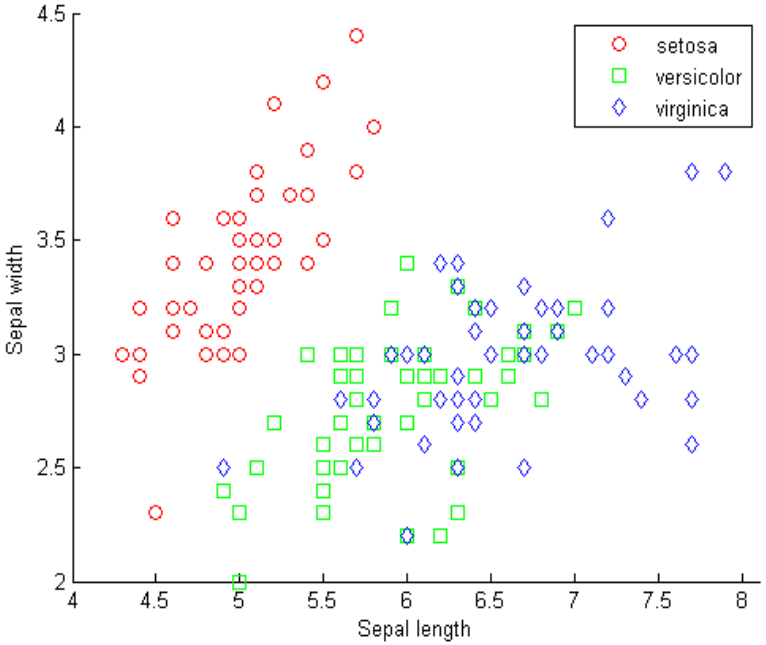
Rozpoznávanie obrazcov šk.r. 2019-20

Bayesovský klasifikátor

Doc. RNDr. Milan Ftáčnik, CSc.

Neseparovateľný prípad

- Bayesovský klasifikátor sa používa pre neseparovateľné prípady
- Vtedy môžu príznakové vektory v tej istej časti príznakového priestoru patriť do 2 rôznych tried
- V takom prípade sa nedá bezchybne zvoliť oddeľujúca nadplocha



Neseparovateľný prípad II

- Klasifikátor je deterministické zariadenie a vektor ***x*** zaradí vždy do tej istej triedy
- Ak sa dopúšťame pri takomto rozhodnutí chyby, klasifikátor nastavujeme tak, aby bola chyba (strata) čo najmenšia
- Ak poznáme pravdepodobnosti objavenia sa triedy $P(\omega_1), P(\omega_2), \ldots, P(\omega_C)$ – aké pravidlo *d(**x**)* zvolíme, aby chyba bola minimálna?

Neseparovateľný prípad III

- Majme aj druhú apriórnu pravdepodobnosť $P(\boldsymbol{x}|\omega_r)$, ktorá vyjadruje podmienenú pravdepodobnosť vektora ***x*** za podmienky známej triedy ω_r
- Ak zvolíme veľkosť straty *0* pri správnej klasifikácii a *1* pri nesprávnej, potom optimálny klasifikátor, ktorý rozhoduje s najmenšou (chybou) stratou je Bayesovský klasifikátor

Bayesovský klasifikátor

- Je založený na dvoch apriórnych (dopredu známych) pravdepodobnostiach: $P(\omega_r)$ ako pravdepodobnosti objavenia sa triedy a $p(\boldsymbol{x}|\omega_r)$ pravdepodobnosti vektora $\boldsymbol{x}$ za podmienky, že je známa trieda ω_r
- Klasifikátor zaraďuje neznámy vektor $\boldsymbol{x}$ do triedy ω_r, ktorá má pre tento vektor najvyššiu aposteriórnu pravdepodobnosť $P(\omega_r|\boldsymbol{x})$

Bayesovský klasifikátor II

- Aposteriórna pravdepodobnosť $P(\omega_r|\boldsymbol{x})$ je podmienená pravdepodobnosť triedy ω_r za podmienky, že na vstupe je príznakový vektor $\boldsymbol{x}$, čiže môžeme povedať, že je to pravdepodobnosť toho, že vektor $\boldsymbol{x}$ patrí do triedy ω_r
- Aposteriórna pravdepodobnosť $P(\omega_r|\boldsymbol{x})$ sa vypočíta z apriórnych pravdepodobností podľa Bayesovho pravidla

Bayesovský klasifikátor III

- Bayesovo pravidlo

$$P(\omega_i|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|\omega_i)P(\omega_i)}{P(\boldsymbol{x})}$$

- V menovateli je nepodmienená pravdepodobnosť $P(\boldsymbol{x})$ – určuje pravdepodobnosť vektora $\boldsymbol{x}$ bez ohľadu na triedu
- Vypočíta sa podľa vety o úplnej pravdepodobnosti $P(\boldsymbol{x}) = \sum_{i=1}^{C} P(\boldsymbol{x}|\omega_i) \,.\, P(\omega_i)$

Bayesovský klasifikátor IV

- Klasifikátor zaradí neznámy vektor $\boldsymbol{x}$ do tej triedy, ktorej aposteriórna pravdepodobnosť je najväčšia, čiže platí

$$\frac{P(\mathbf{x}|\omega_i)P(\omega_i)}{P(\mathbf{x})} \geq \frac{P(\mathbf{x}|\omega_j)P(\omega_j)}{P(\mathbf{x})}$$

- Označuje sa ako MAP (maximum aposterior probability) pravidlo
- Akej chyby sa pri tomto pravidle dopúšťame?

Bayesovský klasifikátor V

- Vzhľadom na rovnaký menovateľ stačí vyhodnocovať čitateľa, čiže rozhodovacie pravidlo MAP je

$$d(\boldsymbol{x}) = \omega_i \leftrightarrow$$

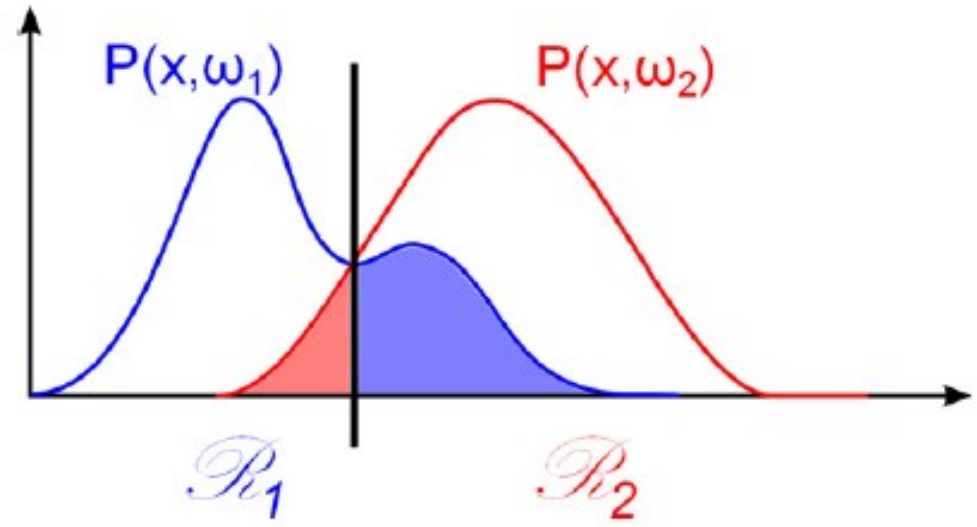
$$P(\boldsymbol{x}|\omega_i)P(\omega_i) = \max_{\forall j=1,2,\ldots,C \ a \ j\neq i} P(\boldsymbol{x}|\omega_j)P(\omega_j)$$

- Klasifikátor vypočíta *C* aposteriórnych pravdepodobností a na základe maximálnej hodnoty rozhodne

Optimalita klasifikátora

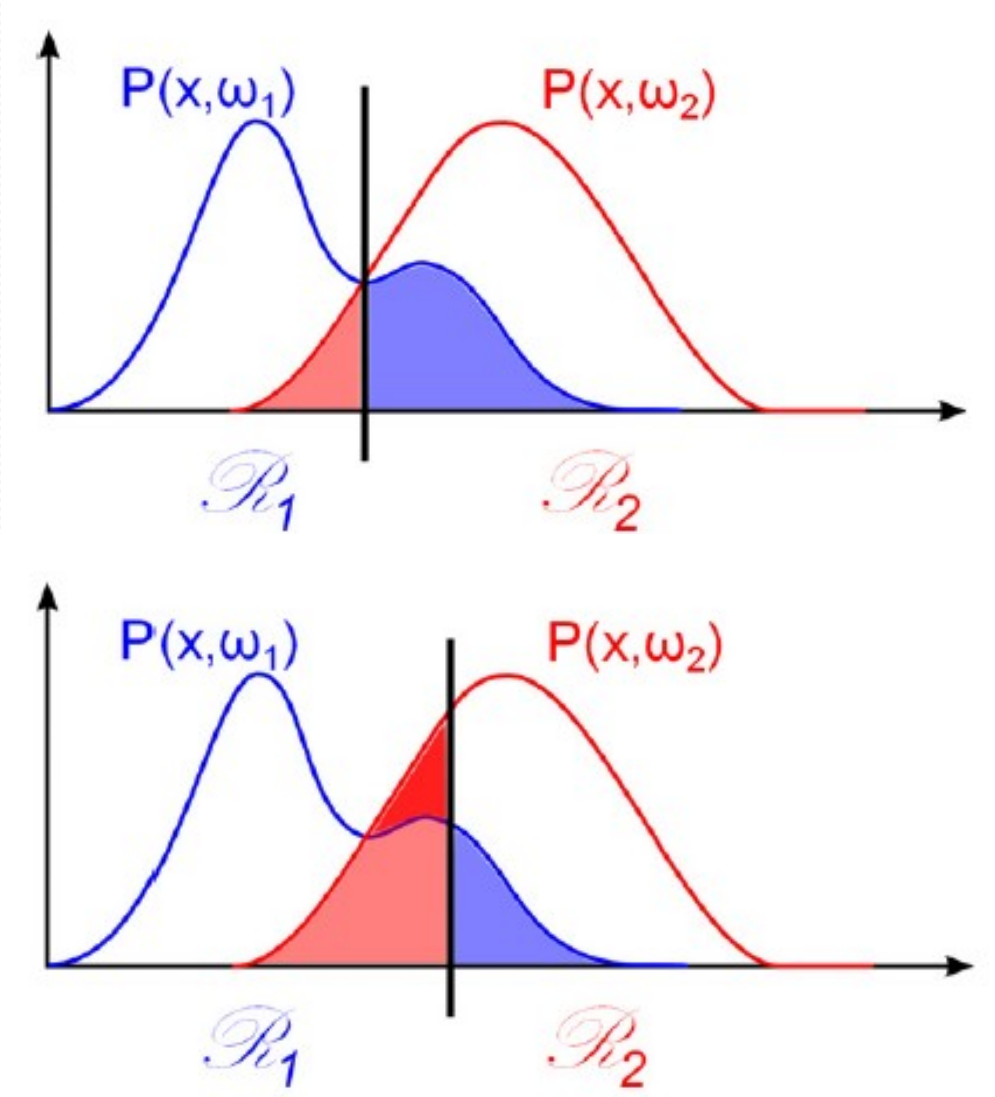
- Chybu vypočítame integrovaním združených pravdepodobností (kde $\mathcal{R}_i$ a $\mathcal{R}_j$ sú rozhodovacie oblasti a $P(A,B) = P(A|B)P(B)$)

$$P(\text{chyba}) = \int_{\mathcal{R}_1} P(\mathbf{x}, \omega_2)\mathrm{d}\mathbf{x} + \int_{\mathcal{R}_2} P(\mathbf{x}, \omega_1)\mathrm{d}\mathbf{x},$$

