Kvadratni klasifikatori

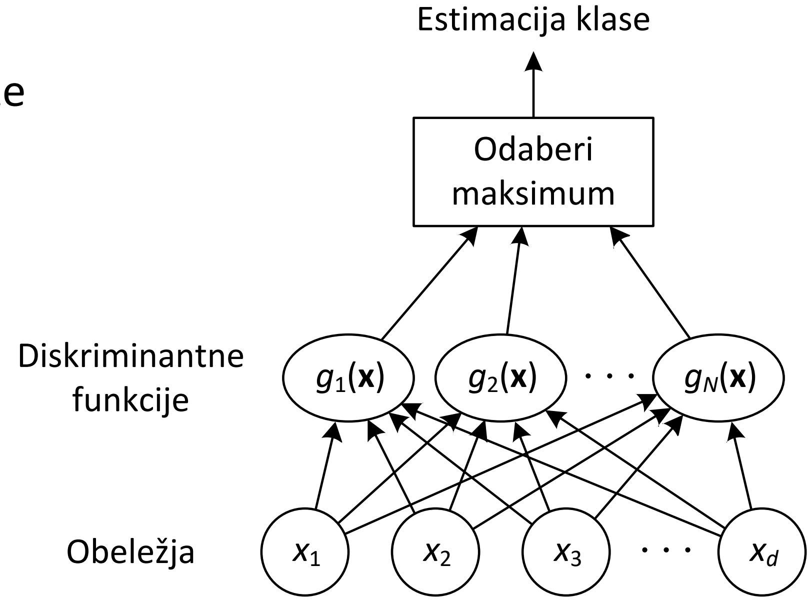
- Oblik diskriminantne funkcije u opštem slučaju
- Slučaj jednakih kovarijansnih matrica
- Karakteristični primeri
  - Slučaj 1: $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$
  - Slučaj 2: $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \sigma_2^2 \ldots \sigma_N^2)$
  - Slučaj 3: $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$
  - Slučaj 4: $\boldsymbol{\Sigma}_i = \sigma_i^2 \mathbf{I}$
  - Slučaj 5: $\boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j$

Diskriminantne funkcije

- MAP pravilo odlučivanja koje minimizuje verovatnoću greške može se formulisati preko familije diskriminantnih funkcija $g_i(\mathbf{x})$
  - Pravilo odlučivanja u tom slučaju glasi:

    „dodeli $\mathbf{x}$ klasi ω_i ako je $g_i(\mathbf{x}) \geq g_j(\mathbf{x})$ za svako $j \neq i$"

    gde je $g_i(\mathbf{x}) = P(\omega_i|\mathbf{x})$
  - U nekim slučajevima diskriminante se mogu svesti na jednostavne izraze
  - Ista monotona transformacija svih diskriminantnih funkcija ne utiče na rezultat, npr:
    - dodavanje konstante
    - logaritmovanje



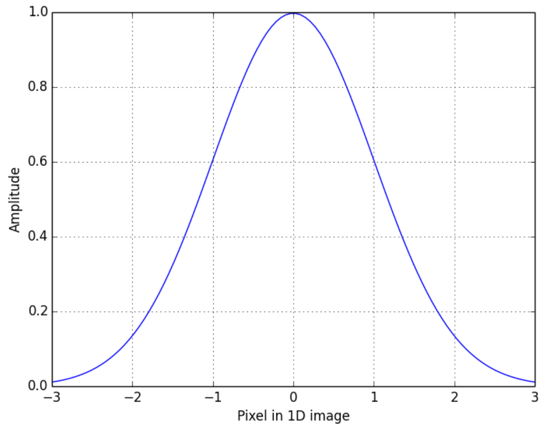
Normalna (Gaussova) raspodela

- Normalna (Gaussova) raspodela nad skalarnom slučajnom promenljivom $\mathcal{N}(\mu, \sigma^2)$ data je izrazom:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

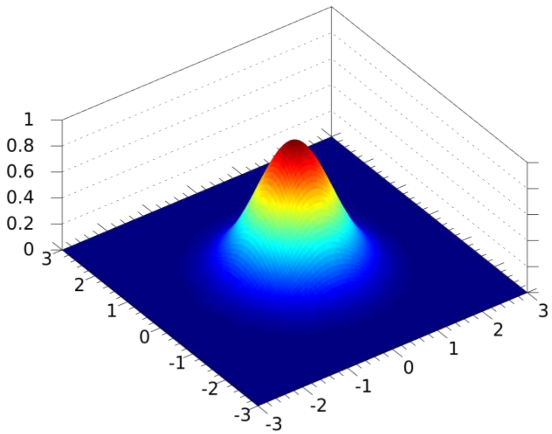
dok za vektorsku slučajnu promenljivu ima oblik:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} |\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$



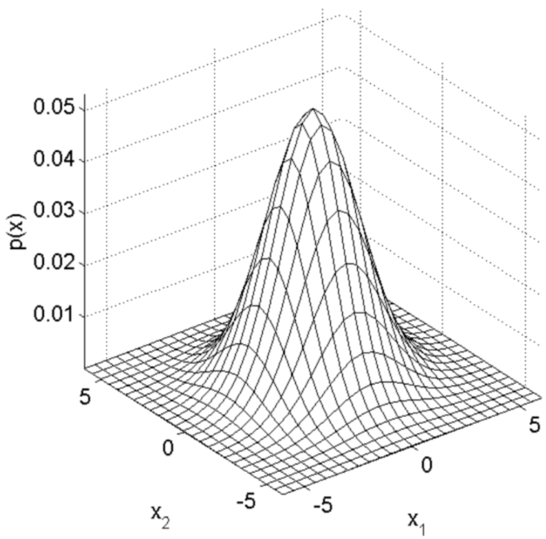
- Svojstva Gaussove raspodele:
  - Jedinstveno je opisana parametrima **μ** i **Σ**
  - Ako su obeležja međusobno nekorelisana ($c_{ik} = 0$), ona su takođe i *nezavisna*
    - Kovarijansna matrica je tada dijagonalna
  - Marginalne i uslovne raspodele su takođe normalne
  - Linearna transformacija **y** = **Ax** Gaussovog slučajnog vektora **y** takođe predstavlja Gaussov slučajni vektor

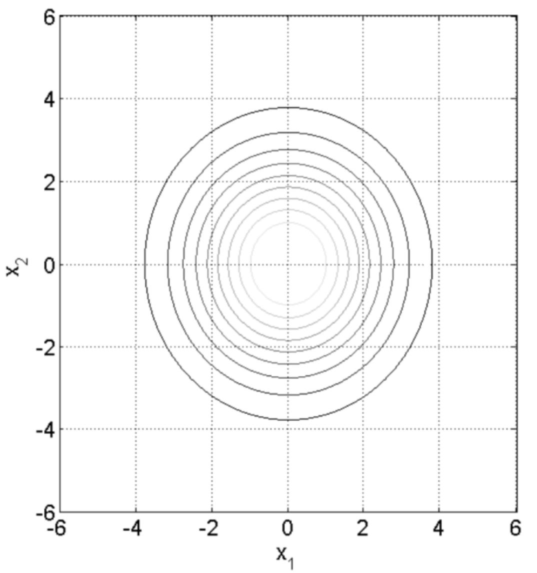
$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{f_{\mathbf{X}}(\mathbf{A}^{-1}\mathbf{y})}{|\mathbf{A}|}$$



Primeri dvodimenzionalne Gaussove raspodele (1)

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \qquad f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi |\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x}}$$

