



# Klasifikace obrazových vzorů

## Pokročilá analýza obrazu

Mgr. Markéta Trnečková, Ph.D.

# Klasifikace

- Klasifikace = přiřazení objektu nebo datového vzoru do jedné z předem definovaných tříd na základě jeho vlastností
  - klasifikace pomocí porovnávání s prototypy
  - klasifikace založená na optimální statistické formulaci
  - klasifikace založená na neuronových sítích
- První dva případy se využívají, pokud je úloha dobře definovaná (známe příznaky)
- Poslední spíše, pokud nemáme předem jasnou definici
- Opět si uvádíme spíše principy a aplikace



# Klasifikace

- Lidé mají velice vyspělé schopnosti rozpoznávání vzorů
- Pro počítače je tato úloha složitá – hůře interpretují komplexní obrazy i zobecňují znalosti
- I přes to jsou velice důležité
- Aplikace – čtení čárových kódů, kontrola kvality výrobků, snímání otisků prstů a jiné
- **Vzor** = prostorové uspořádání příznaků
- **Třída vzorů** = množina vzorů, které sdílejí nějaké společné vlastnosti
- Fáze klasifikace
  - 1 snímání
  - 2 předzpracování
  - 3 extrakce příznaků
  - 4 klasifikace

# Klasifikace

## ■ Snímání

- generování signálů v prostorovém (2D) nebo vyšším rozměru

## ■ Předzpracování

- potlačení šumu
- zlepšení kvality
- rekonstrukce
- segmentace

## ■ Extrakce příznaků

- identifikace vhodných příznaků

## ■ Klasifikace

- klasifikace pomocí porovnávání s prototypy
- klasifikace založená na optimální statistické formulaci
- klasifikace založená na neuronových sítích

# Klasifikace

## ■ Klasifikace pomocí porovnávání s prototypy

- Cílem je vytvořit příznaky natolik jedinečné a snadno detekovatelné, aby samotná klasifikace byla jednoduchým úkolem
- Např. využívání stejných písem při rozpoznávání textů

## ■ Klasifikace založená na optimální statistické formulaci

- Klasifikace formulována v termínech rozhodovací teorie a statistiky
- Přístup je založen na volbě parametrů, u nichž lze ukázat, že vedou k optimálnímu klasifikačnímu výkonu ve statistickém smyslu
- Důraz je kladen jak na použité příznaky, tak na návrh klasifikátoru
- Např. Bayesovské klasifikátory vzorů

## ■ Klasifikace založená na neuronových sítích

- Neuronové sítě mohou pracovat i s předem navrženými příznaky, ale mají jedinečnou schopnost samostatně generovat reprezentace (příznaky) vhodné pro rozpoznávání

# Klasifikace

- Všechny jsou založené na parametrech, které musí být buď specifikovány nebo naučeny z množiny vzorů
- Vzory můžeme mít
  - označené – víme, o jakou třídu se jedná
  - neoznačené – víme, že data představují vzory, ale jejich třídy nejsou známy
- Označená data – např. máme vzorky písmen a u každého víme, o jaké písmeno se jedná
- Neoznačená data – hledáme shluk v datech, středy těchto shluků představují prototypy v datech

## Klasifikace – Označená data

- Data dělíme na
  - trénovací data
  - validační data
  - testovací
- typické rozdělení může být 50 % pro trénování a po 25 % pro validační a testovací množinu
- **Trénování** – proces, při kterém se trénovací množina obrázků používá k určení klasifikátoru
- Pro každý trénovací objekt máme třídu, do které chceme, aby ho klasifikátor zařadil
- Pokud klasifikátor udělá chybu, upraví si parametry
- Po ukončení trénování se validační množina použije k určení správnosti klasifikátoru
- Většinou je potřeba několik iterací trénování a validace, aby výsledek byl co nejlepší
- Jakmile máme klasifikátor, ověřujeme ho v praxi pomocí testovací množiny
- Pokud jsou trénovací a validační množiny skutečně reprezentativní, měly by být výsledky testování podobné
- Pokud jsou výsledky trénování a validace uspokojivé, ale při testování nemáme dobré výsledky, říkáme, že došlo k přeučení systému (klasifikátoru), je nutné ho upravit
- Jedná se o **řízené učení**