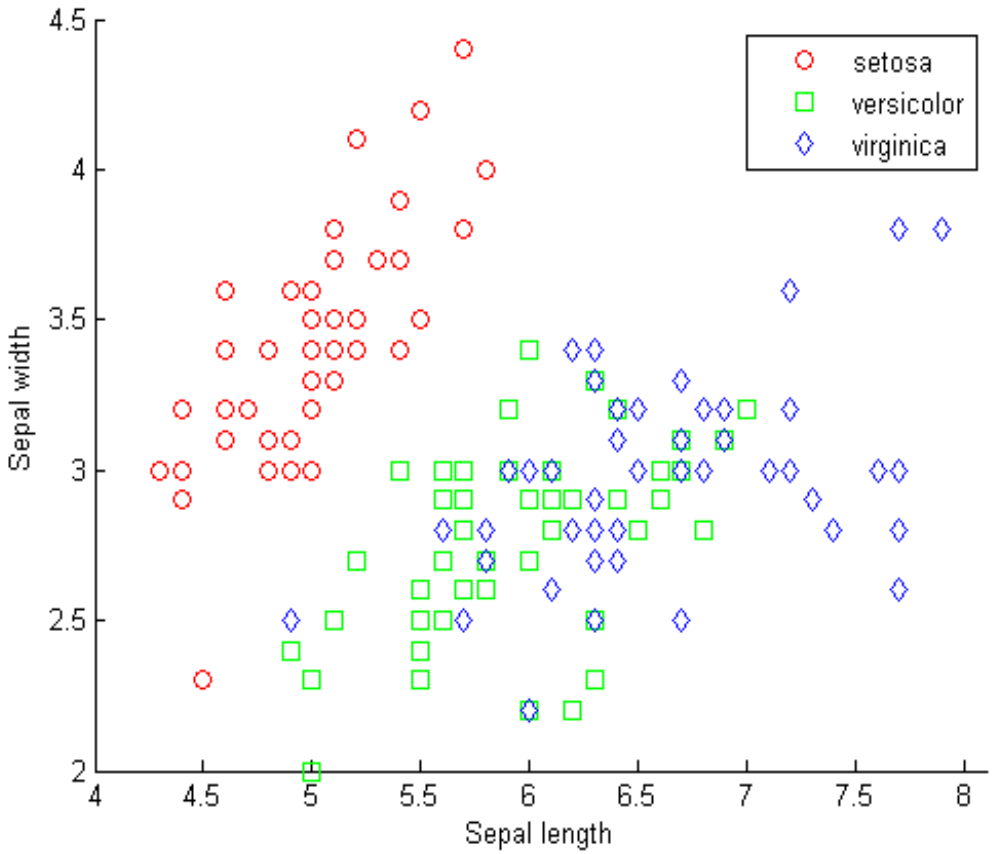


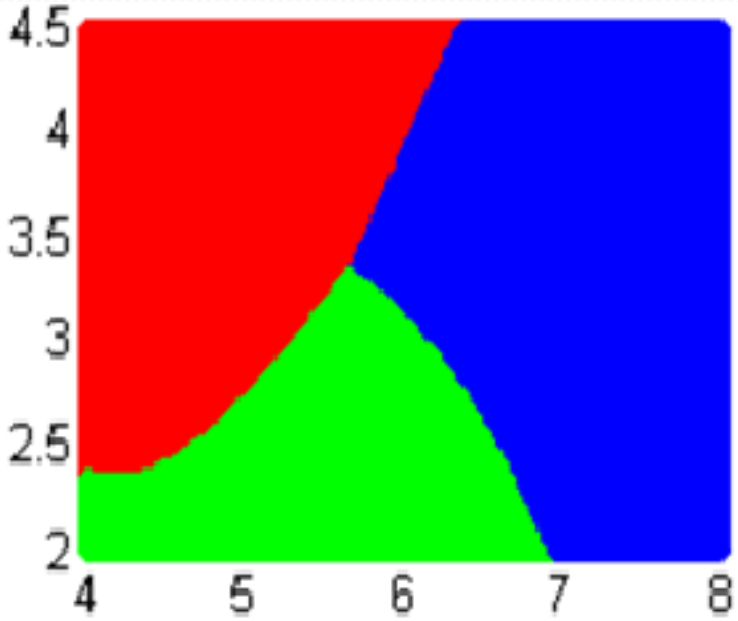
$$\frac{P(\mathbf{x}|\omega_i)P(\omega_i)}{P(\mathbf{x})} \geq \frac{P(\mathbf{x}|\omega_j)P(\omega_j)}{P(\mathbf{x})}$$

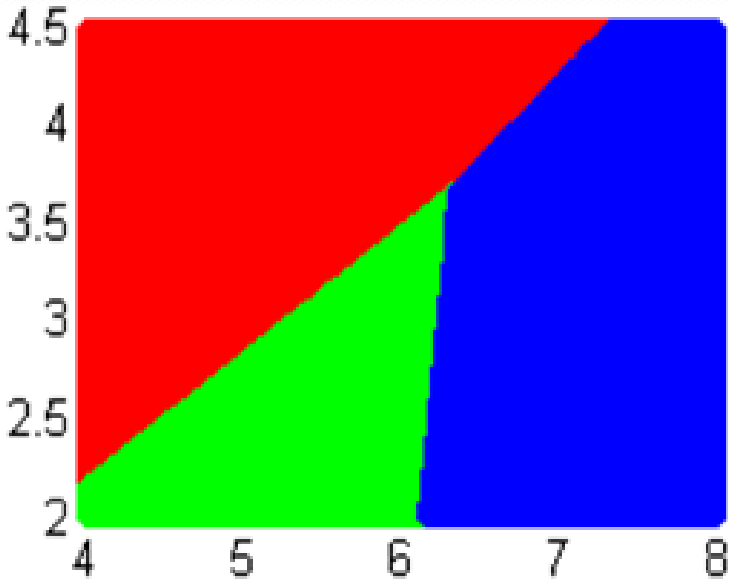
Optimalita klasifikátora II



Porovnanie Bayesovského klasifikátora







Strata Bayesovského klasifikátora

- Vo všeobecnosti nemusí byť strata nesprávnej klasifikácie rovnaká pre všetky triedy
- Napr. pri medicínskych aplikáciách je horšie, ak zaradíme chorého pacienta medzi zdravých ako keď zaradíme zdravého medzi chorých
- Označme $\lambda(\alpha_i|\omega_j)$ ako stratu spojenú s akciou zaradenia objektu do triedy ω_i za podmienky, že patrí do ω_j

Strata Bayesovského klasifikátora II

- Potom očakávanú stratu zo zaradenia vektora $\boldsymbol{x}$ do triedy ω_i môžeme vyjadriť ako

$$R(\alpha_i|\boldsymbol{x}) = \sum_{j=1}^{C} \lambda(\alpha_i|\omega_j)\, P(\omega_j|\boldsymbol{x})$$

- Celková strata cez všetky vektory sa rovná

$$R = \int R(\alpha(x)|x)p(x)dx$$

kde $\alpha(x)$ je jedna akcia na vektore $\boldsymbol{x}$

Strata Bayesovského klasifikátora III

- Ak R je celková strata spojená s rozhodovacím pravidlom, potom sa snažíme ju minimalizovať výberom takej akcie $\alpha(x)$ pre vektor $\boldsymbol{x}$, aby sa minimalizovala strata $R(\alpha_i|\boldsymbol{x})$
- Tak dostaneme stratu R^*, ktorá sa nazýva Bayesovské riziko a je to najmenšia strata, ktorú môžeme dosiahnuť pri informáciách, ktoré sú k dispozícii

Rozhodovacie pravidlo pre 2 triedy

- Zaradíme neznámy vektor do tej triedy, ktorá má menšie podmienenú stratu

$$d(\boldsymbol{x}) = \omega_1 \leftrightarrow R(\alpha_1|\boldsymbol{x}) < R(\alpha_2|\boldsymbol{x})$$

kde $\lambda_{ij} = \lambda(\alpha_i|\omega_j)$

a

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

0-1 strata pre viac tried

- Zvoľme $\lambda(\alpha_i|\omega_j) = 1$ ak $i \neq j$ a inak $= 0$
- Potom

$$R(\alpha_i|\boldsymbol{x}) = \sum_{j=1}^{C} \lambda(\alpha_i|\omega_j)\, P(\omega_j|\boldsymbol{x}) =$$

$$= \sum_{j \neq i} P(\omega_j|\boldsymbol{x}) = 1 - P(\omega_i|\boldsymbol{x})$$

Bayesovský klasifikátor VI

- Potom rozhodovacie pravidlo je

$$d(\boldsymbol{x}) = \omega_i \leftrightarrow P(\omega_i|\boldsymbol{x}) > P(\omega_j|\boldsymbol{x}) \text{ pre } \forall j \neq i$$