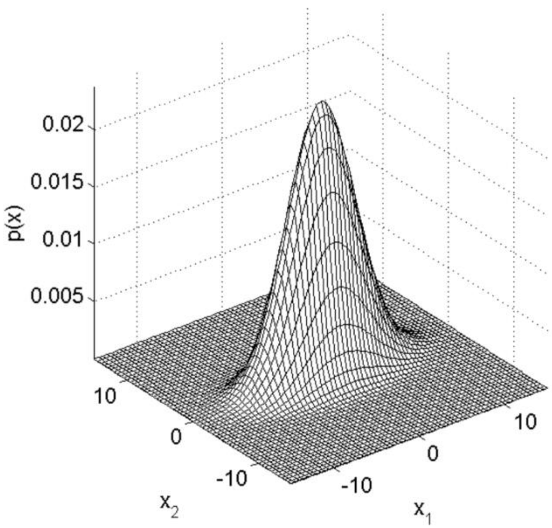


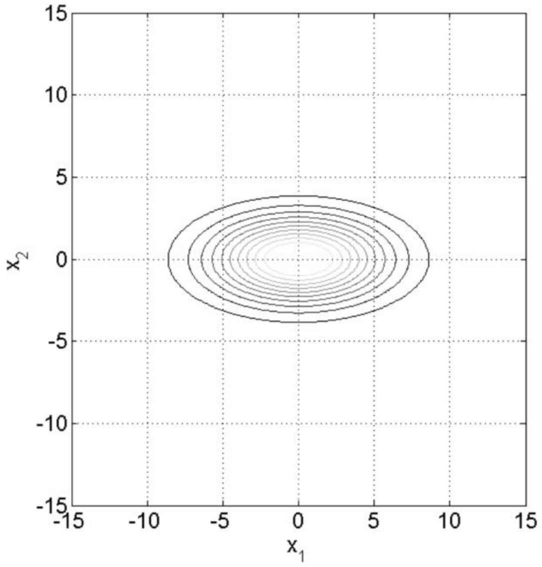


Izokrive su kružnice (ili u opštem slučaju hipersfere)

Primeri dvodimenzionalne Gaussove raspodele (2)

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 15 & 0 \\ 0 & 3 \end{bmatrix} \qquad f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi \,|\, \boldsymbol{\Sigma} \,|^{1/2}} e^{-\frac{1}{2}\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x}}$$

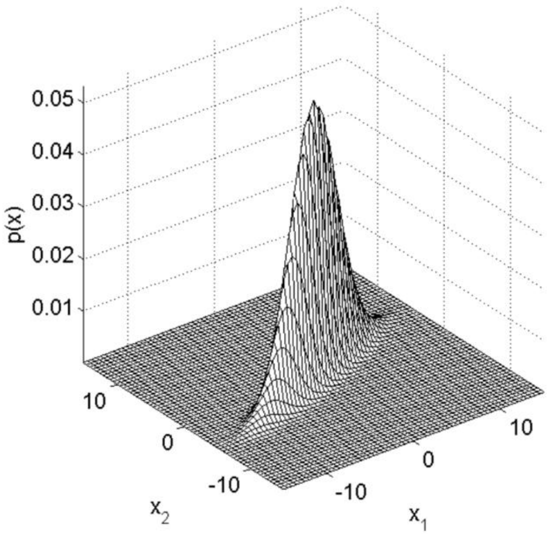


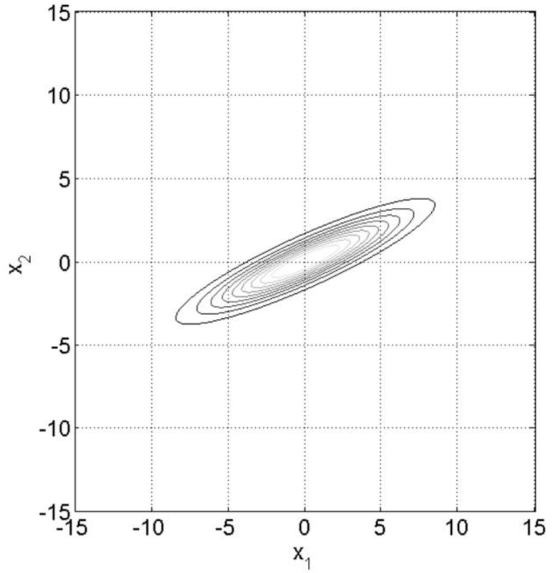


Izokrive su elipse čije su ose paralelne koordinatnim osama

Primeri dvodimenzionalne Gaussove raspodele (3)

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 15 & 6 \\ 6 & 3 \end{bmatrix} \qquad f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi |\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x}}$$





Izokrive su elipse čije su ose paralelne karakterističnim vektorima **Σ**

Bayesovi klasifikatori za klase sa Gaussovom raspodelom

- U slučaju Gaussove raspodele, diskriminantna funkcija za *i*-tu klasu je:

$$g_i(\mathbf{x}) = \mathrm{P}(\omega_i | \mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)\mathrm{P}(\omega_i)}{p(\mathbf{x})} = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}_i|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)} \cdot \frac{\mathrm{P}(\omega_i)}{p(\mathbf{x})}$$

odnosno, nakon logaritmovanja i uklanjanja članova koji su u svim klasama isti:

$$\begin{aligned} g_i(\mathbf{x}) &= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \ln \mathrm{P}(\omega_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| \\ &= -\frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \ln \mathrm{P}(\omega_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| \end{aligned}$$

- Prethodni izraz predstavlja kvadratnu funkciju po **x**, pa se ovakvi klasifikatori nazivaju *kvadratnim*

Kvadratni klasifikatori (primer)

- Naći diskriminantne funkcije i odrediti oblik odgovarajućih granica odlučivanja za slučaj dvodimenzionalnih uzoraka koji u svakoj od K klasa podležu Gaussovoj raspodeli opisanoj sledećim parametrima:

$$\boldsymbol{\mu}_i = \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \end{bmatrix} \qquad \boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_i^2 & 0 \\ 0 & \sigma_i^2 \end{bmatrix} \qquad i = 1,2,\ldots,K$$

- Rešenje:
  - Pošto je inverzna matrica kovarijansne matrice jednaka $\boldsymbol{\Sigma}_i^{-1} = \begin{bmatrix} 1/\sigma_i^2 & 0 \\ 0 & 1/\sigma_i^2 \end{bmatrix}$, opšti izraz za diskriminantnu funkciju može se uprostiti:

$$g_i(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \ln \mathrm{P}(\omega_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i|$$

$$= -\frac{x_1^2 + x_2^2}{2\sigma_i^2} + \frac{x_1\mu_{i1} + x_2\mu_{i2}}{\sigma_i^2} - \frac{\mu_{i1}^2 + \mu_{i2}^2}{2\sigma_i^2} + \ln \mathrm{P}(\omega_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i|$$

$$= -\frac{1}{2\sigma_i^2}((x_1 - \mu_{i1})^2 + (x_2 - \mu_{i2})^2) + \ln \mathrm{P}(\omega_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i|$$

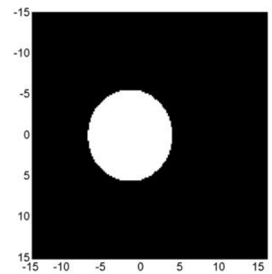
što znači da su granice odlučivanja $g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0$ krive drugog reda (elipse, parabole ili hiperbole)