## Klasifikace – Neznačená data

- Cílem je odhalit strukturu dat, typicky pomocí shlukování (clustering)
- Výsledkem jsou skupiny vzorů s podobnými vlastnostmi
- Středky shluků mohou sloužit jako prototypy tříd
- Využití např. při průzkumné analýze dat nebo předzpracování.
- Vzory se učí **neřízeným učením**
- Výsledek není jednoznačný – různé algoritmy nebo parametry vedou k různým shlukům
- Počet tříd (shluků) často není znám předem a musí se odhadnout
- Obtížné vyhodnocení kvality bez referenčních (označených) dat
- Často slouží jako krok před řízeným učením (např. inicializace nebo redukce dat)

My se budeme zabývat pouze řízeným učením

# Vzory a třídy vzorů

## ■ Vzory:

- kvantitativní – vzory uspořádané ve formě vektorů vzorů
- strukturální – tvořeny symboly uspořádanými ve formě řetězců, stromů nebo grafů

## ■ Vektory vzorů:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

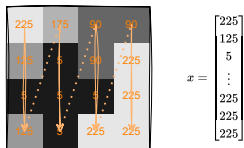
$x_i$  reprezentuje  $i$ -tý deskriptor příznaku

$n$  celkový počet deskriptorů

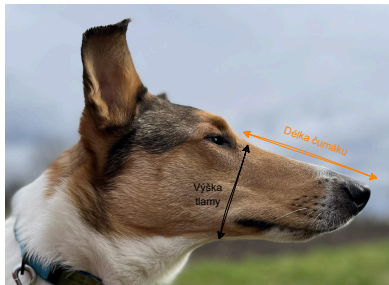
- Vektor vzoru lze chápat jako bod v  $n$ -rozměrném eukleidovském prostoru
- Třídou vzorů lze interpretovat jako „hypermrak“ bodů v tomto prostoru
- Pro účely rozpoznávání je žádoucí, aby třídy vzorů byly co nejvíce kompaktní a co nejvíce od sebe vzdálené

## Vzory a třídy vzorů

- Vektory vzorů můžeme vytvořit přímo z intenzit – vektorizací obrazu

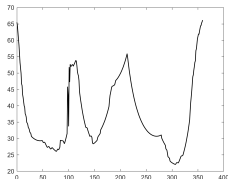


- Častěji se místo intenzit používají příznaky



## Vzory a třídy vzorů

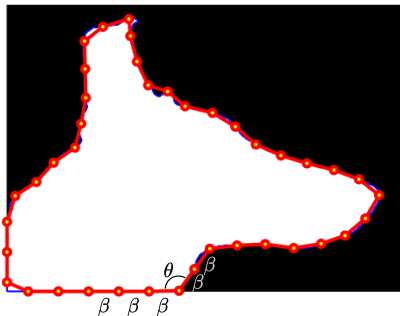
- Vyšší úroveň – deskriptory příznaků (například vektor popisující hranici)
- Vhodné zejména v řízeném prostředí (kontrola kvality)



- Vektory obsahující více příznaků hranic nebo oblastí (kompaktnost, kruhovitost, excentricita, ...)
- Texturové příznaky
- Při práci s celými obrazy se využívá zejména SIFT nebo CNN

## Strukturální vzory

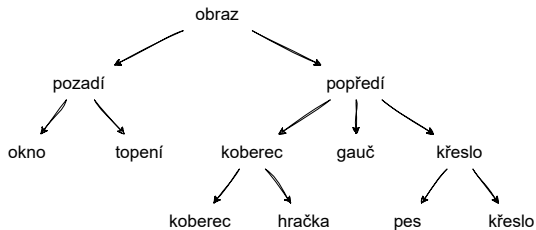
- Objekty jsou reprezentovány **řetězci**, pomocí nichž kódujeme tvar
- Ve zpracování obrazu se využívají méně často
- Jsou důležité v aplikacích, kde je kladen důraz na tvar
- Např. řetězec tvořený úseky určité délky  $\beta$  a vnitřními úhly  $\theta$



$$\alpha = \dots \beta\beta\beta\theta\beta\beta\dots$$

# Strukturální vzory

- **Strom** – vhodný pro popis celého obrazu pomocí jeho komponent
- Hierarchie oblastí v obraze (vztah „složený z ...“)



