7$

(iv)



$a_1 = 2 \quad a_2 = 12$

$b_1 = 5 \quad b_2 = 7$

Slučaji (ii) – (iv) rezultirat će klasifikaciju s pogreškom (engl. nonzero classification error)

Primjer 5: $c = 3$; $d = 2$

Gaussova razdioba

$$p(\vec{x}) = (2\pi)^{-\frac{d}{2}} |\Sigma_i|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x}-\vec{\mu}_i)\right]$$

Pretpostavimo : $P(\omega_1) = P(\omega_2) = P(\omega_3)$

i

$$\Sigma_i = \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix} \quad ; \ i = 1, 2, 3$$

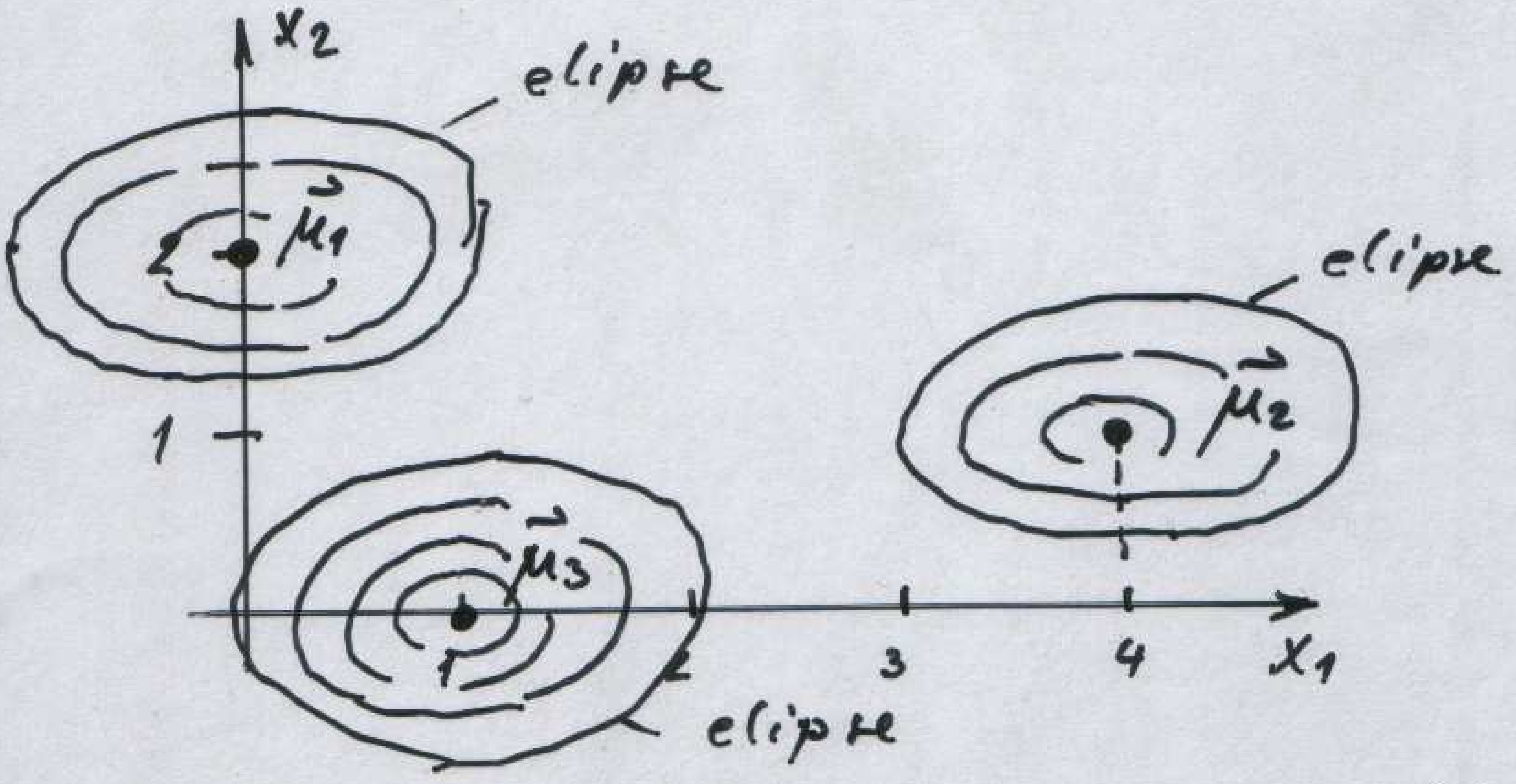
$$p(\vec{x}|\omega_i) = N(\mu_i, \Sigma_i) \qquad \vec{\mu}_1 = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \quad \vec{\mu}_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$g_i(\vec{x}) = \log\{p(\vec{x}|\omega_i)\} \qquad \vec{\mu}_2 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

$$g_i(\vec{x}) = -\frac{1}{2}\underbrace{(\vec{x}-\vec{\mu}_i)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_i)} - \left(\frac{d}{2}\right)\log(2\pi) - \frac{1}{2}\log|\Sigma|$$

$$d_i^2 = (\vec{x}-\vec{\mu}_i)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_i)$$

Mahalanobisova udaljenost !!