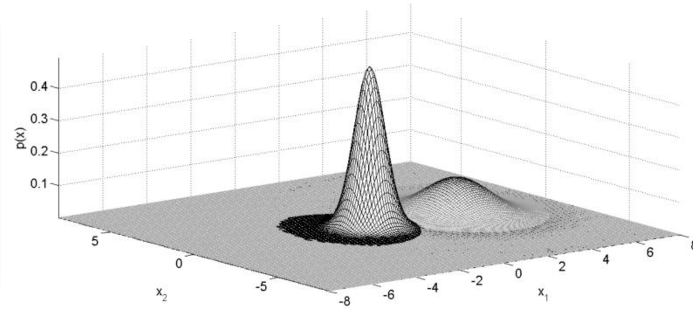
Kvadratni klasifikatori (primer)

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

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0{,}3 & 0 \\ 0 & 0{,}35 \end{bmatrix} \qquad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1{,}2 & 0 \\ 0 & 1{,}85 \end{bmatrix}$$

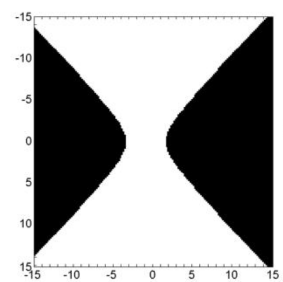


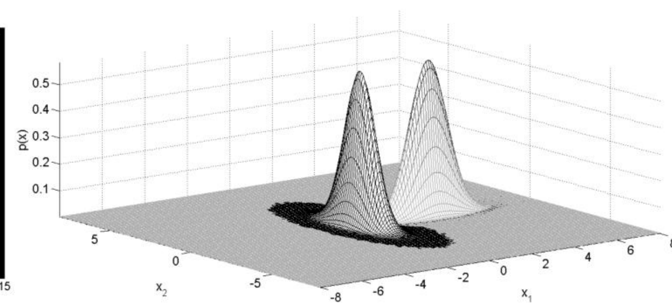


$$P(\omega_1) = P(\omega_2)$$

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 3{,}2 \\ 0 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0{,}1 & 0 \\ 0 & 0{,}75 \end{bmatrix} \qquad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 0{,}75 & 0 \\ 0 & 0{,}1 \end{bmatrix}$$





Slučajevi istih kovarijansnih matrica

- U opštem izrazu za diskriminantnu funkciju:

$$g_i(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}_i^{-1}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \frac{1}{2}\boldsymbol{\mu}_i^T\mathbf{\Sigma}_i^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^T\mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \ln \mathrm{P}(\omega_i) - \frac{1}{2}\ln|\mathbf{\Sigma}_i|$$

  jedini kvadratni član je $-\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}_i^{-1}\mathbf{x}$, i ako je $\mathbf{\Sigma}_i$ ista u svim klasama, on nestaje (a s njim i član $-\frac{1}{2}\ln|\mathbf{\Sigma}_i|$), i diskriminantna funkcija postaje *linearna*
  - granice odlučivanja postaju *hiperravni*
- Diskriminantna funkcija svodi se na:

$$g_i(\mathbf{x}) = \mathbf{x}^T\mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\boldsymbol{\mu}_i^T\mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \ln \mathrm{P}(\omega_i)$$

$$= \mathbf{w}_i^T\mathbf{x} + w_{i0}\,,$$

  pri čemu je:

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

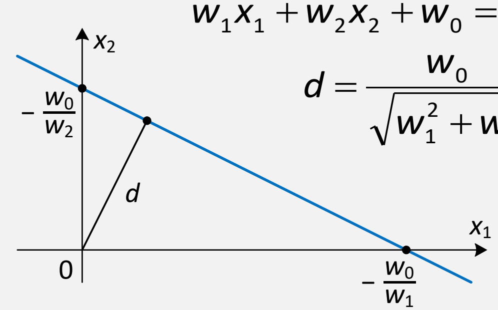
$$w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^T\mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \ln \mathrm{P}(\omega_i)$$

Jednačina prave u ravni
(podsećanje)

$$\mathbf{w}^T\mathbf{x} + w_0 = 0$$

$$w_1x_1 + w_2x_2 + w_0 = 0$$

$$d = \frac{w_0}{\sqrt{w_1^2 + w_2^2}}$$



Slučaj 1: $\mathbf{\Sigma}_i = \mathbf{\Sigma} = \sigma^2\mathbf{I}$

- Iste kovarijansne matrice u svim klasama
- Međusobno nekorelisana, a time i nezavisna obeležja
- Jednake varijanse svih obeležja

$$g_i(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{\Sigma}_i^{-1}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\mathbf{\Sigma}_i^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \ln P(\omega_i) - \frac{1}{2}\ln|\mathbf{\Sigma}_i|$$

Slučaj 1: $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = \sigma^2\mathbf{I}$

- Iste kovarijansne matrice u svim klasama
- Međusobno nekorelisana, a time i nezavisna obeležja
- Jednake varijanse svih obeležja

- Diskriminantna funkcija svodi se na:

$$g_i(\mathbf{x}) = \mathbf{w}_i^{\mathrm{T}}\mathbf{x} + w_{i0}\,,$$

pri čemu je:

$$\mathbf{w}_i = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i = \frac{\boldsymbol{\mu}_i}{\sigma^2}, \quad w_{i0} = -\frac{1}{2\sigma^2}\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\mu}_i + \ln \mathrm{P}(\omega_i)$$

$$g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0 \;\Rightarrow\; (\mathbf{w}_i - \mathbf{w}_j)^{\mathrm{T}}\mathbf{x} + w_{i0} - w_{j0} = 0$$

$$\Rightarrow\; (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}\mathbf{x} + \sigma^2(w_{i0} - w_{j0}) = 0$$

Slučaj 1: $\mathbf{\Sigma}_i = \mathbf{\Sigma} = \sigma^2\mathbf{I}$

- Iste kovarijansne matrice u svim klasama
- Međusobno nekorelisana, a time i nezavisna obeležja
- Jednake varijanse svih obeležja

- Diskriminantna funkcija svodi se na:

$$g_i(\mathbf{x}) = \mathbf{w}_i^{\mathrm{T}}\mathbf{x} + w_{i0}\,,$$

pri čemu je:

$$\mathbf{w}_i = \mathbf{\Sigma}^{-1}\boldsymbol{\mu}_i = \frac{\boldsymbol{\mu}_i}{\sigma^2}, \quad w_{i0} = -\frac{1}{2\sigma^2}\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\mu}_i + \ln \mathrm{P}(\omega_i)$$

- Granice odlučivanja su *hiperravni*:

$$g_{ij}(\mathbf{x}) = g_i(\mathbf{x}) - g_j(\mathbf{x}) = \boldsymbol{\mu}_{ij}^{\mathrm{T}}(\mathbf{x} - \mathbf{x}_0) = 0\,,$$

pri čemu je $\boldsymbol{\mu}_{ij}$ normala na $g_{ij}(\mathbf{x})$, i $g_{ij}(\mathbf{x})$ prolazi kroz tačku $\mathbf{x}_0$:

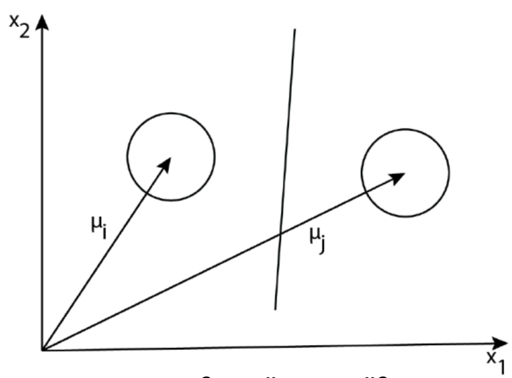
$$\boldsymbol{\mu}_{ij} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j\,, \quad \mathbf{x}_0 = \frac{\boldsymbol{\mu}_i + \boldsymbol{\mu}_j}{2} - \sigma^2 \frac{\boldsymbol{\mu}_i + \boldsymbol{\mu}_j}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2} \cdot \ln \frac{\mathrm{P}(\omega_1)}{\mathrm{P}(\omega_2)}$$

Slučaj 1: $\mathbf{\Sigma}_i = \mathbf{\Sigma} = \sigma^2\mathbf{I}$

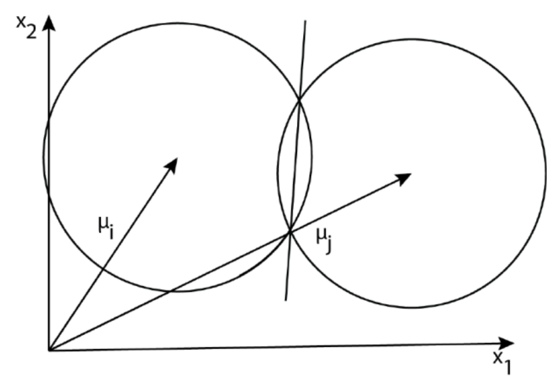
- Ako su klase jednako verovatne, hiperravan odlučivanja prolazi kroz tačku koja predstavlja sredinu između $\boldsymbol{\mu}_i$ i $\boldsymbol{\mu}_j$, i diskriminantna funkcija se može svesti na:

$$g_i(\mathbf{x}) = -\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}(\mathbf{x}-\boldsymbol{\mu}_i),$$

  što predstavlja *klasifikator na bazi minimalnog rastojanja od srednje vrednosti*
- Ako je varijansa σ^2 mala u poređenju sa $\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2$ (slika levo):
  - položaj granice odlučivanja ne zavisi značajno od odnosa $P(\omega_i)$ i $P(\omega_j)$



$\sigma^2 \ll \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2$



$\sigma^2 \not\ll \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2$

Slučaj 1: $\mathbf{\Sigma}_i = \mathbf{\Sigma} = \sigma^2\mathbf{I}$ (primer)

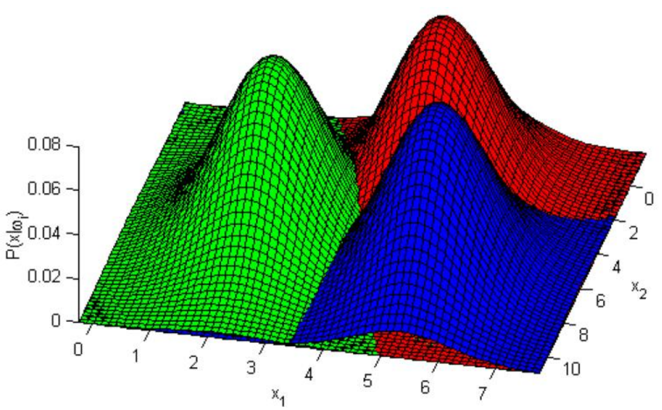
- Dvodimenzionalni vektor obeležja, tri klase:

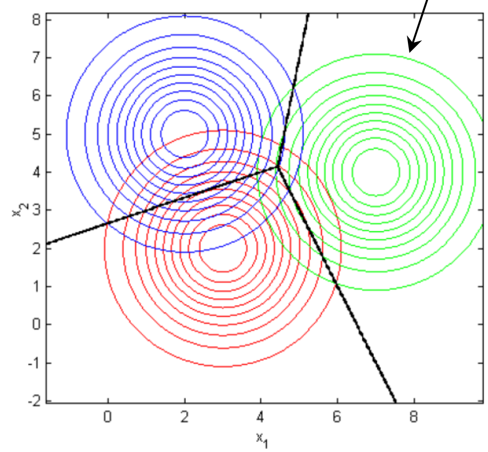
$$P(\omega_1) = P(\omega_2) = P(\omega_3)$$

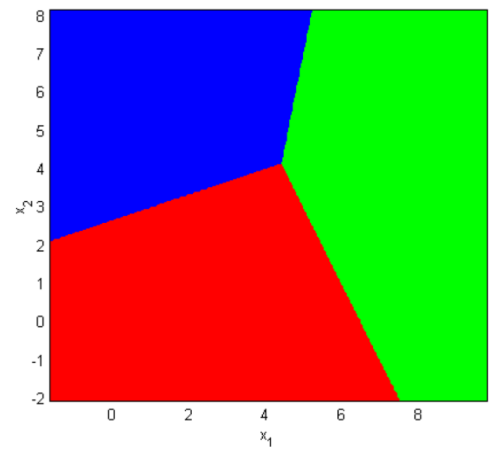
$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 4 \end{bmatrix} \qquad \boldsymbol{\mu}_3 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$\mathbf{\Sigma}_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \qquad \mathbf{\Sigma}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \qquad \mathbf{\Sigma}_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Geometrijska mesta tačaka s jednakom verovatnoćom su kružnice







Slučaj 2: $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$

- Iste kovarijansne matrice u svim klasama
- Međusobno nekorelisana, a time i nezavisna obeležja
- **Različite varijanse pojedinih obeležja**

- Diskriminantna funkcija može se napisati u obliku:

$$
\begin{aligned}
g_i(\mathbf{x}) &= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln \mathrm{P}(\omega_i) \\
&= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}\begin{bmatrix}\sigma_1^{-2} & & \\ & \ddots & \\ & & \sigma_d^{-2}\end{bmatrix}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{1}{2}\ln\begin{vmatrix}\sigma_1^{2} & & \\ & \ddots & \\ & & \sigma_d^{2}\end{vmatrix} + \ln \mathrm{P}(\omega_i) \\
&= -\frac{1}{2}\sum_{k=1}^{d}\frac{(x_k-\mu_{ik})^2}{\sigma_k^2} - \frac{1}{2}\ln\prod_{k=1}^{d}\sigma_k^2 + \ln \mathrm{P}(\omega_i) \\
&= -\frac{1}{2}\sum_{k=1}^{d}\frac{x_k^2 - 2x_k\mu_{ik} + \mu_{ik}^2}{\sigma_k^2} - \frac{1}{2}\ln\prod_{k=1}^{d}\sigma_k^2 + \ln \mathrm{P}(\omega_i)
\end{aligned}
$$

Slučaj 2: $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$

- Nakon eliminacije članova koji su u svim klasama isti, dobija se:

$$g_i(\mathbf{x}) = -\frac{1}{2}\sum_{k=1}^{d}\frac{-2x_k\mu_{ik} + \mu_{ik}^2}{\sigma_k^2} + \ln P(\omega_i)$$

- Dobijena diskriminanta je ponovo linearna, odnosno, granice odlučivanja su ponovo hiperravni
  - Geometrijska mesta tačaka sa konstantnom verovatnoćom su hiperelipsoidi čije su ose paralelne sa koordinatnim osama
  - Ako su klase jednako verovatne, i član $\ln P(\omega_i)$ može se izostaviti

Slučaj 2: $\mathbf{\Sigma}_i = \mathbf{\Sigma} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$ (primer)