## Klasifikace vzorů pomocí porovnávání s prototypy

- Srovnání neznámého vzoru se sadou prototypů (každý prototyp reprezentuje nějakou třídu)
- Následné přiřazení třídy, která je vzoru nejvíce „podobná“
- Jedna třída může mít více prototypů  
Každý prototyp patří jen do jedné třídy.
- Jednotlivé metody se liší definicí podobnosti
  - Minimální vzdálenost
  - Korelace

## Klasifikace vzorů pomocí porovnávání s prototypy

- Srovnání neznámého vzoru se sadou prototypů (každý prototyp reprezentuje nějakou třídu)
- Následné přiřazení třídy, která je vzoru nejvíce „podobná“
- Jedna třída může mít více prototypů  
Každý prototyp patří jen do jedné třídy.
- Jednotlivé metody se liší definicí podobnosti
  - Minimální vzdálenost
  - Korelace

# Klasifikace vzorů pomocí porovnávání s prototypy

## Klasifikátor s minimální vzdáleností

- Jeden z nejjednodušších klasifikátorů
- Založen na vzdálenosti mezi neznámým vektorem a každým prototypem tříd
- Vektor se přiřadí k třídě, podle nejmenší vzdálenosti
- Prototypové vektory klasifikátoru s minimální vzdáleností jsou obvykle střední vektory jednotlivých tříd vzorů

$$m_j = \frac{1}{n_j} \sum_{x \in c_j} x, \quad j = 1, 2, \dots, N_c$$

$n_j$  počet vektorů vzorů použitých pro výpočet  $j$ -tého středního vektoru

$c_j$   $j$ -tá třída vzorů

$N_c$  je počet tříd

# Klasifikace vzorů pomocí porovnávání s prototypy

## Klasifikátor s minimální vzdáleností

- Vzdálenost založená na Eukleidovské vzdálenosti

$$D_j(x) = \|x - m_j\|, \quad j = 1, 2, \dots, N_c$$

$$\|a\| = (a^T a)^{1/2} \text{ eukleidovská norma}$$

- Klasifikátor přiřadí neznámý vzor  $x$  do třídy  $c_i$ , pokud platí  $D_i(x) < D_j(x)$  pro  $j = 1, 2, \dots, N_c, j \neq i$

- Výběr nejmenší vzdálenosti je ekvivalentní vyhodnocení funkcí

$$d_j(x) = m_j^T x - \frac{1}{2} m_j^T m_j, \quad j = 1, 2, \dots, N_c$$

přiřazení neznámého vzoru  $x$  do třídy, jejíž prototyp dává největší hodnotu  $d_j$

- Funkce  $d_j$  se nazývá **rozhodovací nebo diskriminační funkce**

- Rozhodovací hranice oddělující třídy  $c_i$  a  $c_j$  je dána hodnotami  $x$ , pro které platí

$$d_i(x) = d_j(x)$$

nebo ekvivalentně

$$d_i(x) - d_j(x) = 0$$

- Rovnice hranice

$$d_{ij}(x) = d_i(x) - d_j(x) = (m_i - m_j)^T x - \frac{1}{2}(m_i^T m_i - m_j^T m_j) = 0$$