- Teraz vyjadríme Bayesovský klasifikátor v jazyku diskriminačných funkcií
- Definujeme

$$g_i(\boldsymbol{x}) = -R(\alpha_i|\boldsymbol{x}) \text{ pre všeobecný prípad a}$$
$$g_i(\boldsymbol{x}) = P(\boldsymbol{x}|\omega_i)P(\omega_i) \text{ pre prípad straty 0-1}$$

Bayesovský klasifikátor VII

- Môžeme tiež napísať pre 0-1 stratu

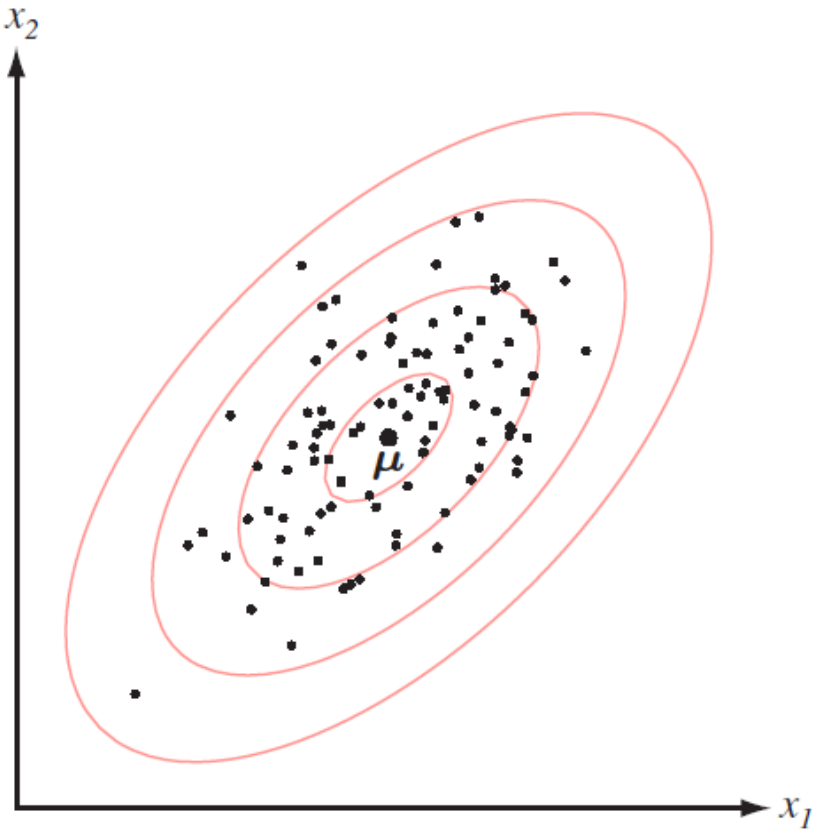
$$g_i(\boldsymbol{x}) = \ln P(\boldsymbol{x}|\omega_i) + \ln P(\omega_i)$$

- Pre dve triedy nám stačí jedna diskriminačná funkcia $g(\boldsymbol{x})$, pri ktorej rozhodujeme podľa znamienka
- Potom $g(\boldsymbol{x}) = P(\omega_1|\boldsymbol{x}) - P(\omega_2|\boldsymbol{x})$, alebo aj

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

Aké sú rozhodovacie oblasti?

- Predpokladajme, že podmienená pravdepodobnosť $P(\boldsymbol{x}|\omega_i)$ má normálne rozdelenie
- V 2D príznakovom priestore sú vektory $\boldsymbol{x}$ rozmiestnené okolo strednej hodnoty μ, pričom elipsy zodpovedajú rovnakej hustote pravdepodobnosti



Aké sú rozhodovacie oblasti II

- Pripomeňme diskriminačnú funkciu
  $g_i(\boldsymbol{x}) = \ln P(\boldsymbol{x}|\omega_i) + \ln P(\omega_i)$
- Ak $P(\boldsymbol{x}|\omega_i) \sim N(\mu_i, \Sigma_i)$, tak môžeme aproximovať $P(\boldsymbol{x}|\omega_i)$ ako

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

- Po logaritmovaní bude diskriminačná funkcia

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\mu_\mathbf{i})^t \boldsymbol{\Sigma_i}^{-1}(\mathbf{x}-\mu_\mathbf{i}) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma_i}| + \ln P(\omega_i)$$

Aké sú rozhodovacie oblasti III

- Našim cieľom je eliminovať z diskriminačnej funkcie konštantné členy a posúdiť, ako budú vyzerať oddeľujúce nadplochy
- Môžu nastať tri prípady:
  - Štatisticky nezávislé príznaky, rovnaká variancia v každej triede
  - Identické kovariancie v každej triede
  - Ľubovoľné kovariancie

Prvý prípad

- Platí $\Sigma_i = \sigma^2 I$
- Po vynechaní konštánt dostaneme funkciu

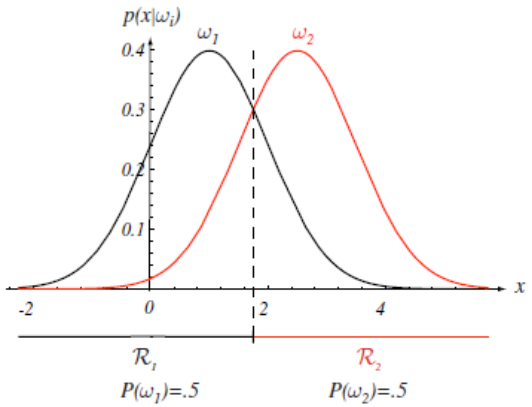
$$g_i(x) = \frac{1}{\sigma^2}\mu_i{}^t \mathbf{x} - \frac{1}{2\sigma^2}\mu_i^t\mu_i + \ln P(\omega_i)$$

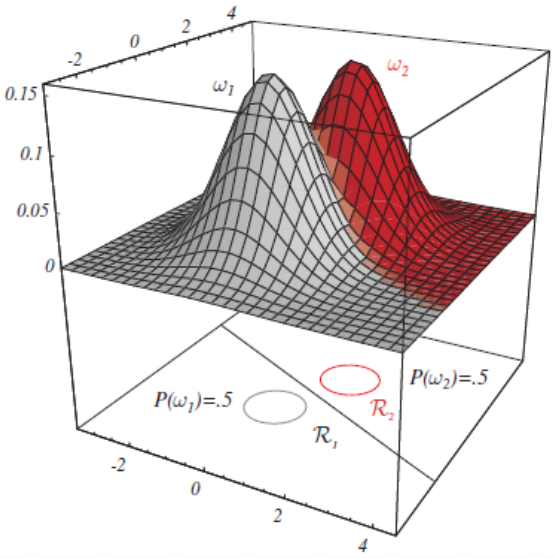
čo je vlastne lineárny klasifikátor

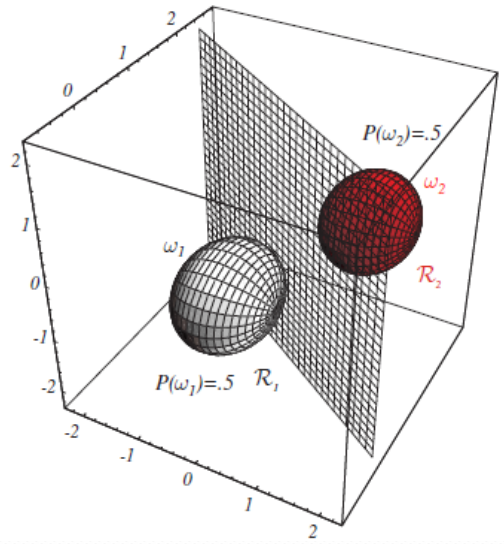
- Ak sa $P(\omega_i)$ rovnajú, tak nadrovina polí úsečku μ_i a μ_j, ak nie, nadrovina sa posúva k jednej hodnote

Prvý prípad II

- Pre prípad, že sa $P(\omega_i)$ rovnajú, tak nadrovina je kolmá na spojnicu stredov dvoch tried
- Ukážka pre 1D, 2D a 3D prípad