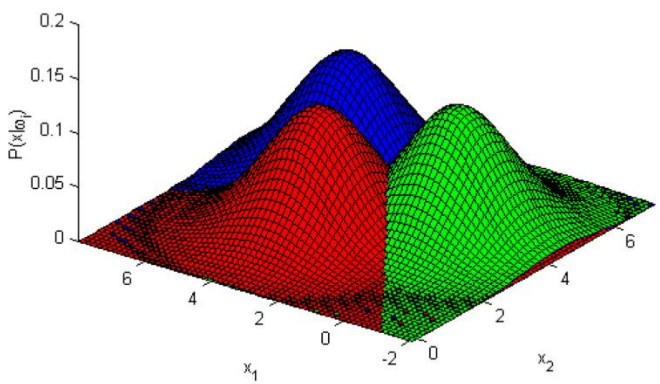
- Dvodimenzionalni vektor obeležja, tri klase:

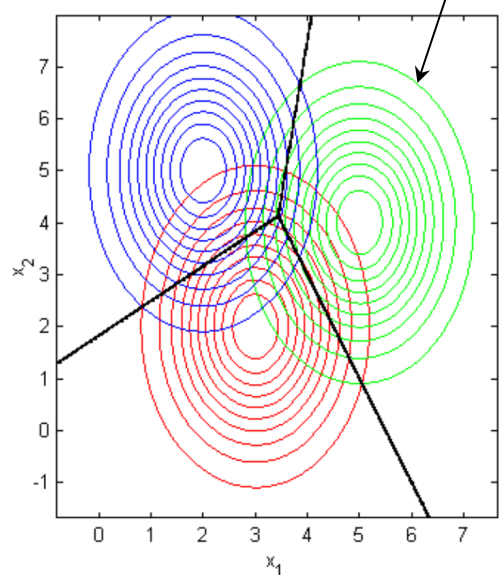
$$P(\omega_1) = P(\omega_2) = P(\omega_3)$$

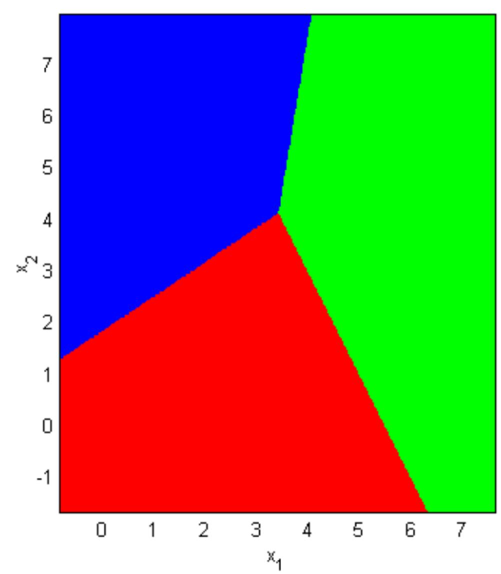
$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 5 \\ 4 \end{bmatrix} \qquad \boldsymbol{\mu}_3 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$\mathbf{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \qquad \mathbf{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \qquad \mathbf{\Sigma}_3 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

Geometrijska mesta tačaka s jednakom verovatnoćom su elipse







Slučaj 3: $\mathbf{\Sigma}_i = \mathbf{\Sigma} \neq \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$

- Iste kovarijansne matrice u svim klasama
- Različite varijanse pojedinih obeležja
- **Postoji korelacija između pojedinih obeležja**

- Diskriminantna funkcija mogla bi se uprostiti kao i u prethodnom slučaju:

$$g_i(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{\Sigma}_i^{-1}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\mathbf{\Sigma}_i^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \ln \mathrm{P}(\omega_i) - \frac{1}{2}\ln|\mathbf{\Sigma}_i|$$

$$g_i(\mathbf{x}) = \boldsymbol{\mu}_i^{\mathrm{T}}\mathbf{\Sigma}_i^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \ln \mathrm{P}(\omega_i)$$

odnosno:

$$g_i(\mathbf{x}) = \mathbf{w}_i^{\mathrm{T}}\mathbf{x} + w_{i0}, \quad \mathbf{w}_i = \mathbf{\Sigma}^{-1}\boldsymbol{\mu}_i, \quad w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_i + \ln \mathrm{P}(\omega_i)$$

što znači da su granice odlučivanja su ponovo hiperravni

  - Geometrijsko mesto tačaka s konstantnom verovatnoćom su hiperelipsoidi čije su ose paralelne karakterističnim vektorima kovarijansne matrice $\mathbf{\Sigma}$

Slučaj 3: $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} \neq \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$

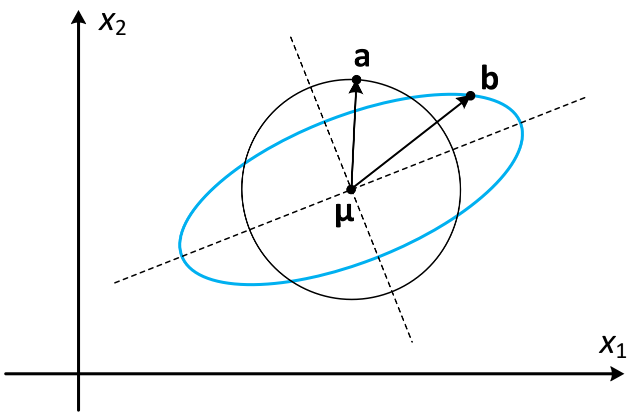
- U ovom slučaju zanimljiva je i interpretacija diskriminantne funkcije kod koje se kvadratni član ostavlja:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \ln P(\omega_i)$$

- Kvadratni član $(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)$ odgovara kvadratu tzv. *Mahalanobisovog rastojanja* vektora **x** od $\boldsymbol{\mu}_i$:

$$\boxed{\|\mathbf{x}-\mathbf{y}\|^2_{\boldsymbol{\Sigma}^{-1}} = (\mathbf{x}-\mathbf{y})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{y})}$$

- Mahalanobisovo rastojanje koristi tzv. $\boldsymbol{\Sigma}^{-1}$ normu
  - $\boldsymbol{\Sigma}^{-1}$ se može posmatrati kao neizotropni skalirajući faktor
  - Za $\boldsymbol{\Sigma} = \mathbf{I}$ Mahalanobisovo rastojanje svodi se na euklidsko
- Za klase s jednakim verovatnoćama dobija se *klasifikator na bazi minimalnog Mahalanobisovog rastojanja* od srednje vrednosti



$$\|\mathbf{a}-\boldsymbol{\mu}\|_2 = C$$

$$\|\mathbf{b}-\boldsymbol{\mu}\|_{\boldsymbol{\Sigma}^{-1}} = C$$

Slučaj 3: $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} \neq \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$ (primer)