# Klasifikace vzorů pomocí porovnávání s prototypy

Klasifikátor s minimální vzdáleností



Kruhovitost  
Solidita

0.55  
0.82

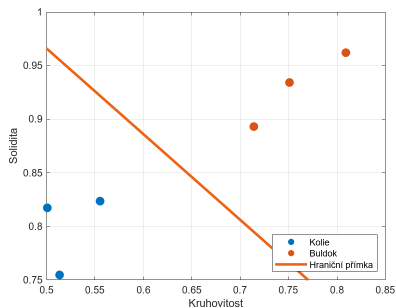
0.50  
0.81

0.51  
0.75

0.81  
0.96

0.75  
0.93

0.71  
0.89



# Klasifikace vzorů pomocí porovnávání s prototypy

## Klasifikátor s minimální vzdáleností

- Klasifikátor funguje dobře, pokud je vzdálenost mezi středními hodnotami velká ve srovnání s rozptylem
- Klíčové je najít vhodné příznaky, které dobře odlišují třídy
- Pokud máme kontrolované prostředí, je tato úloha snadná a nevyžaduje pokročilejší metody

### Příklad

Jaké jiné příznaky by byly vhodné pro rozlišení buldoků a kolíí?

# Klasifikace vzorů pomocí porovnávání s prototypy

## Klasifikátor s použitím korelace

- Korelace masky  $w$  s obrazem  $f(x, y)$   
 $(w \star f)(x, y) = \sum_s \sum_t w(s, t) f(x + s, y + t)$
- Tato hodnota se počítá pro každý posun masky po obraze
- Korelace nabývá nejvyšších hodnot v oblastech, kde jsou  $f$  a  $w$  shodné nebo téměř shodné
- Velká citlivost na změny amplitudy jedné nebo obou funkcí
- Pro normalizaci korelace vůči změnám amplitudy používáme korelační koeficient

$$g(x, y) = \frac{\sum_s \sum_t [w(s, t) - \bar{w}] [f(x + s, y + t) - \bar{f}_{xy}]}{\left\{ \sum_s \sum_t [w(s, t) - \bar{w}]^2 \sum_s \sum_t [f(x + s, y + t) - \bar{f}_{xy}]^2 \right\}^{1/2}}$$

$\bar{w}$  je průměrná hodnota masky

$\bar{f}_{xy}$  je průměrná hodnota funkce  $f$  v oblasti odpovídající poloze  $w$

- $w$  v této úloze nazýváme **šablona** (template, prototypový podobraz)
- Korelace se označuje jako **template matching**

# Klasifikace vzorů pomocí porovnávání s prototypy

## Klasifikátor s použitím korelace

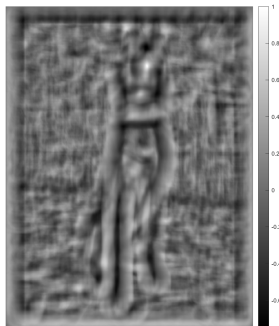
- $g(x, y)$  nabývá hodnot v intervalu  $[-1, 1]$
- Maximální hodnota  $g$  nastává tehdy, když normalizované  $w$  a odpovídající normalizovaná oblast v  $f$  jsou totožné



Obrázek



Šablona



Mapa podobnosti



Nalezené maximum

# Klasifikace vzorů pomocí porovnávání s prototypy

Klasifikátor s použitím korelace

## Příklad

Korelace je poměrně pomalý proces, zejména pro velké šablony. Jak tento proces můžeme urychlit?



## Porovnávání příznaků SIFT

- SIFT počítá množinu invariantních příznaků, které lze použít pro porovnávání mezi známými (prototypovými) a neznámými obrazy
- SIFT vytváří pro každou lokální oblast obrazu 128rozměrné vektory příznaků
- Vzhledem k velkému počtu příznaků je hledání přesných shod výpočetně náročné
- Místo toho se používá metoda best-bin-first, která dokáže s vysokou pravděpodobností identifikovat nejbližší sousedy při omezeném výpočetním úsilí
- Používá se zobecněná metoda Houghovy transformace

# Porovnávání strukturálních prototypů

- Místo vektorů, které popisují obraz kvantitativně budeme pracovat se strukturálními vzory
- Techniky jsou podobné, jako už dříve představené, ale pracujeme s jiným formátem dat
- Různé přístupy, kde zvolené reprezentace:
  - Porovnávání tvarových čísel
  - Porovnávání řetězců