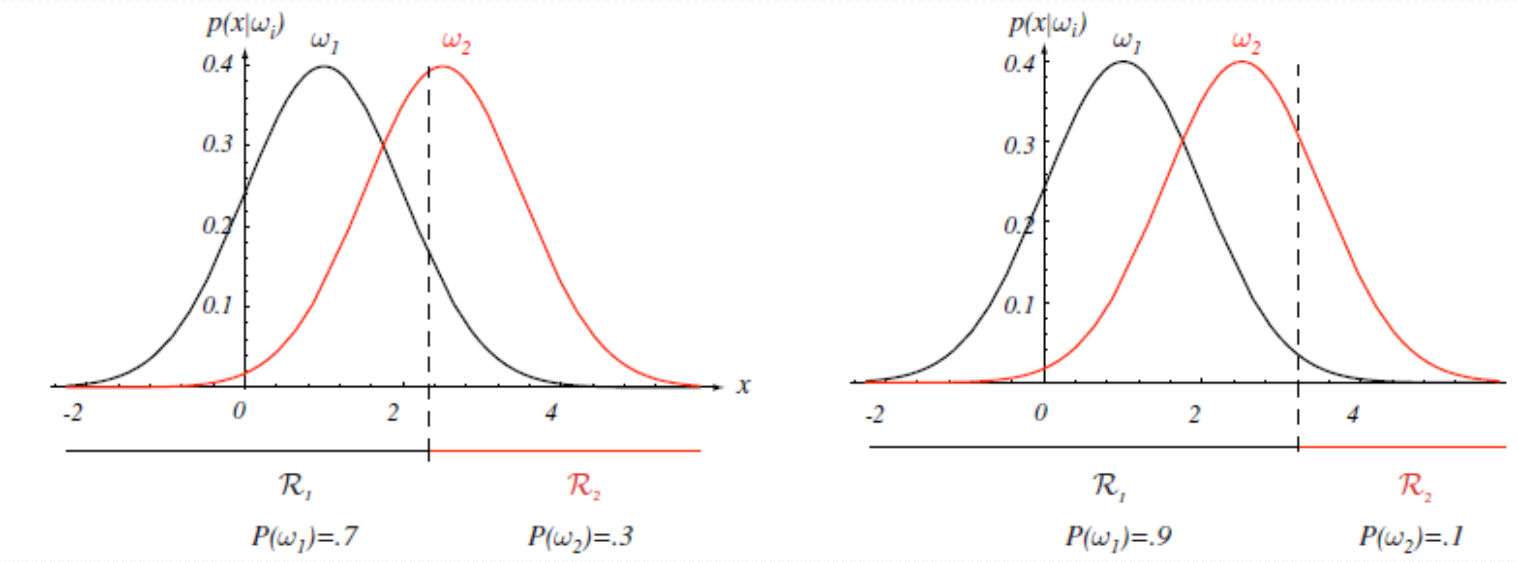






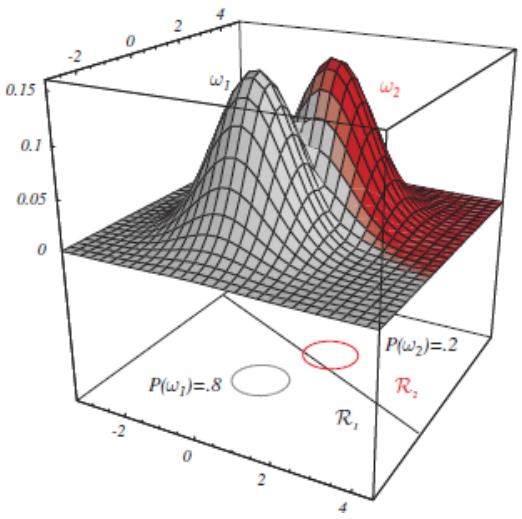
Prvý prípad III

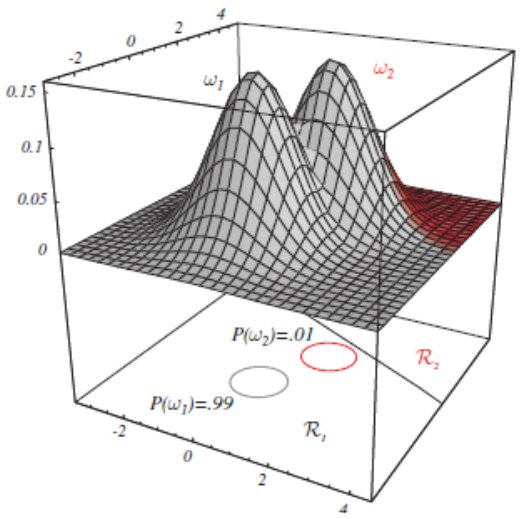
- Ak sa $P(\omega_i)$ nerovnajú, nadrovina sa posúva
- Ukážka pre 1D prípad

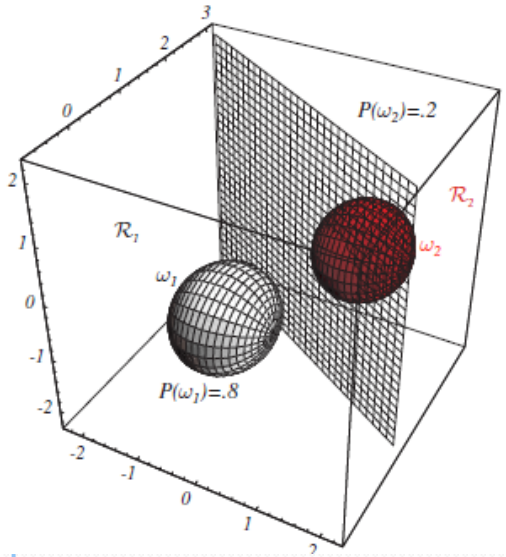


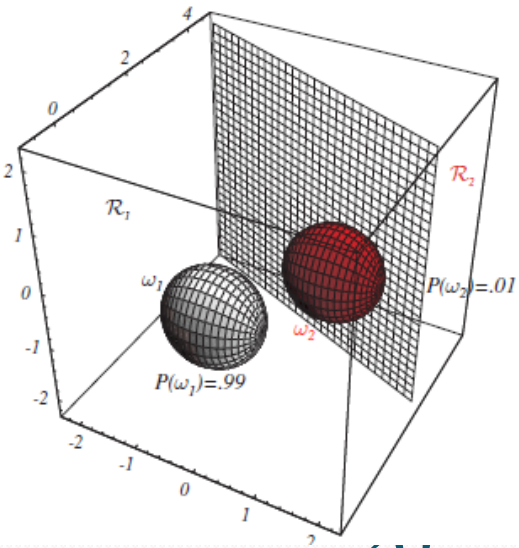
Prvý prípad IV

- Ukážka pre 2D a 3D prípad









Druhý prípad

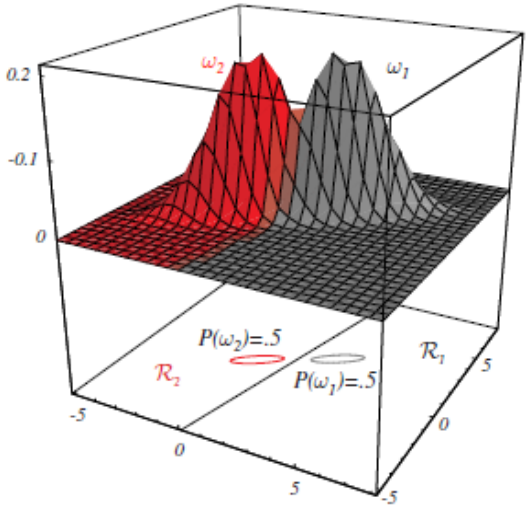
- Identické kovariancie $\Sigma_i = \Sigma$
- Po vynechaní konštantných členov opäť dostaneme lineárnu diskriminačnú funkciu

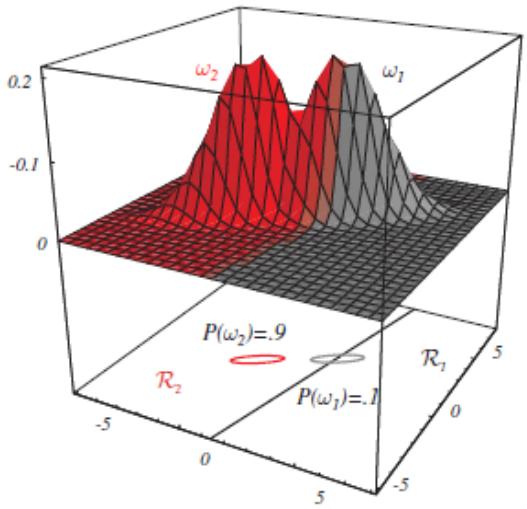
$$g_i(\mathbf{x}) = (\Sigma^{-1}\mu_i)^t\mathbf{x} \quad -\frac{1}{2}\mu_i^t\Sigma^{-1}\mu_i + \ln P(\omega_i)$$

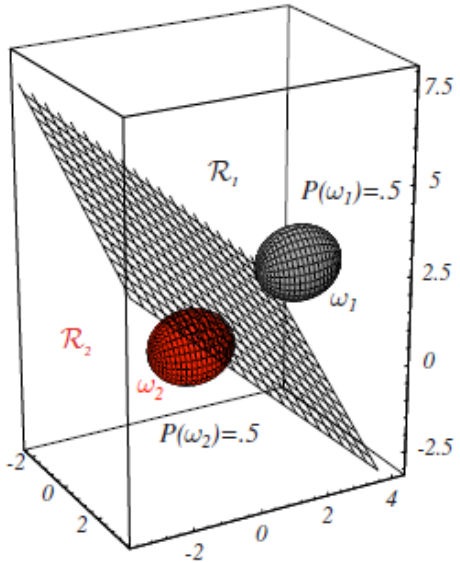
- Ak sa $P(\omega_i)$ rovnajú, tak nadrovina prechádza stredom úsečky μ_i a μ_j, ale nie je na ňu kolmá