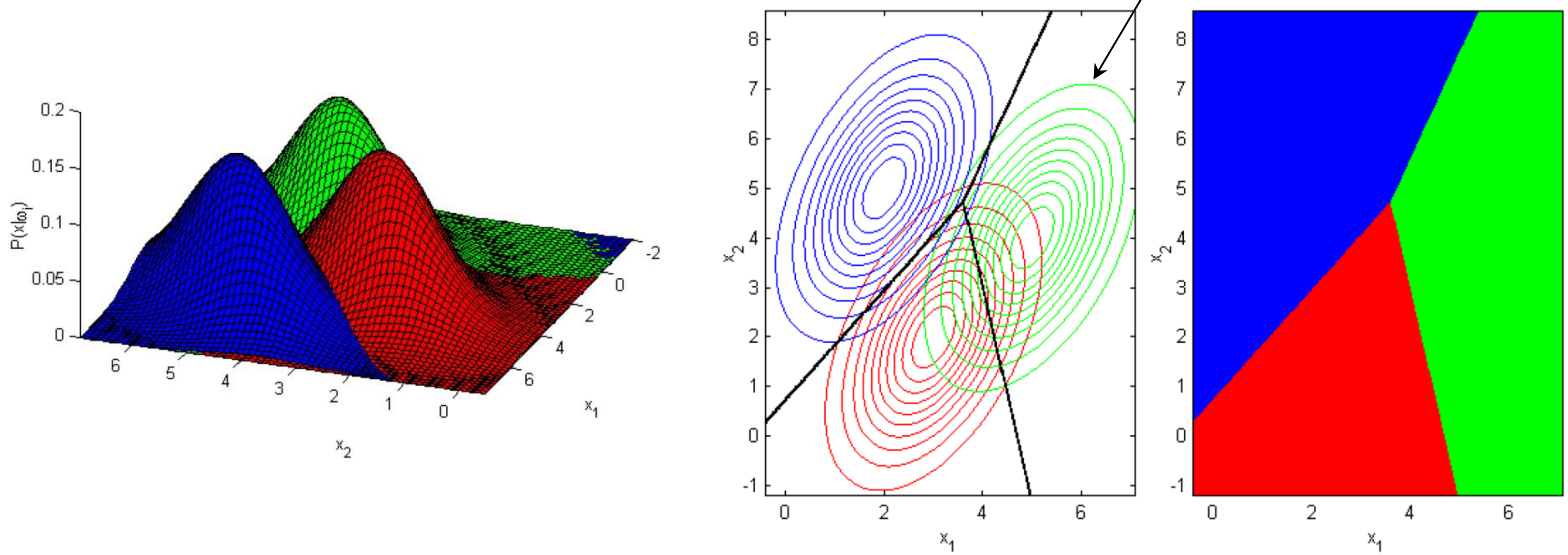
- Dvodimenzionalni vektor obeležja, tri klase:

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

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 5 \\ 4 \end{bmatrix} \qquad \boldsymbol{\mu}_3 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 2 \end{bmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 2 \end{bmatrix} \quad \boldsymbol{\Sigma}_3 = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 2 \end{bmatrix}$$

Granica odlučivanja između klasa ω_i i ω_j više nije normalna na $\boldsymbol{\mu}_{ij} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j$, već na $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{ij}$



Slučajevi različitih kovarijansnih matrica

- U opštem izrazu za diskriminantnu funkciju:

$$g_i(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \ln \mathrm{P}(\omega_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i|$$

  sada nije moguće izostaviti kvadratni član $-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x}$ jer kovarijansna matrica $\boldsymbol{\Sigma}_i$ nije ista u svim klasama

  - diskriminantne funkcije ne mogu se svesti na linearne

Slučaj 4: $\boldsymbol{\Sigma}_i = \sigma_i^2\mathbf{I}$

- Različite kovarijansne matrice u pojedinim klasama
- Svaka od njih je dijagonalna, a varijanse svih obeležja unutar jedne klase su jednake

- Diskriminantna funkcija svodi se na:

$$
\begin{aligned}
g_i(\mathbf{x}) &= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \ln \mathrm{P}(\omega_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| \\
&= -\frac{1}{2\sigma_i^2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}(\mathbf{x}-\boldsymbol{\mu}_i) + \ln \mathrm{P}(\omega_i) - \frac{N}{2}\ln \sigma_i^2
\end{aligned}
$$

  i dobijeni izraz se ne može dalje uprostiti
- Granice odlučivanja su hiperelipsoidi

Slučaj 4: $\mathbf{\Sigma}_i = \sigma_i^2\mathbf{I}$ (primer)

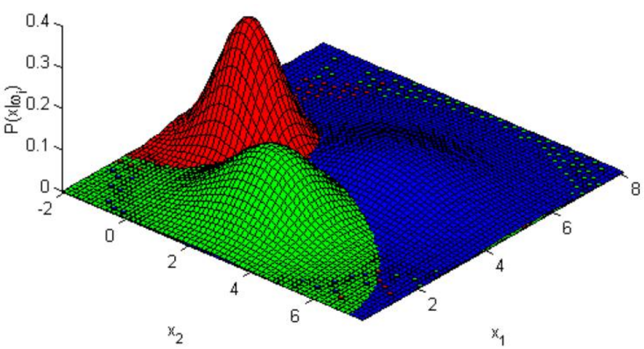
- Dvodimenzionalni vektor obeležja, tri klase:

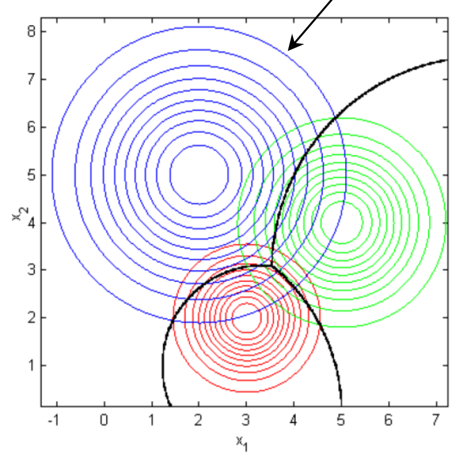
$$P(\omega_1) = P(\omega_2) = P(\omega_3)$$

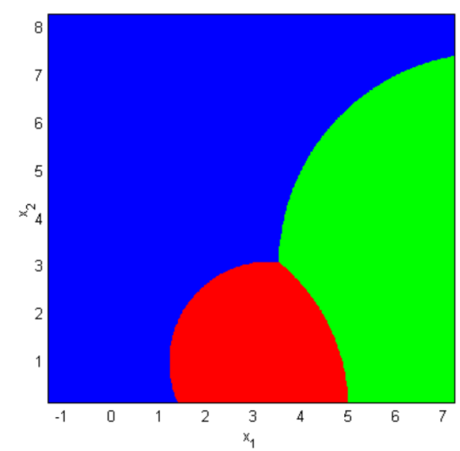
$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 5 \\ 4 \end{bmatrix} \qquad \boldsymbol{\mu}_3 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$\mathbf{\Sigma}_1 = \begin{bmatrix} 0,5 & 0 \\ 0 & 0,5 \end{bmatrix} \qquad \mathbf{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{\Sigma}_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Geometrijska mesta tačaka s jednakom verovatnoćom su kružnice







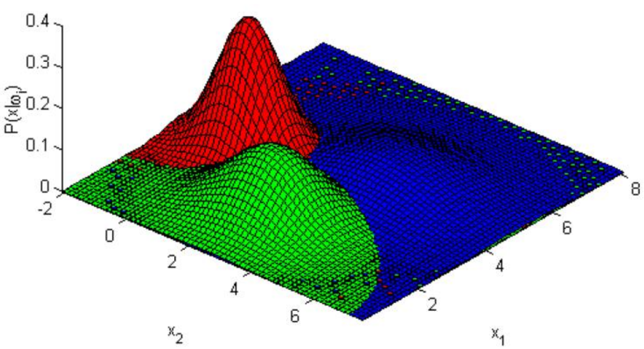
Slučaj 4: $\mathbf{\Sigma}_i = \sigma_i^2\mathbf{I}$ (primer)