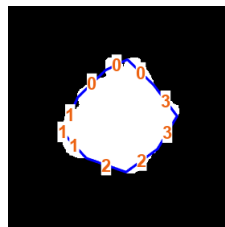
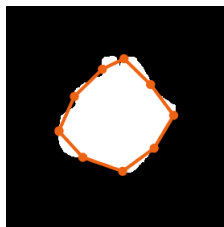
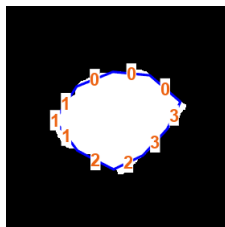
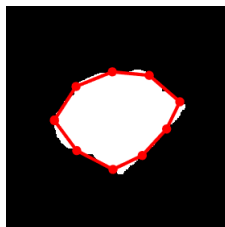
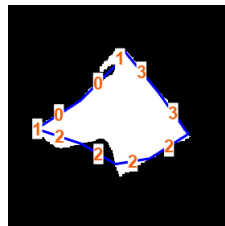
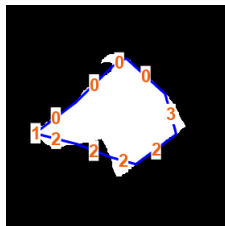
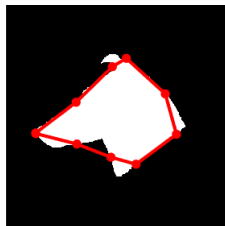
# Porovnávání strukturálních prototypů

## Porovnávání tvarových čísel

- Podobný klasifikátoru s minimální vzdáleností
- Hranice popsané tvarovými čísly
- Stupeň podobnosti  $k$  mezi dvěma hranicemi oblastí definován jako nejvyšší řád, pro který se jejich tvarová čísla ještě shodují
- Příklad:  $a$  a  $b$  označují tvarová čísla uzavřených hranic reprezentovaných čtyřsměrovými řetězovými kódy
- Dva tvary mají stupeň podobnosti  $k$ , pokud
$$s_j(a) = s_j(b), \quad j = 4, 6, 8, \dots, k,$$
$$s_j(a) \neq s_j(b), \quad j = k + 2, k + 4, \dots$$
 $s$  označuje tvarové číslo a index udává řád tvaru
- Vzdálenost mezi dvěma tvary  $a$  a  $b$  je definována jako převrácená hodnota jejich stupně podobnosti
$$D(a, b) = \frac{1}{k}$$
- Pro porovnávání dvou tvarů lze použít buď  $k$ , nebo  $D$

# Porovnávání strukturálních prototypů

Porovnávání tvarových čísel



# Porovnávání strukturálních prototypů

## Porovnávání řetězců

- Dvě hranice oblastí  $a$  a  $b$  jsou zakódovány do řetězců symbolů

$$a_1 a_2 \dots a_n$$

$$b_1 b_2 \dots b_m$$

- Nechť  $\alpha$  označuje počet shod mezi těmito dvěma řetězci  
shoda nastane na  $k$ -té pozici, pokud  $a_k = b_k$

- Počet neshodujících se symbolů

$$\beta = \max(a, b) - \alpha$$

označuje délku (počet symbolů) řetězce

- $\beta = 0$  právě tehdy, když  $a$  a  $b$  jsou totožné

- **Efektivní míra podobnosti** je dána

$$R = \frac{\alpha}{\beta} = \frac{\alpha}{\max(\arg a, \arg b) - \alpha}$$

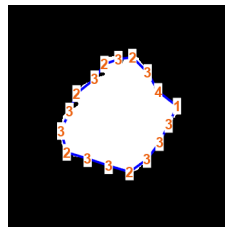
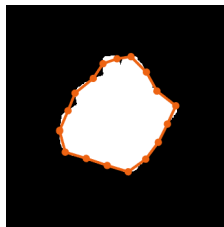
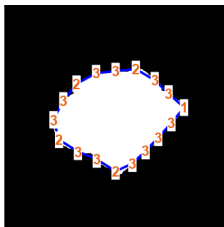
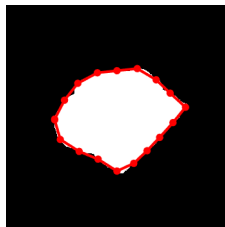
- Porovnávání probíhá symbol po symbolu  $\rightarrow$  důležitá volba počátečních symbolů