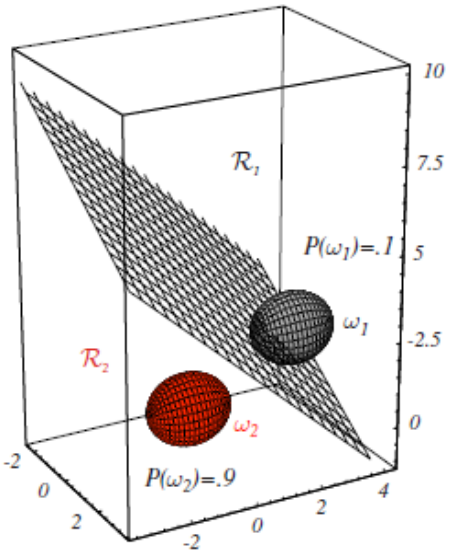
Druhý prípad II

- Ukážka pre 2D a 3D prípad pre rovnaké, ale asymetrické gaussovské rozdelenia









Tretí prípad

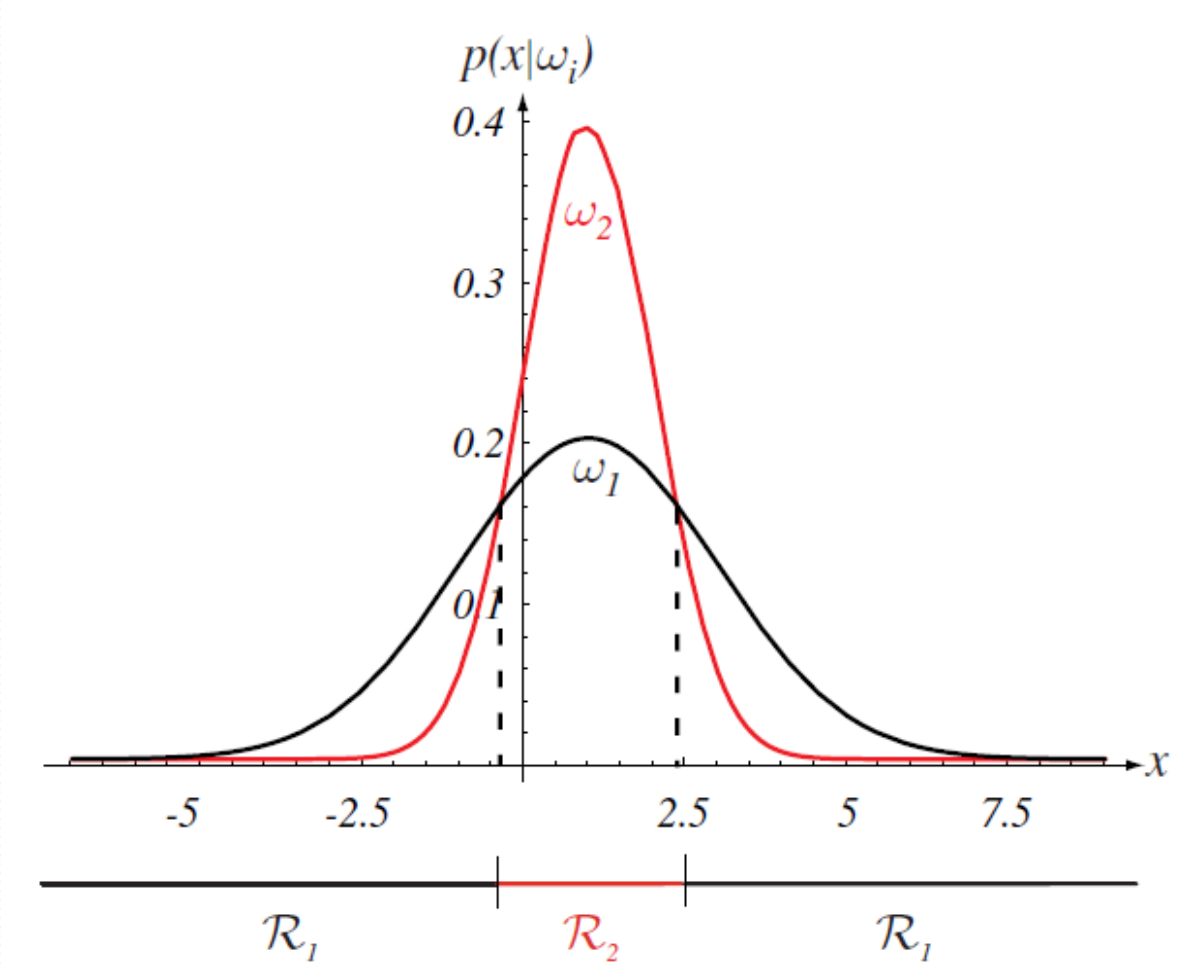
- Ľubovoľné Σ_i
- Diskriminačná funkcia bude kvadratická

$$g_i(x) = x^t\left(-\frac{1}{2}\Sigma_i^{-1}\right)x + \left(\Sigma_i^{-1}\mu_i\right)^t x - \frac{1}{2}\mu_i^t\Sigma_i^{-1}\mu_i - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

- Potom sú oddeľujúce nadplochy: dvojice nadplôch, hypersféry, hyperelipsoidy, hyperparaboloidy, pričom nemusia byť súvislé, a to ani v 1D

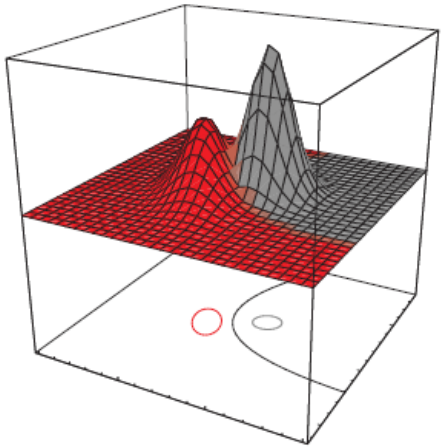
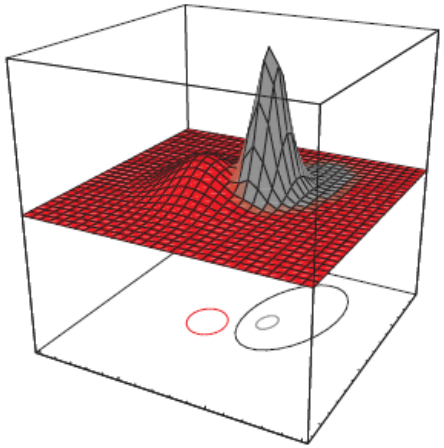
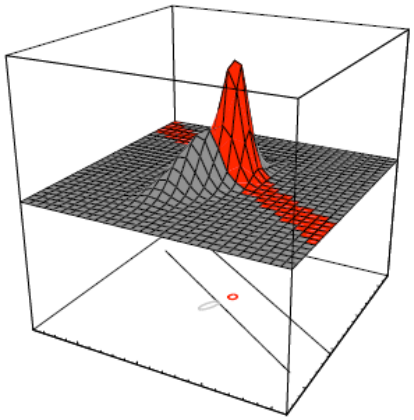
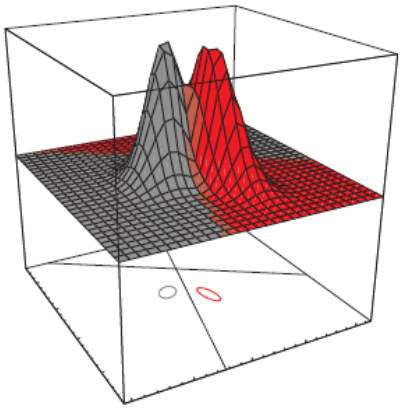
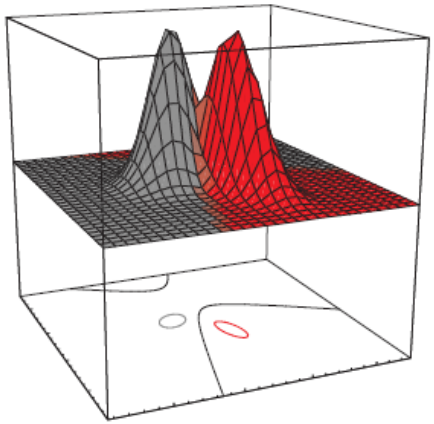
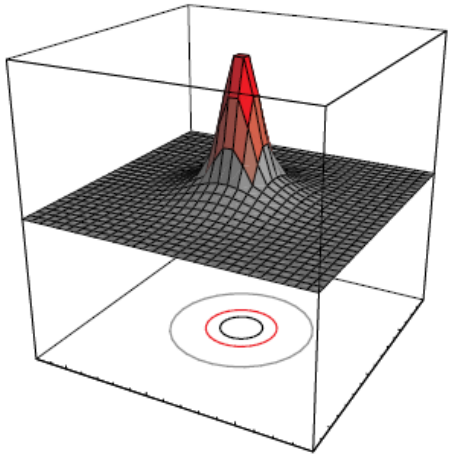
Tretí prípad II

