$f_k(\vec{x})$ - the class-conditional density of $\vec{X}$ in class $\{G = k\}$

π_k - the prior probability of class k

не се знаят, избират се, в зависимост от избора може да се смени резултатът

$$\sum_{k=1}^{K} \pi_k = 1$$

Bayes theorem:

$$P(G = k \mid \vec{X} = \vec{x}) = \frac{\pi_k f_k(\vec{x})}{\sum_{\ell=1}^{K} f_\ell(\vec{x}) \pi_\ell}$$

Many techniques are based on models for the class densities:

- linear & quadratic discriminant analysis with gaussian densities
- mixtures of gaussian allow for nonlinear boundaries
- nonparametric density estimates for each class
- Naive Bayes - the inputs are conditionally $\perp\!\!\!\perp$ in each class

Now let us consider the multivariate gaussian

$$f_k(x) = \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} e^{-\frac{1}{2}(\vec{x} - \vec{\mu}_k)' \Sigma_k^{-1} (\vec{x} - \vec{\mu}_k)}$$

Here Σ_k is the covariance matrix

$\vec{\mu}_k$ is the mean vector

$p \to R^p$ - размерност на пр-вото от данни (не на обем на извадката!)

Linear discriminant analysis: (LDA)

$$\Sigma_k = \Sigma \quad \forall k$$

Compare two classes k & ℓ!

$$\log \frac{P(G=k \mid \vec{X}=\vec{x})}{P(G=\ell \mid \vec{X}=\vec{x})} = \log \frac{\pi_k f_k(\vec{x})}{\cancel{\sum_{j=1}^{K} \pi_j f_j(\vec{x})}} \cdot \frac{\cancel{\sum_{j=1}^{K} \pi_j f_j(\vec{x})}}{\pi_\ell f_\ell(\vec{x})} =$$

$$= \log \frac{f_k(\vec{x})}{f_\ell(\vec{x})} + \log \frac{\pi_k}{\pi_\ell} =$$

$$= \log \frac{\pi_k}{\pi_\ell} + \frac{1}{2}(\vec{x}-\vec{\mu}_\ell)' \Sigma^{-1} (\vec{x}-\vec{\mu}_\ell) - \frac{1}{2}(\vec{x}-\vec{\mu}_k)' \Sigma^{-1} (\vec{x}-\vec{\mu}_k)$$

$$= \log \frac{\pi_k}{\pi_\ell} + \frac{1}{2}\Big[\cancel{\vec{x}' \Sigma^{-1} \vec{x}} - \vec{x}' \Sigma^{-1} \vec{\mu}_\ell - \vec{\mu}_\ell' \Sigma^{-1} \vec{x} + \vec{\mu}_\ell' \Sigma^{-1} \vec{\mu}_\ell - \cancel{\vec{x}' \Sigma^{-1} \vec{x}} + \vec{x}' \Sigma^{-1} \vec{\mu}_k + \vec{\mu}_k' \Sigma^{-1} \vec{x} - \vec{\mu}_k' \Sigma^{-1} \vec{\mu}_k \Big] =$$

nice!

$$= \log \frac{\pi_k}{\pi_\ell} + \frac{1}{2}\Big[\vec{x}' \Sigma^{-1} (\vec{\mu}_k - \vec{\mu}_\ell) + (\mu_k - \mu_\ell)' \Sigma^{-1} \vec{x} + \vec{\mu}_\ell' \Sigma^{-1} \vec{\mu}_\ell - \vec{\mu}_k' \Sigma^{-1} \vec{\mu}_k + \underbrace{\vec{\mu}_\ell' \Sigma^{-1} \vec{\mu}_k}_{= \vec{\mu}_k' \Sigma^{-1} \vec{\mu}_\ell} - \vec{\mu}_\ell' \Sigma^{-1} \vec{\mu}_k \Big] =$$

$$= \log \frac{\pi_k}{\pi_\ell} + \vec{x}' \Sigma^{-1} (\vec{\mu}_k - \vec{\mu}_\ell) - \frac{1}{2} (\vec{\mu}_k + \vec{\mu}_\ell)' \Sigma^{-1} (\vec{\mu}_k - \vec{\mu}_\ell)$$