- Buď hrubá síla (volí se všechny možné počáteční body u druhého porovnávaného řetězce)



# Porovnávání strukturálních prototypů

## Porovnávání řetězců



$A = 2\ 3\ 2\ 3\ 3\ 2\ 3\ 3\ 1\ 3\ 3\ 3\ 3\ 2\ 3\ 3\ 2\ 3$

$B = 2\ 3\ 2\ 3\ 2\ 3\ 2\ 3\ 4\ 1\ 3\ 3\ 3\ 2\ 3\ 3\ 2\ 3$

$= |||| \dots | \dots |||||$

$a = 13,$

$b = 5,$

$R = 2.6000,$

$\text{shift} = 0$

# Optimální Bayesovské statistické klasifikátory

- Pravděpodobnostní přístup ke klasifikaci vzorů
- Pravděpodobnost hraje velkou roli, když se v generování vzorů objevuje náhoda
- Hledáme optimální klasifikaci ve smyslu, že vede k nejnižší pravděpodobnosti chyb
- Opět jen představíme myšlenku

# Optimální Bayesovské statistické klasifikátory

- $p(c_i|x)$  – pravděpodobnost, že vektor vzoru  $x$  pochází ze třídy  $c_i$
- Pokud klasifikátor rozhodne, že  $x$  pochází ze třídy  $c_j$ , zatímco ve skutečnosti pochází ze třídy  $c_i$ , vzniká ztráta (loss)  $L_{ij}$
- Průměrná ztráta při přiřazení  $x$  do třídy  $c_j$   
$$r_j(x) = \sum_{k=1}^{N_c} L_{kj} p(c_k|x)$$
protože vzor  $x$  může patřit do libovolné z  $N_c$  možných tříd
- $r_j(x)$  nazýváme **podmíněné průměrné riziko** nebo **ztráta**
- Z Bayesova pravidla víte, že  $p(a|b) = \frac{p(a)p(b|a)}{p(b)}$   
$$r_j(x) = \frac{1}{p(x)} \sum_{k=1}^{N_c} L_{kj} p(x|c_k) P(c_k)$$
 $p(x|c_k)$  je hustota pravděpodobnosti (PDF) vzorů ze třídy  $c_k$   
 $P(c_k)$  je pravděpodobnost výskytu třídy  $c_k$  (apriorní pravděpodobnost)
- $\frac{1}{p(x)}$  je kladná a pro všechna  $r_j(x)$  stejná, lze jí vypustit  
$$r_j(x) = \sum_{k=1}^{N_c} L_{kj} p(x|c_k) P(c_k)$$

# Optimální Bayesovské statistické klasifikátory

- Pro neznámý vzor může klasifikátor vybrat jednu z  $N_c$  možných tříd
- Pokud pro každý vzor  $x$  vypočte  $r_1(x), r_2(x), \dots, r_{N_c}(x)$  a přiřadí vzor třídě s nejmenší ztrátou, bude celková průměrná ztráta minimální
- Klasifikátor, který minimalizuje celkovou průměrnou ztrátu, se nazývá **Bayesovský klasifikátor**
- Přiřadí neznámý vzor  $x$  do třídy  $c_i$ , jestliže  $r_i(x) < r_j(x)$  pro všechna  $j = 1, 2, \dots, N_c$ ,  $j \neq i$
- Jinými slovy,  $x$  je přiřazen do třídy  $c_i$ , jestliže
$$\sum_{k=1}^{N_c} L_{ki} p(x|c_k) P(c_k) < \sum_{k=1}^{N_c} L_{kj} p(x|c_k) P(c_k)$$
pro všechna  $j \neq i$
- Ztráta se obvykle nastavuje na 0 při správném rozhodnutí a na 1 při nesprávném
$$L_{ij} = 1 - \delta_{ij}$$
$$\delta_{ij} = 1 \text{ pro } i = j \text{ a } \delta_{ij} = 0 \text{ pro } i \neq j$$
- Průměrná ztráta
$$r_j(x) = \sum_{k=1}^{N_c} (1 - \delta_{kj}) p(x|c_k) P(c_k) = p(x) - p(x|c_j)P(c_j)$$