- Ukážka 1D



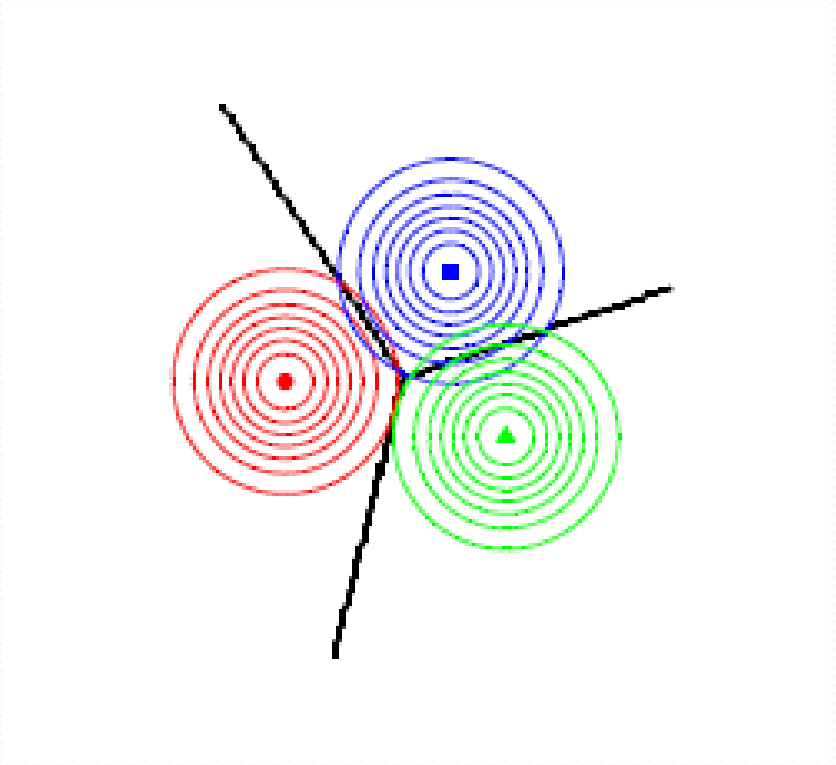
Tretí prípad III

- Ukážky 3D

Rozhodovacie oblasti pre 3 triedy

- Štatisticky nezávislé
- Rovnaká variancia

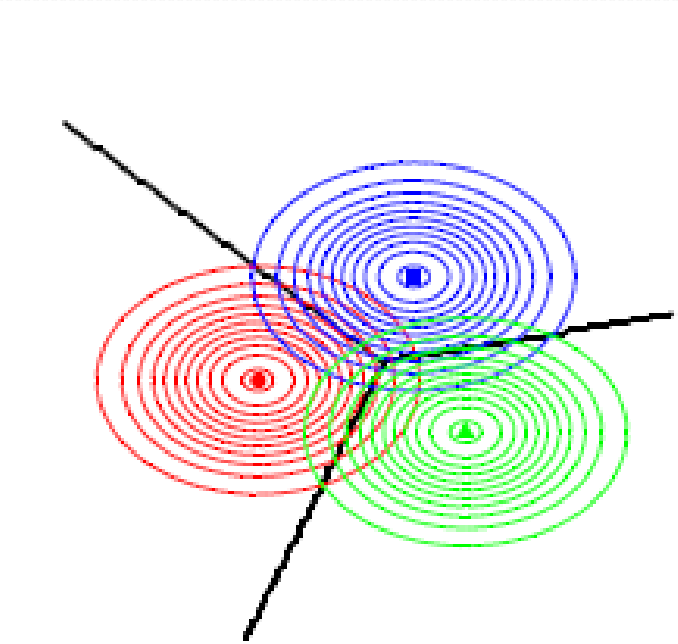


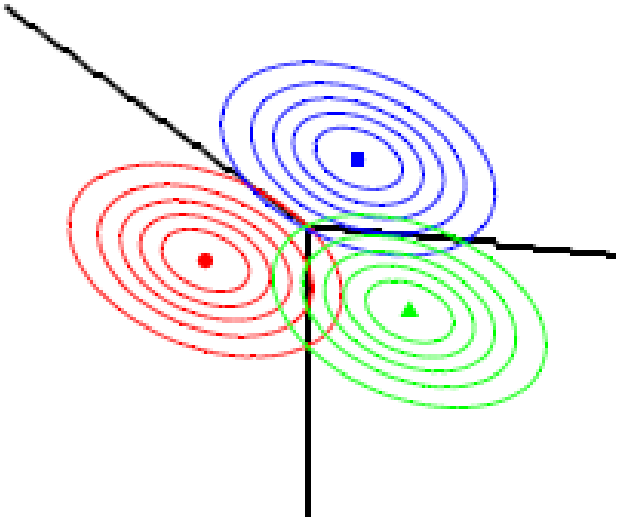
Rozhodovacie oblasti pre 3 triedy II

- Rovnaká kovariančná matica

- Rôzne variancie

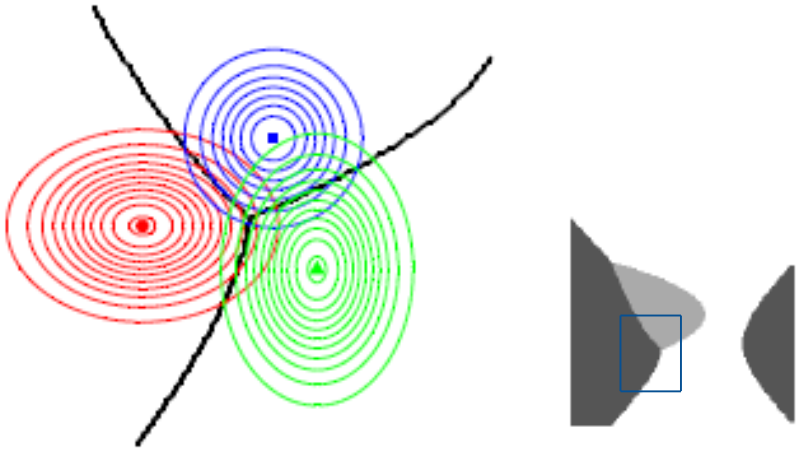
Nie diagonálna



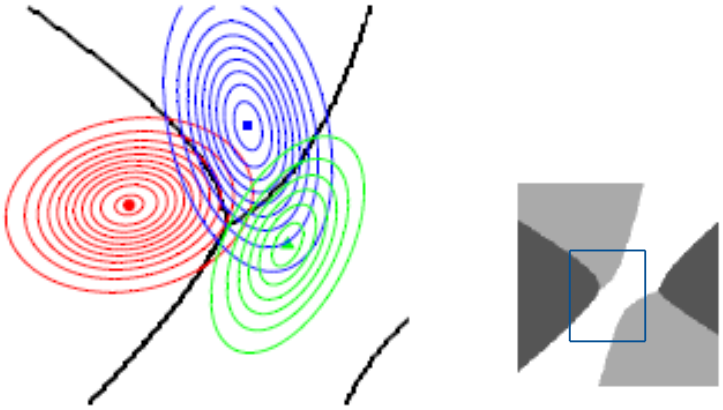


Rozhodovacie oblasti pre 3 triedy III

- Diagonálna kovariančná matica, rôzna



Rôzna kovariančná matica



Učenie klasifikátora

- Kým pri lineárnom klasifikátore sme odhadovali optimálnu hodnotu skalárov, pri Bayesovskom klasifikátore odhadujeme dve apriórne rozdelenia pravdepodobnosti
- Vo všeobecnosti ich nie je jednoduché určiť z trénovacej množiny
- Preto zjednodušíme klasifikátor na naivný Bayesovský klasifikátor

Naivný Bayesovský klasifikátor

- Vychádza z predpokladu, že podmienené pravdepodobnosti jednotlivých príznakov sú štatisticky nezávislé
- T.j. hodnota každého príznaku vektora $\boldsymbol{x}$ v danej triede nie je ovplyvnená hodnotami ostatných príznakov
- Naivný preto, že nezávislosť príznakov vo väčšine reálnych aplikácií neplatí

Naivný Bayesovský klasifikátor II

- Napríklad, ovocie považujeme za jablko, ak je červené, okrúhle a jeho priemer je cca 10 cm
- Tu sa predpokladá, že každý príznak prispieva k pravdepodobnosti, že je to jablko nezávisle od ostatných bez ohľadu na ich možné korelácie
- Na základe predpokladu platí, že

$$P(\boldsymbol{x}|\omega_i) = P(x_1, x_2, \dots, x_D|\omega_i) = P(x_1|\omega_i)P(x_2|\omega_i) \dots P(x_D|\omega_i)$$

Naivný Bayesovský klasifikátor III

- Uvažujme Bayesovo pravidlo:

$$P(\omega_i|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|\omega_i)P(\omega_i)}{P(\boldsymbol{x})}$$

- V čitateli je združená pravdepodobnosť $P(\omega_i, x_1, x_2, \dots, x_D)$ a rovná sa $P(x_1, x_2, \dots, x_D, \omega_i)$
- S využitím definície podmienenej pravdepodobnosti a reťazcového pravidla môžeme napísať, že

Naivný Bayesovský klasifikátor IV

- $P(x_1, x_2, \dots, x_D, \omega_i) = P(x_1|x_2, \dots, x_D, \omega_i)P(x_2, \dots, x_D, \omega_i)$
$= P(x_1|x_2, \dots, x_D, \omega_i)P(x_2|x_3, \dots, x_D, \omega_i)P(x_3, \dots, x_D, \omega_i) =$
$\dots$
$= P(x_1|x_2, \dots, x_D, \omega_i)P(x_2|x_3, \dots, x_D, \omega_i) \dots P(x_{D-1}|x_D, \omega_i).$
$P(x_D|\omega_i)P(\omega_i)$

- Potom nezávislosť príznakov znamená, že

$$P(x_i|x_{i+1}, \dots, x_D, \omega_i) = P(x_i|\omega_i)$$

- Potom platí, že

$$P(\omega_i|\boldsymbol{x}) \propto P(\omega_i, x_1, x_2, \dots, x_D) = P(\omega_i) \prod_{i=1}^{D} P(x_i|\omega_i)$$

Učenie naivného klasifikátora

- Nastavenie parametrov klasifikátora (teda apriórnych pravdepodobností) pre kategorické príznaky sa vypočíta takto:

$$P(x_k|\omega_i) = \frac{N^{i,k}}{N^i} \text{ a } P(\omega_i) = \frac{N^i}{N}, \text{ pričom}$$

N je počet všetkých vzoriek v trénovacej množine

N^i je počet tých vzoriek, ktoré patria do triedy ω_i

$N^{i,k}$ je počet tých, ktoré patria do ω_i a príznak má hodnotu x_k

Príklad 1

- Mám červený Touareg, ukradnú mi ho?

$P(Yes|Red, SUV, Domestic)$?

Zistíme, že $P(Yes) = P(No)$

$P(Red, SUV, Domestic|Yes) = ?$

$P(Red, SUV, Domestic|No) = ?$

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

$$P(\boldsymbol{x}|\omega_i) = \prod_{k=1}^{D} P(x_k|\omega_i)$$

$P(Red|Yes), P(SUV|Yes), P(Domestic|Yes), P(Red|No),$
$P(SUV|No), P(Domestic|No)$

Príklad 2

- Do ktorej triedy patrí tento zákazník?

$X =$
$(age \leq 30,$
$income = medium$
$Student = Yes$
$Credit\ Rating = Fair)$

age	income	student	credit rating	buys computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

Učenie naivného klasifikátora II

- Pre spojité dáta sa snažíme odhadnúť z trénovacej množiny typ pravdepodobnostného rozdelenia a jeho parametre – vtedy používame parametrické metódy
- Ak odhadujeme apriórnu pravdepodobnosť na základe výskytu vzoriek v malom okolí nejakej oblasti, vtedy používame neparametrické metódy

Parametrické metódy

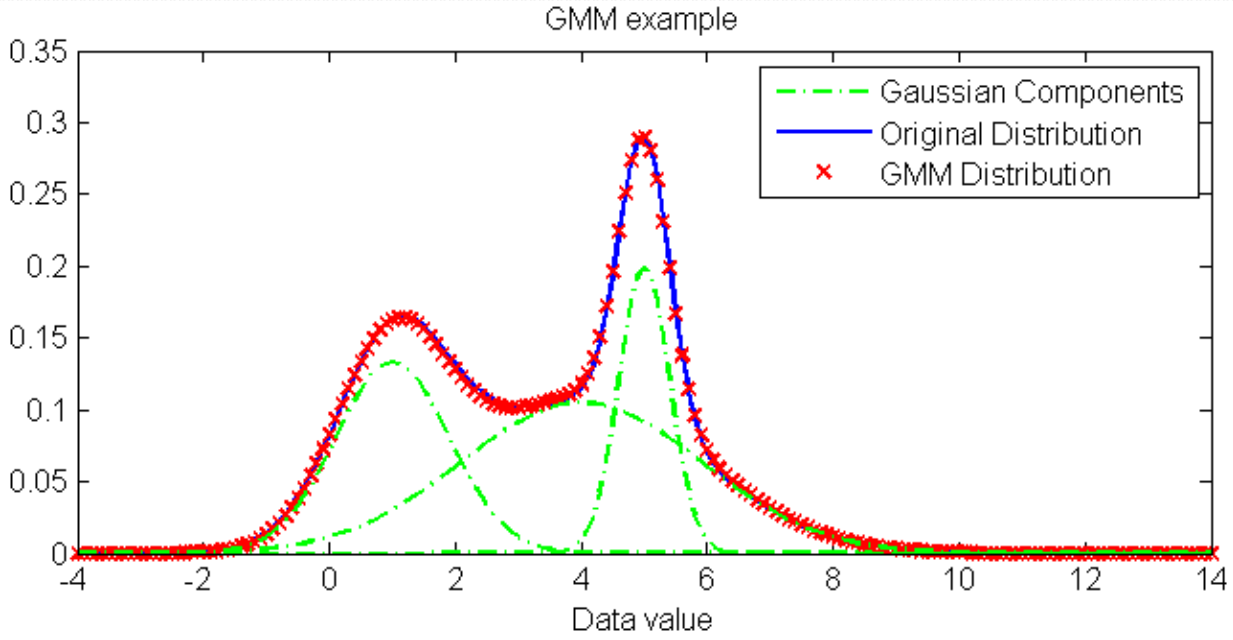
- Predpokladáme, že dáta majú unimodálne rozdelenie – a to buď normálne alebo iné a odhadneme jeho parametre
- Pri normálnom rozdelení vypočítame výberový priemer a výberovú kovarianciu
- Potom určíme apriórnu pravdepodobnosť

$$P(\mathbf{x} \mid \omega_i) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp\left(-\tfrac{1}{2}(\mathbf{x} - \overline{\mathbf{x}}_i)^T \Sigma_i^{-1} (\mathbf{x} - \overline{\mathbf{x}}_i)\right)$$