Ако $P(G=k \mid \vec{X}=\vec{x}) = P(G=l \mid \vec{X}=\vec{x})$, то

$$\log \frac{P(G=k \mid \vec{X}=\vec{x})}{P(G=l \mid \vec{X}=\vec{x})} = \log 1 = 0$$

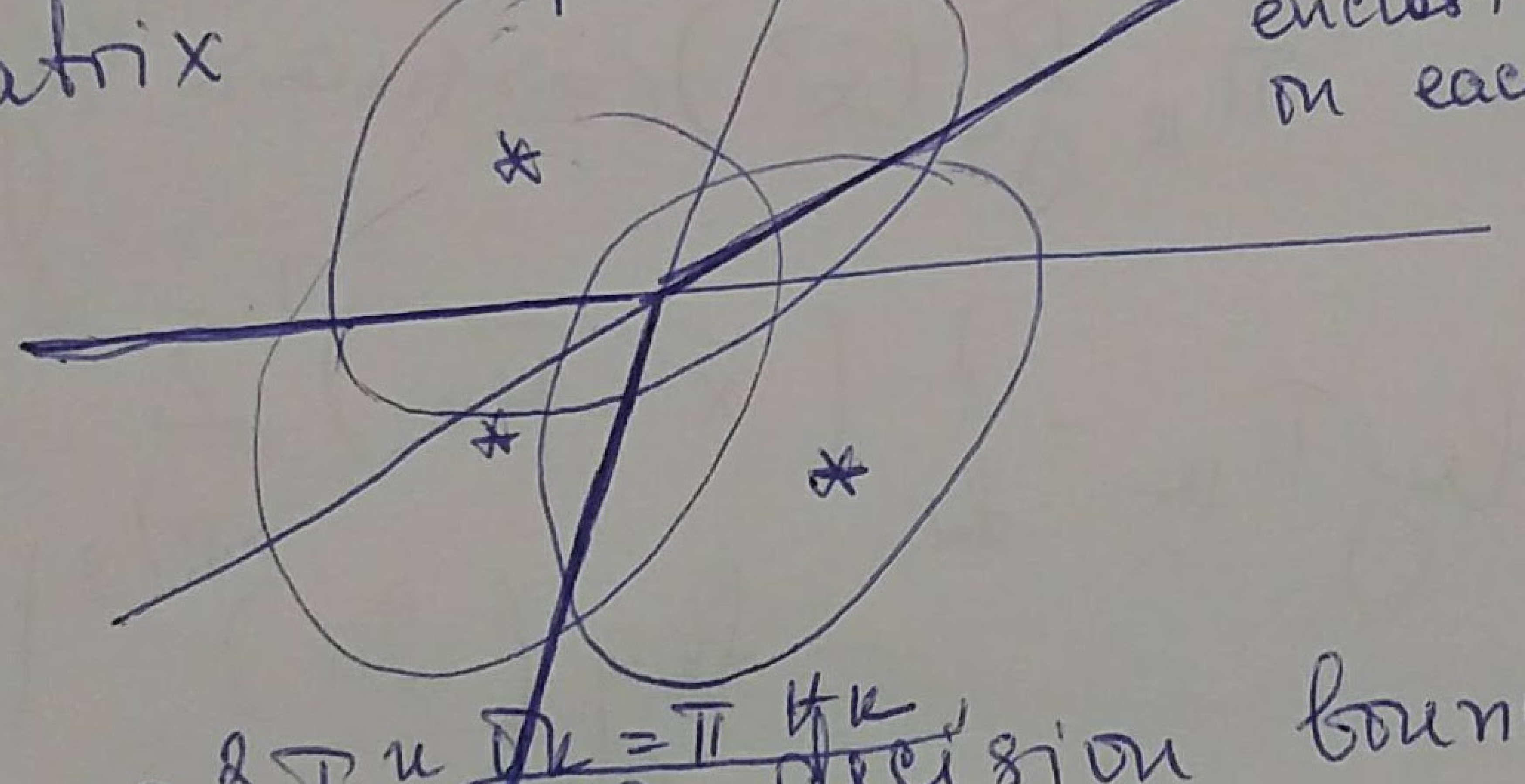
$$(*) \Rightarrow \vec{x}' \Sigma^{-1} (\vec{\mu}_k - \vec{\mu}_l) - \frac{1}{2} (\vec{\mu}_k + \vec{\mu}_l)' \Sigma^{-1} (\vec{\mu}_k - \vec{\mu}_l) + \log \frac{\pi_k}{\pi_l} = 0$$