## Optimální Bayesovské statistické klasifikátory

- Bayesovský klasifikátor pak přiřadí vzor  $x$  do třídy  $c_i$ , jestliže pro všechna  $j \neq i$  platí  $p(x) - p(x|c_i)P(c_i) < p(x) - p(x|c_j)P(c_j)$
- Bayesovský klasifikátor pro ztrátovou funkci 0–1 tedy počítá rozhodovací funkce tvaru  $d_j(x) = p(x|c_j)P(c_j)$ ,  $j = 1, 2, \dots, N_c$
- Přiřadí vzor do třídy  $c_i$ , pokud  $d_i(x) > d_j(x)$  pro všechna  $j \neq i$

## Optimální Bayesovské statistické klasifikátory

- Aby byla optimalita Bayesových rozhodovacích funkcí splněna, musí být známy hustoty pravděpodobnosti vzorů v každé třídě i pravděpodobnosti výskytu jednotlivých tříd
- Pokud jsou všechny třídy stejně pravděpodobné, pak  $P(c_j) = \frac{1}{N_c}$
- Odhad hustot pravděpodobnosti  $p(x|c_j)$  je obtížný
- Pokud jsou vektory vzorů  $n$ -rozměrné, pak  $p(x|c_j)$  je funkcí  $n$  proměnných
- Není-li její tvar znám, je nutné použít vícerozměrné odhadovací metody, které jsou v praxi obtížně použitelné, zejména pokud není k dispozici dostatek reprezentativních vzorů
- Z těchto důvodů se Bayesovský klasifikátor často používá s předpokladem analytického tvaru hustot
- Úloha se pak redukuje na odhad parametrů z trénovacích dat
- Nejčastěji se předpokládá Gaussovská hustota pravděpodobnosti

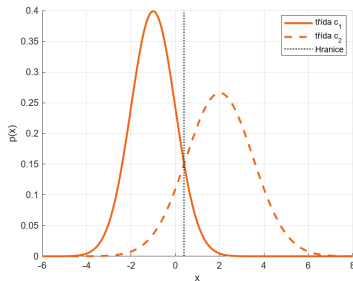
## Optimální Bayesovské statistické klasifikátory

- Aby byla optimalita Bayesových rozhodovacích funkcí splněna, musí být známy hustoty pravděpodobnosti vzorů v každé třídě i pravděpodobnosti výskytu jednotlivých tříd
- Pokud jsou všechny třídy stejně pravděpodobné, pak  $P(c_j) = \frac{1}{N_c}$
- Odhad hustot pravděpodobnosti  $p(x|c_j)$  je obtížný
- Pokud jsou vektory vzorů  $n$ -rozměrné, pak  $p(x|c_j)$  je funkcí  $n$  proměnných
- Není-li její tvar znám, je nutné použít vícerozměrné odhadovací metody, které jsou v praxi obtížně použitelné, zejména pokud není k dispozici dostatek reprezentativních vzorů
- Z těchto důvodů se Bayesovský klasifikátor často používá s předpokladem analytického tvaru hustot
- Úloha se pak redukuje na odhad parametrů z trénovacích dat
- Nejčastěji se předpokládá Gaussovská hustota pravděpodobnosti

# Gaussovské třídy

## 1D případ

- Dvě třídy ( $N_c = 2$ ),  $n = 1$
- Gaussovské hustoty:  $m_1, m_2, \sigma_1, \sigma_2$
- **Rozhodovací funkce:**
$$d_j(x) = \frac{P(c_j)}{\sqrt{2\pi}\sigma_j} e^{-\frac{(x-m_j)^2}{2\sigma_j^2}}$$
- Hranice: bod  $x_0$ , kde  $d_1(x_0) = d_2(x_0)$
- Pokud je stejná pravděpodobnost obou tříd, pak  $x_0$  je rozhodovací hodnoty
- Body, které se nachází vlevo od  $x_0$  přísluší 1. třídě, body vpravo 2.
- Pokud je  $c_1$  pravděpodobnější, pak se  $x_0$  posune vpravo, jinak vlevo