Parametrické metódy II

- Ak dáta majú multimodálne rozdelenie, tak použijeme GMM – Gaussian mixture model

$$P(\mathbf{x} \mid \omega_i) = \sum_{m=1}^{M} \frac{w_m}{\sqrt{(2\pi)^d |\Sigma_m|}} \exp\left(-\tfrac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_m)^T \Sigma_m^{-1} (\mathbf{x} - \bar{\mathbf{x}}_m)\right)$$



Gaussovský naivný Bayes

- Ak máme spojitý príznak x, rozdelíme ho do tried a vypočítame strednú hodnotu μ_k a rozptyl σ_k v triede ω_k
- Pre nameranú hodnotu v rozdelenie podmienenej pravdepodobnosti vypočítame ako

$$P(x = v|\omega_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}}$$

Mnohočlenový naivný Bayes

- V tomto modeli príznakové vektory reprezentujú frekvencie, s ktorými sú určité udalosti generované mnohočlenom $(p_1, \dots, p_D)$, kde p_i je pravdepodobnosť, že sa objaví udalosť i
- Potom príznakový vektor $\boldsymbol{x} = (x_1, \dots, x_D)$ je histogramom, kde x_i je počet, koľkokrát sa objavuje udalosť i v konkrétnej inštancii
- Tento model sa používa na klasifikáciu dokumentov a udalosť je koľkokrát sa objavilo slovo v dokumente

Mnohočlenový naivný Bayes II

- Potom podmienená pravdepodobnosť je

$$P(\boldsymbol{x}|\omega_k) = \frac{(\sum_i x_i)!}{\prod_i x_i!} \prod_i p_{ki}^{\ x_i}$$

- Ak sa vyjadrí v logaritmickom priestore, tak sa Bayes stane lineárnym klasifikátorom

$$logP(\omega_k|\boldsymbol{x}) \propto \log(P(\omega_k) \prod_{i=1}^{D} p_{ki}^{\ x_i}) =$$

$$= logP(\omega_k) + \sum_{i=1}^{D} x_i \, log \, p_{ki} = b + \boldsymbol{w}_k^T \boldsymbol{x}$$

Bernoulliho naivný Bayes

- V tomto modeli sú príznaky binárne premenné
- Ak x_i je booleovská premenná, vyjadrujúca (ne)prítomnosť *i*-teho termu zo slovníka (pri rozpoznávaní krátkych textov), potom

$$P(\boldsymbol{x}|\omega_k) = \prod_{i=1}^{D} {p_{ki}}^{x_i}(1-p_{ki})^{(1-x_i)}$$

kde p_{ki} je pravdepodobnosť, že trieda ω_k generuje term x_i

Neparametrické metódy

- Pravdepodobnosť, že vybraný vektor bude z oblasti $\mathcal{R}$ je $P = \int_{\mathcal{R}} p(\boldsymbol{x})d\boldsymbol{x}$
- Pre dostatočne malé okolie $\mathcal{R}$ platí:

$$P = \int_{\mathcal{R}} p(\boldsymbol{x})d\boldsymbol{x} = p(\boldsymbol{x}) \int_{\mathcal{R}} d\boldsymbol{x} = p(\boldsymbol{x})V$$

kde V je obsah $\mathcal{R}$

Neparametrické metódy II

- Majme n nezávislých výberov $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ z rozdelenia $p(\boldsymbol{x})$ a nech M z nich patrí do $\mathcal{R}$
- Potom $P = M/n$ a tiež $\hat{p}(\boldsymbol{x}) = \frac{M/n}{V}$
- Problém je ak $V \to 0$
- Na praktickú realizáciu používame metódu Parzenovho okna

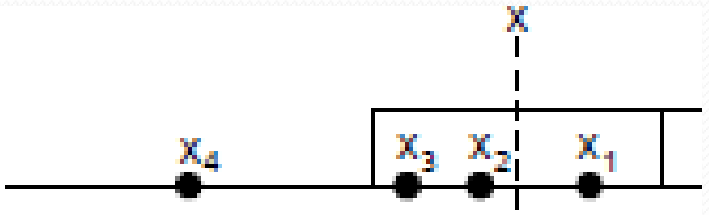
Parzenovo okno

- Uvažujme oblasť $\mathcal{R}$ a D-rozmernú jednotkovú hyperkocku (ako jadro al. kernel) definovanú:

$$k(\mathbf{u}) = \begin{cases} 1 & |u_j| \le 1/2; \quad j = 1,\ldots,D \\ 0 & \text{inak,} \end{cases}$$

- Počet vzoriek v kocke so stranou h a stredom v $\boldsymbol{x}$

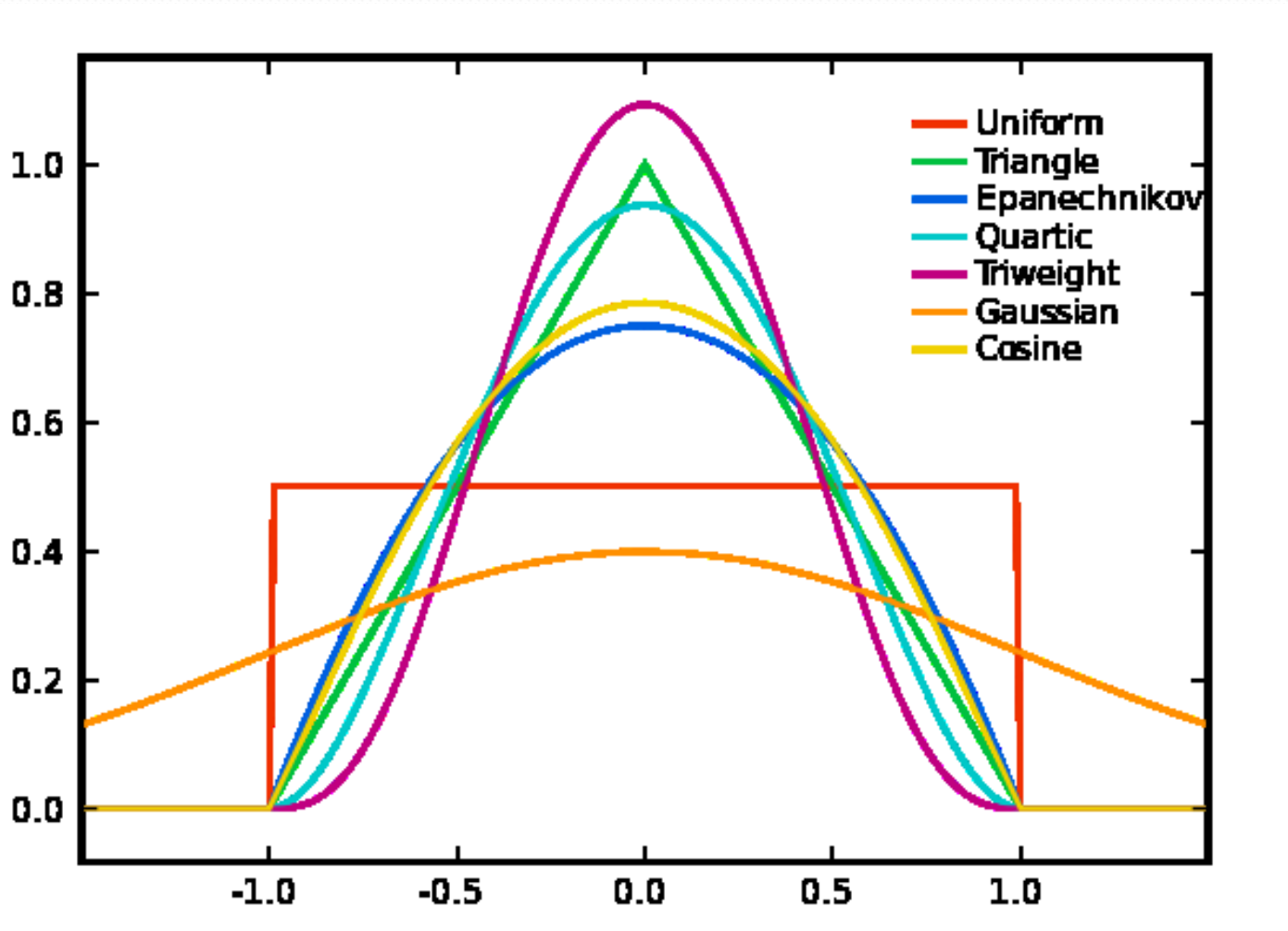
$$M = \sum_{i=1}^{n} k\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)$$



$$\hat{p}(\mathbf{x}) = \frac{M}{nV} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^D}k\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)$$