е уравнение на хиперравнина в R^p

$\Rightarrow$ all the decision boundaries are linear

If we divide R^p into regions that are classified as class 1, 2, etc, these regions will be separated by hyperplanes

Пример: 3 клас и $p=2$, Normal distr. with common covariance matrix



Ако $\Sigma = \sigma^2 I$ и $\pi_k = \pi$ $\forall k$, то decision boundaries са перпендикуляри вектори на правите, свързващи средните $\perp$

На практика не знаем параметрите на Гаусовите р-я и ги оценяваме с training set:

$$\hat{\pi}_k = \frac{N_k}{N} \quad - \quad N_k \text{ е броят на набл. в } k\text{-тия клас}, \; N \text{ е общият бр. набл.}$$

$$\hat{\vec{\mu}}_k = \frac{\sum_{g_i=k} \vec{x}_i}{N_k}$$

$$\hat{\Sigma} = \sum_{k=1} \sum_{g_i=k} (\vec{x}_i - \hat{\vec{\mu}}_k)_{n\times 1} (\vec{x}_i - \hat{\vec{\mu}}_k)'_{1\times n} / N-K$$

Ние гледаме дали при конкретни ст-сти на $\vec{x}$ (*) е > 0 или < 0

$\Rightarrow$ получаваме наредба на $P(G=k \mid \vec{X}=\vec{x})$

Избираме клас, за който P е max. Това е еквивалентно да разгледаме линейните дискриминантни функции (linear discriminant funct.)

$$\delta_k(\vec{x}) = \vec{x}' \Sigma^{-1} \vec{\mu}_k - \frac{1}{2} \vec{\mu}_k' \Sigma^{-1} \vec{\mu}_k + \log \pi_k$$

и да изберем decision rule

$$G(x) = \arg\max_k \delta_k(\vec{x})$$

$$P(G=k \mid \vec{X}=\vec{x}) = \frac{\pi_k f_k(\vec{x})}{\sum_l f_l(\vec{x}) \pi_l}$$ това е общо

$\Rightarrow$ търсим $\max \pi_k f_k(\vec{x}) \Leftrightarrow \max \log \pi_k f_k(\vec{x})$

Но

$$\log \pi_k f_k(\vec{x}) = \log \pi_k - \frac{1}{2}(\vec{x} - \vec{\mu}_k)' \Sigma^{-1} (\vec{x} - \vec{\mu}_k) - \log (2\pi)^{p/2} |\Sigma_k|^{1/2} =$$

$$= \log \pi_k - \frac{1}{2} \vec{\mu}_k' \Sigma^{-1} \vec{\mu}_k + \vec{x}' \Sigma^{-1} \vec{\mu}_k = \delta_k(\vec{x})$$

$$- \frac{1}{2} \vec{x}' \Sigma^{-1} \vec{x} - \log (2\pi)^{p/2} |\Sigma_k|^{1/2}$$ $\rightarrow$ общо $\forall$

$\Rightarrow \log \pi_k f_k(\vec{x}) \rightarrow \max$

$\Updownarrow$

$\delta_k(\vec{x}) \longrightarrow \max$

Quadratic discriminant analysis (QDA):

- Σ_k are not equal $\Rightarrow$ the pieces quadratic in $\vec{x}$ remain

$\Rightarrow$ Разглеждаме quadratic discriminant functions

$$\delta_k(\vec{x}) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(\vec{x}-\vec{\mu}_k)'\Sigma_k^{-1}(\vec{x}-\vec{\mu}_k) + \log \pi_k$$

The decision boundary between each pair of classes k and l is described by a quadratic equation

$$x : \delta_k(\vec{x}) = \delta_l(\vec{x})$$

Ако p-голямо - много параметри!