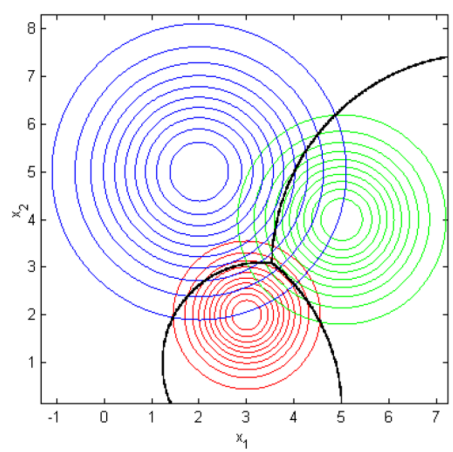
- Dvodimenzionalni vektor obeležja, tri klase:

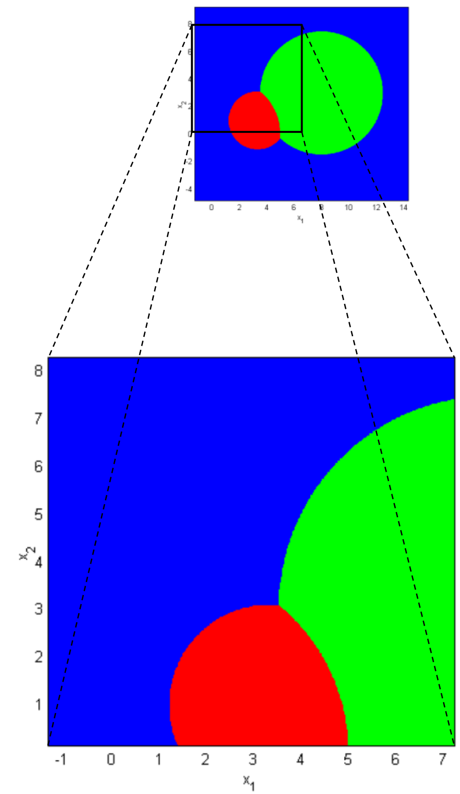
$$P(\omega_1) = P(\omega_2) = P(\omega_3)$$

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 5 \\ 4 \end{bmatrix} \qquad \boldsymbol{\mu}_3 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$\mathbf{\Sigma}_1 = \begin{bmatrix} 0{,}5 & 0 \\ 0 & 0{,}5 \end{bmatrix} \qquad \mathbf{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{\Sigma}_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$







Slučaj 5: $\boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j$

- Najopštiji slučaj
- Različite klase mogu imati različite kovarijansne matrice
- Kovarijansne matrice pojedinih klasa ne moraju biti dijagonalne

- Opšti izraz za diskriminantnu funkciju:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \ln \mathrm{P}(\omega_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i|$$

ne može se dalje uprostiti, ali se može predstaviti u kvadratnom obliku:

$$g_i(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{W}_i\mathbf{x} + \mathbf{w}_i^{\mathrm{T}}\mathbf{x} + w_{i0},$$

$$\mathbf{W}_i = -\frac{1}{2}\boldsymbol{\Sigma}_i^{-1}, \quad \mathbf{w}_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i, \quad w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln\mathrm{P}(\omega_i)$$

  - Geometrijska mesta tačaka sa konstantnom verovatnoćom za svaku klasu ω_i su hiperelipsoidi čije su ose paralelne karakterističnim vektorima kovarijansne matrice $\boldsymbol{\Sigma}_i$
  - Granice regiona odlučivanja su opet kvadratne funkcije: hiperelipsoidi ili hiperparaboloidi
  - Kvadratni član u diskriminantnoj funkciji proporcionalan je Mahalanobisovom rastojanju za klasnu kovarijansnu matricu $\boldsymbol{\Sigma}_i$

Slučaj 5: $\boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j$ (primer)

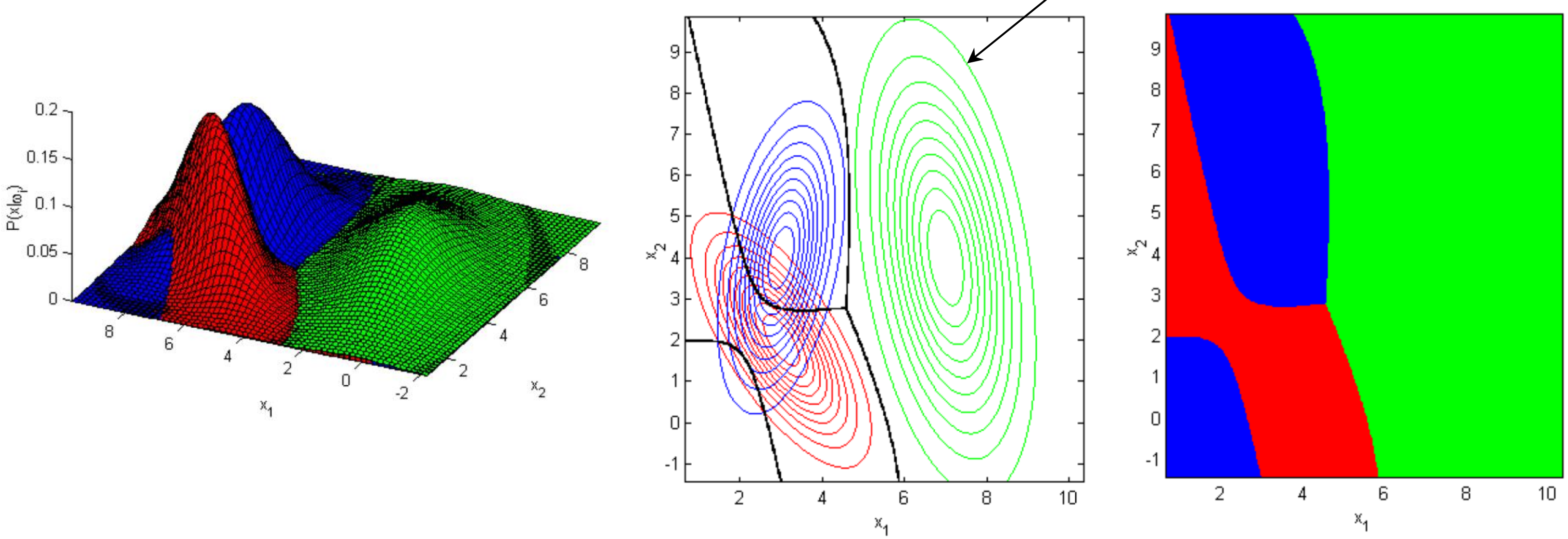
- Dvodimenzionalni vektor obeležja, tri klase:

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

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 5 \\ 4 \end{bmatrix} \qquad \boldsymbol{\mu}_3 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & -1 \\ -1 & 7 \end{bmatrix} \quad \boldsymbol{\Sigma}_3 = \begin{bmatrix} 0{,}5 & 0{,}5 \\ 0{,}5 & 3 \end{bmatrix}$$

Geometrijska mesta tačaka s jednakom verovatnoćom su hiperelipsoidi



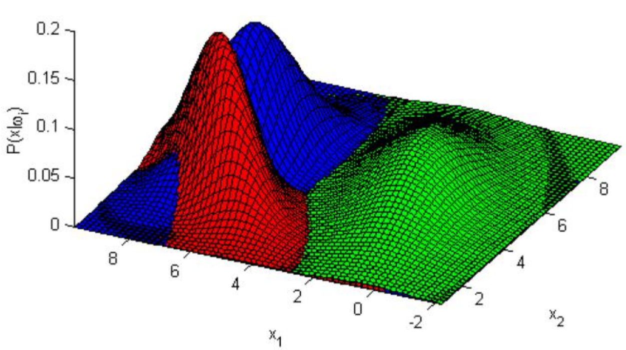
Slučaj 5: $\boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j$ (primer)