Parzenovo okno II

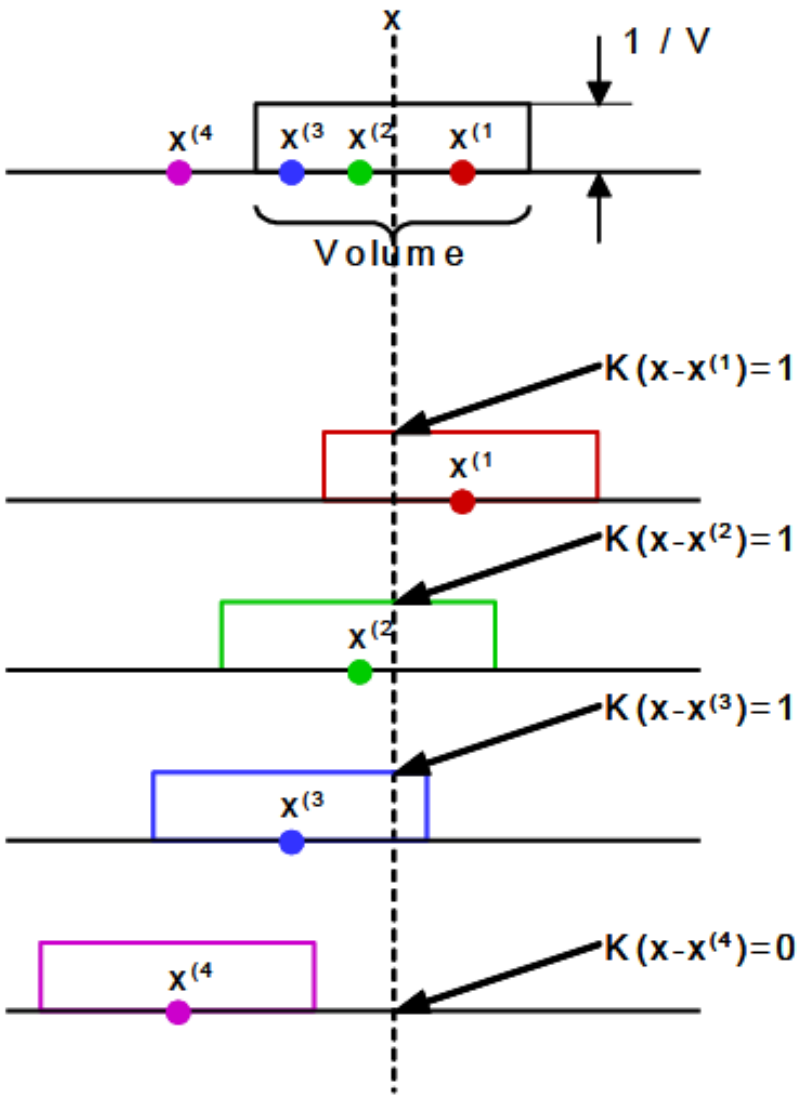
- Spojité jadrá



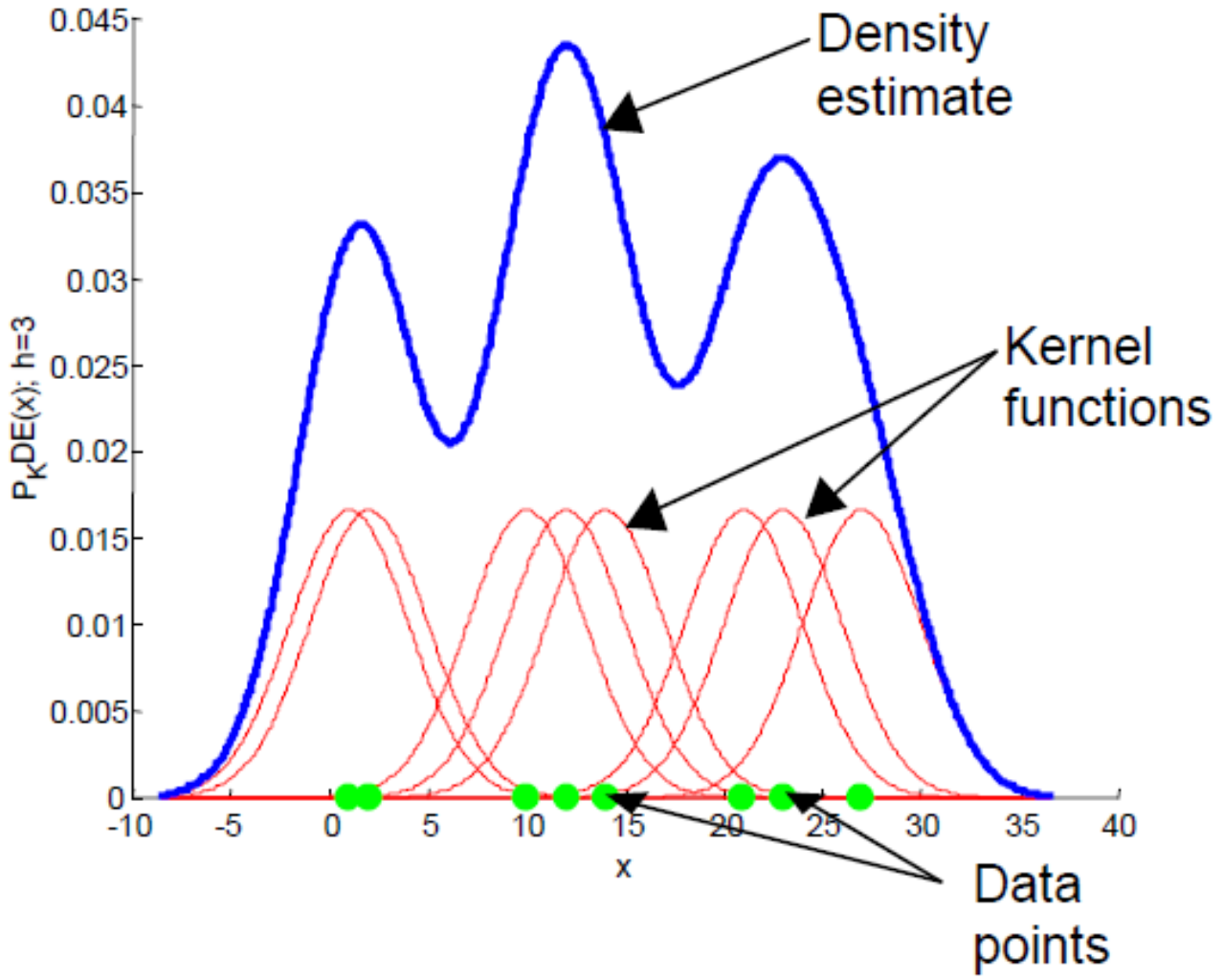
Parzenovo okno III

- Každá vzorka prispieva k hustote v bodoch, ktoré padnú do okna okolo vzorky
- Preto stačí dať okno na vzorky a spočítať

$$p(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^D}k\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)$$

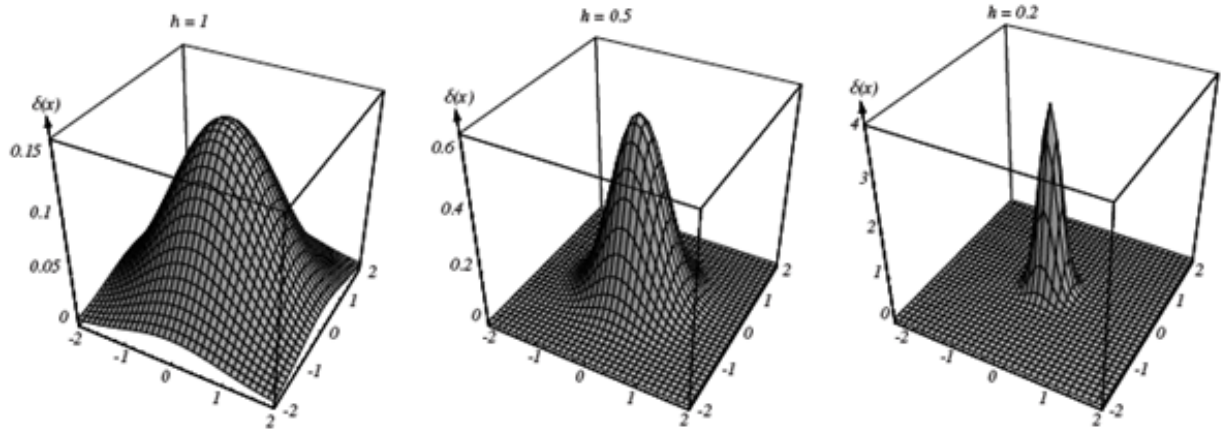


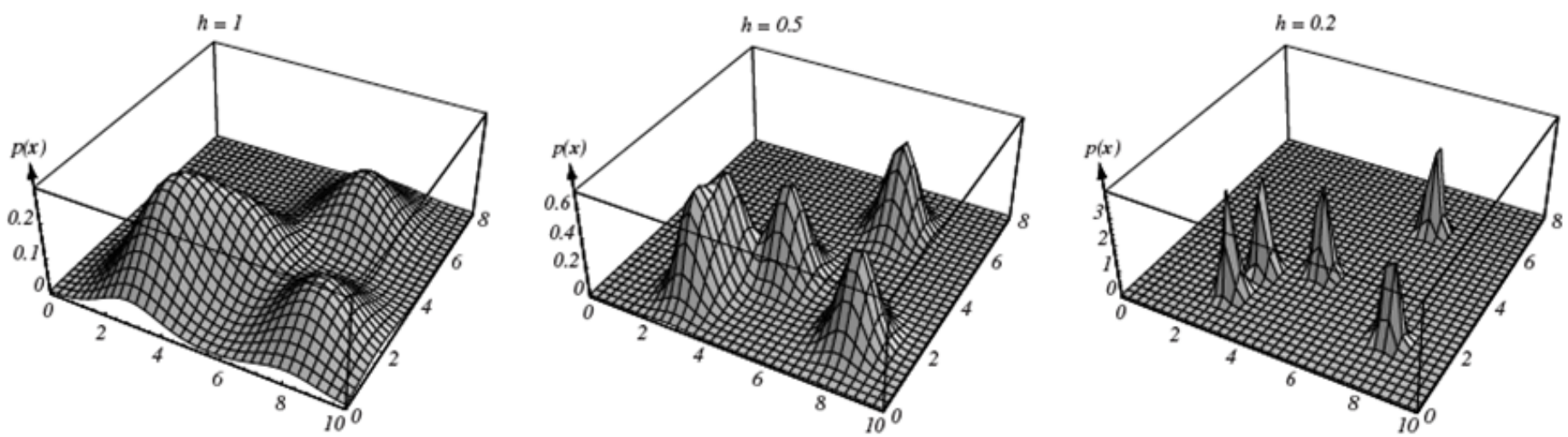
Parzenovo okno IV



Parzenovo okno V

- Veľkosť okna *h* ovplyvňuje amplitúdu aj šírku
- Ukážka kruhového symetrického Parzenovho okna pre rôzne *h*





Výsledky klasifikátora

- Bayesovské metódy zovšeobecňujú omnoho lepšie s malým počtom vzoriek
- Majú veľkú výpočtovú náročnosť pre veľký počet vzoriek
- Aj keď naivný Bayes neodráža úplne reálne vzťahy medzi príznakmi, dokáže relatívne dobre oddeliť od seba triedy, ktoré sú rozdielne