Funkcije gustoće vjerojatnosti

$$\Sigma^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

Oblik funkcija gustoća razdiobe vjerojatnosti:

$$(\vec{x}-\vec{\mu}_i)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_i) = k$$

k - pozitivna (proizvoljna) konstanta

za ω_1:

$$\left[\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 0 \\ 2 \end{pmatrix}\right]^T \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \left[\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 0 \\ 2 \end{pmatrix}\right] = k$$

$$\begin{pmatrix} x_1 \\ x_2-2 \end{pmatrix}^T \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2-2 \end{pmatrix} = k$$

$$[x_1 \,;\, 2(x_2-2)] \begin{pmatrix} x_1 \\ x_2-2 \end{pmatrix} = k$$

$$x_1^2 + 2(x_2-2)^2 = k \quad \text{ili}$$

$$\left(\frac{x_1}{\sqrt{2}}\right)^2 + \left(\frac{x_2-2}{1}\right)^2 = \frac{k}{2} \qquad \text{elipsa}$$

za ω_2:

$$(x_1-4)^2 + 2(x_2-1)^2 = k \quad \text{ili}$$

$$\left(\frac{x_1-4}{\sqrt{2}}\right)^2 + \left(\frac{x_2-1}{1}\right)^2 = \frac{k}{2} \qquad \text{elipsa}$$

za ω_3:

$$(x_1-1)^2 + 2x_2^2 = k \quad \text{ili}$$

$$\left(\frac{x_1-1}{\sqrt{2}}\right)^2 + \left(\frac{x_2}{1}\right)^2 = \frac{k}{2} \qquad \text{elipsa}$$

$$g_i(\vec{x}) = -(\vec{x}-\vec{\mu}_i)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_i)$$

$$g_i(\vec{x}) = -\|\vec{x}-\vec{\mu}_i\|^2_{\Sigma^{-1}}$$

$$g_i(\vec{x}) = \underbrace{-\vec{x}^T \Sigma^{-1} \vec{x}}_{\text{ne zavisi od } i} + 2\vec{\mu}_i^T \Sigma^{-1} \vec{x} - \vec{\mu}_i^T \Sigma^{-1} \vec{\mu}_i$$

Σ je simetrična!

$$g_i(\vec{x}) = (\Sigma^{-1}\vec{\mu}_i)^T \vec{x} - \frac{1}{2} \vec{\mu}_i^T \Sigma^{-1} \vec{\mu}_i$$

$$\boxed{g_i(\vec{x}) = \vec{w}_i^T \vec{x} - w_{i0}}$$

$$\vec{w}_i = \Sigma^{-1} \vec{\mu}_i$$

$$w_{i0} = +\frac{1}{2} \|\vec{\mu}_i\|^2_{\Sigma^{-1}}$$

Dobivamo:

$$g_1(\vec{x}) = \left[\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix}\right]^T \vec{x} - \frac{1}{2} \begin{pmatrix} 0 \\ 2 \end{pmatrix}^T \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

$$g_1(\vec{x}) = (0 \ \ 4) \cdot \vec{x} - \left[(0 \ \ 2) \underset{2\times 1}{\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}} \underset{2\times 2}{\begin{pmatrix} 0 \\ 2 \end{pmatrix}}\right] \cdot \frac{1}{2}$$

$$g_1(\vec{x}) = (0, 4) \cdot \vec{x} - \left[\underset{1\times 2}{\begin{bmatrix} 0 & 4 \end{bmatrix}} \underset{2\times 1}{\begin{bmatrix} 0 \\ 2 \end{bmatrix}}\right] \cdot \frac{1}{2}$$

$$g_1(\vec{x}) = (0, 4) \cdot \vec{x} - \frac{8}{2}$$

$$g_1(\vec{x}) = 4x_2 - 4$$

Na sličan način:

$$g_2(\vec{x}) = (4 \ \ 2) \cdot \vec{x} - 9 = 4x_1 + 2x_2 - 9$$

$$g_3(\vec{x}) = (1 \ \ 0) \cdot \vec{x} - \frac{1}{2} = x_1 - \frac{1}{2}$$

VIIa

Hiperravnine : $c=3$

$(c^2-c)/2 = 3$ separatibilne hiperravnine

$$g_i(\vec{x}) = g_j(\vec{x})$$

$$\vec{W}_i^T \vec{x} - W_{i0} = \vec{W}_j^T \vec{x} - W_{j0}$$

$$\vec{W}_{ij}^T \vec{x} - W_{ij0} = 0 \text{ , gdje je}$$

$$\vec{W}_{ij} = \vec{W}_i - \vec{W}_j = \Sigma^{-1}(\vec{\mu}_i - \vec{\mu}_j)$$

$$W_{ij0} = W_{i0} - W_{j0}$$

$$\vec{W}_{12} = \vec{W}_1 - \vec{W}_2 = \begin{pmatrix} 0 \\ 4 \end{pmatrix} - \begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} -4 \\ 2 \end{pmatrix}$$

$$W_{120} = -4 - (-9) = 5$$

H_{12}: $4x_1 - 2x_2 - 5 = 0$

slično:

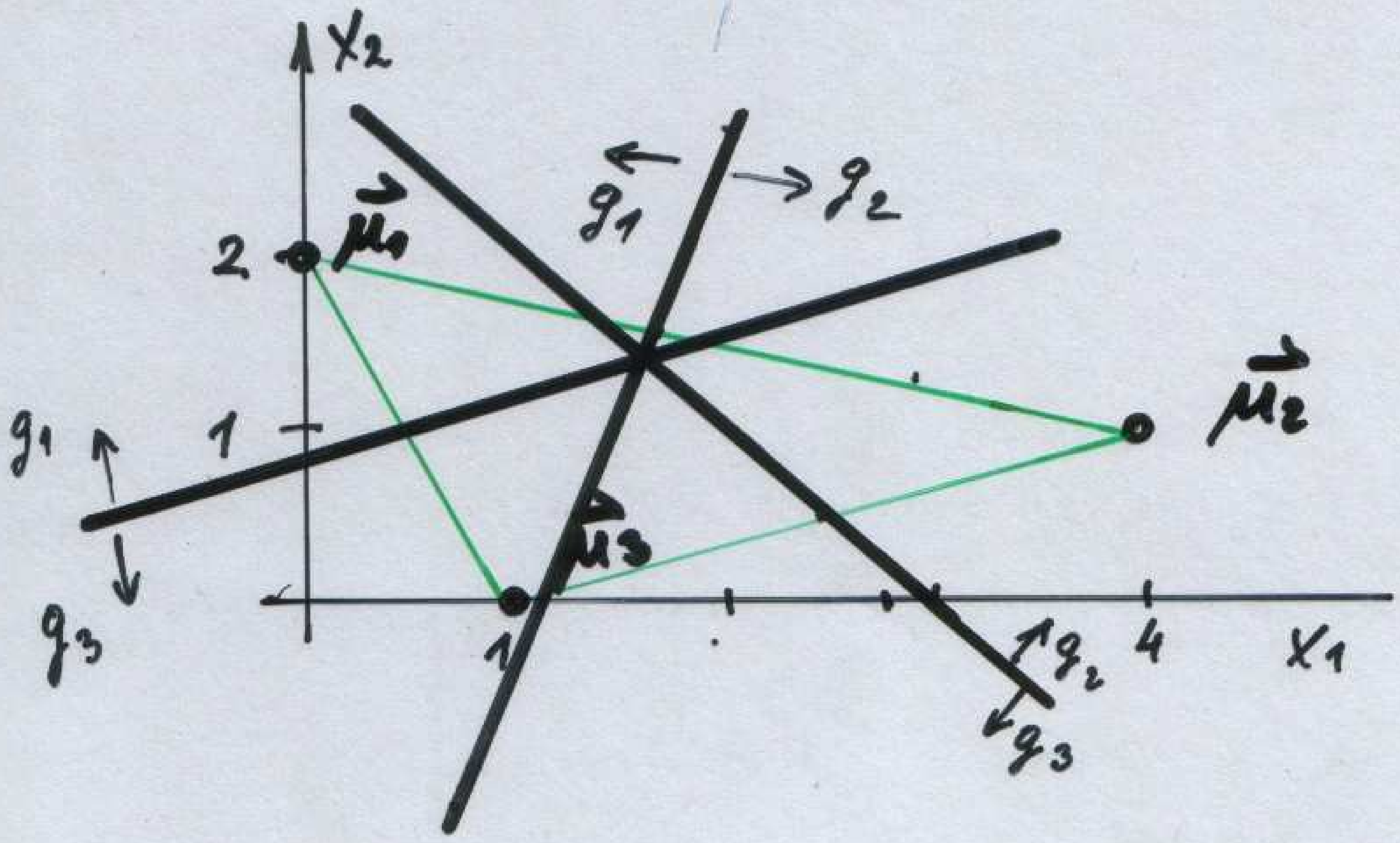
$$\vec{W}_{23} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

H_{23}: $6x_1 + 4x_2 - 17 = 0$

$$\vec{W}_{31} = \begin{pmatrix} -1 \\ 4 \end{pmatrix}$$

H_{31}: $2x_1 - 8x_2 + 7 = 0$

v16a



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 $\{\omega_1, \dots, \omega_c\}$ konačan skup od c stanja (razreda, kategorija)

i

3) uvođenje akcije umjesto samo odluke o stanju prirode

- skup $\{\alpha_1, \alpha_2, \dots, \alpha_a\}$ konačan skup od a mogućih akcija

i

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-komponentni vektor značajki

$p(\vec{x} | \omega_j)$ - funkcija uvjetne razdiobe gustoće vjerojatnosti