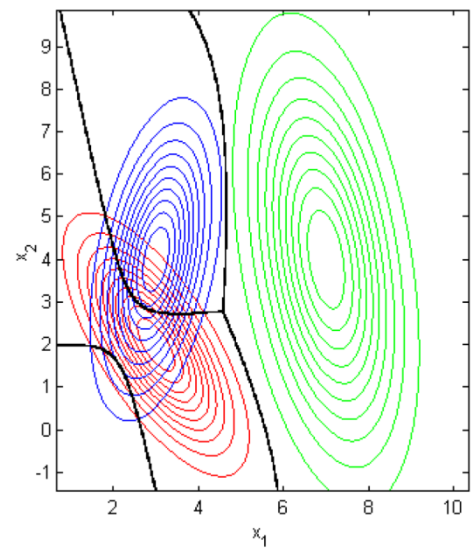
- Dvodimenzionalni vektor obeležja, tri klase:

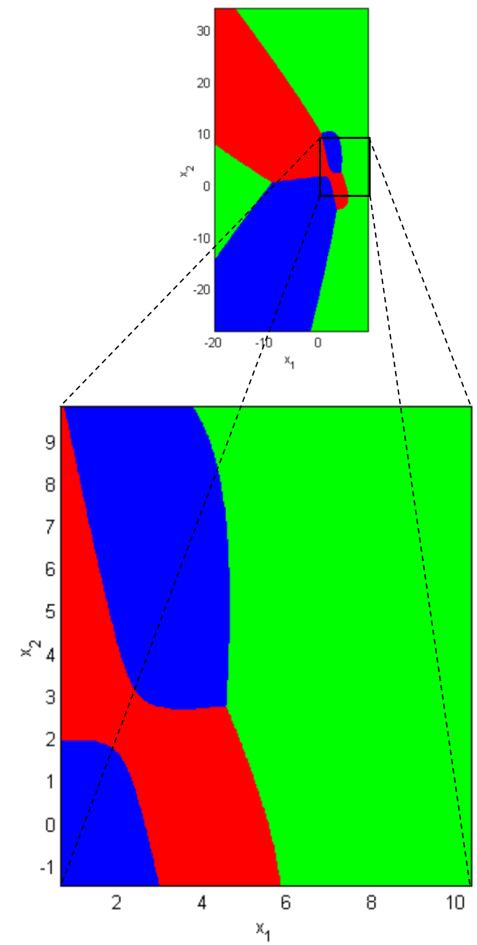
$$P(\omega_1) = P(\omega_2) = P(\omega_3)$$

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 5 \\ 4 \end{bmatrix} \qquad \boldsymbol{\mu}_3 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & -1 \\ -1 & 7 \end{bmatrix} \quad \boldsymbol{\Sigma}_3 = \begin{bmatrix} 0,5 & 0,5 \\ 0,5 & 3 \end{bmatrix}$$







Primer

- Odrediti diskriminantnu funkciju za dve klase sa trodimenzionalnom Gaussovom raspodelom, definisane sledećim podacima:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{bmatrix}, \quad \mathrm{P}(\omega_2) = 2\mathrm{P}(\omega_1)$$

- Rešenje:

$$g_i(\mathbf{x}) = -\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}(\mathbf{x}-\boldsymbol{\mu}_i) + \ln \mathrm{P}(\omega_i) = -2\begin{bmatrix} x_1-\mu_{1i} \\ x_2-\mu_{2i} \\ x_3-\mu_{3i} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} x_1-\mu_{1i} \\ x_2-\mu_{2i} \\ x_3-\mu_{3i} \end{bmatrix} + \ln \mathrm{P}(\omega_i)$$

$$g_1(\mathbf{x}) = -2\begin{bmatrix} x_1-0 \\ x_2-0 \\ x_3-0 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} x_1-0 \\ x_2-0 \\ x_3-0 \end{bmatrix} + \ln\frac{1}{3} = -2(x_1^2 + x_2^2 + x_3^2) + \ln\frac{1}{3}$$

$$g_2(\mathbf{x}) = -2\begin{bmatrix} x_1-1 \\ x_2-1 \\ x_3-1 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} x_1-1 \\ x_2-1 \\ x_3-1 \end{bmatrix} + \ln\frac{2}{3} = -2((x_1-1)^2 + (x_2-1)^2 + (x_3-1)^2) + \ln\frac{2}{3}$$

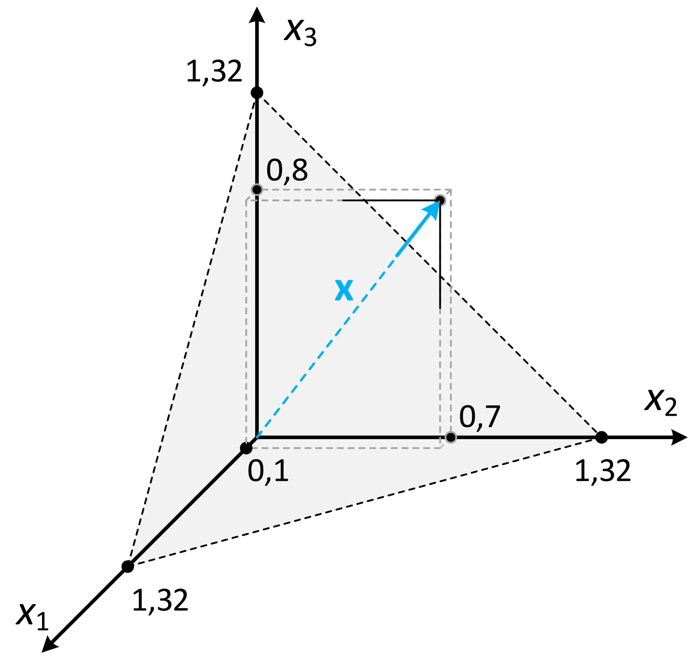
Primer

$$g_1(\mathbf{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} g_2(\mathbf{x}) \;\Leftrightarrow\; -2(x_1^2 + x_2^2 + x_3^2) + \ln\frac{1}{3} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} -2((x_1-1)^2 + (x_2-1)^2 + (x_3-1)^2) + \ln\frac{2}{3}$$

$$\Leftrightarrow\; x_1 + x_2 + x_3 \underset{\omega_1}{\overset{\omega_2}{\gtrless}} \frac{6-\ln 2}{4} = 1{,}32$$

- Jednakost $x_1 + x_2 + x_3 = 1{,}32$ određuje ravan koja predstavlja granicu odlučivanja
  - Kada bi bilo potrebno klasifikovati npr. uzorak $\mathbf{x} = [0{,}1\;\; 0{,}7\;\; 0{,}8]^T$, trebalo bi uraditi sledeće:

$$0{,}1 + 0{,}7 + 0{,}8 = 1{,}6 \underset{\omega_1}{\overset{\omega_2}{\gtrless}} 1{,}32 \Rightarrow \mathbf{x} \in \omega_2$$



Zaključci

- Bayesov klasifikator za klase sa normalnom raspodelom je u opštem slučaju *kvadratni klasifikator*
- Ako su kovarijansne matrice različitih klasa jednake, klasifikator je *linearan*
- Korišćenje bilo kog od opisanih klasifikatora počiva na određenim statističkim pretpostavkama
  - Na pitanje da li su te pretpostavke tačne retko se može odgovoriti u praksi
  - U većini slučajeva može se samo odgovoriti na pitanje: „da li klasifikator rešava problem ili ne?“