# Gaussovské třídy – $n$ -rozměrný případ

## $n$ -rozměrný případ

- **Gaussovská hustota**  $j$ -tého vzoru

$$p(x|c_j) = \frac{1}{(2\pi)^{n/2}|C_j|^{1/2}} e^{-\frac{1}{2}(x-m_j)^T C_j^{-1}(x-m_j)}$$

$$m_j \text{ střední vektor } m_j = \frac{1}{n_j} \sum x$$

$$C_j \text{ kovarianční matice } C_j = \frac{1}{n_j} \sum xx^T - m_j m_j^T$$

- Kovariance

- diagonála = rozptyly vzorů
- mimo diagonálu = kovariance  $x_k$  a  $x_j$

- Vícerozměrná Gaussovská hustota se redukuje na součin jednorozměrných Gaussovských hustot jednotlivých složek vektoru  $x$ , jsou-li mimosložkové prvky kovarianční matice nulové, což nastává, když jsou složky vektoru  $x_k$  a  $x_j$  nekorelované

## Logaritmická rozhodovací funkce

- Díky exponenciálnímu tvaru Gaussové PDF, můžeme využít přirozený logaritmus

$$d_j(x) = \ln p(x | c_j) + \ln P(c_j)$$

- Logaritmus je monotónně rostoucí a tedy

$$d_j(x) = \ln P(c_j) - \frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |C_j| - \frac{1}{2} (x - m_j)^T C_j^{-1} (x - m_j)$$

Člen  $\frac{n}{2} \ln 2\pi$  je stejný pro všechny třídy a je možné ho vynechat

- **Bayesovská rozhodovací funkce pro Gaussové třídy s 0-1 loss funkcí**

$$d_j(x) = \ln P(c_j) - \frac{1}{2} \ln |C_j| - \frac{1}{2} (x - m_j)^T C_j^{-1} (x - m_j)$$



## Zjednodušené případy

- Pokud jsou kovarianční matice všech tříd stejné ( $C_j = C$ ) → **Lineární rozhodovací funkce**  $d_j(x) = \ln P(c_j) + x^T C^{-1} m_j - \frac{1}{2} m_j^T C^{-1} m_j$
- Pokud navíc  $C = I$  (všechny mají stejné pravděpodobnosti) → **Klasifikátor minimální vzdálenosti**:  
 $d_j(x) = m_j^T x - \frac{1}{2} m_j^T m_j$

# Poznámky

- Optimalita vyžaduje znalost  $p(x|c_j)$  a  $P(c_j)$
- Apriorní pravděpodobnosti obvykle známé
- Často se předpokládá **Gaussovské rozdělení**
- Parametry se odhadují z trénovacích dat
- Přesnost závisí na **shodě modelu s realitou**
- Prahování (Otsuova metoda)  $\approx$  Bayesovská klasifikace

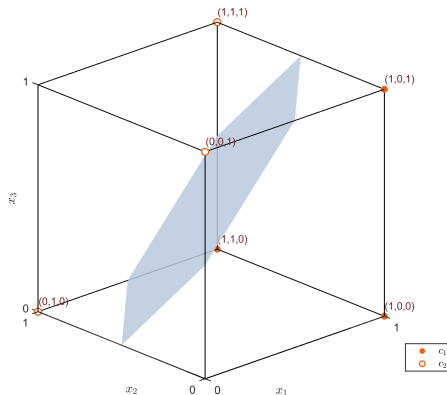
# Bayesovský klasifikátor – Gaussovský příklad

- Dvě třídy:  $c_1$ ,  $c_2$
- Předpoklad:
  - Gaussovské třídy vzorů
  - stejné apriorní pravděpodobnosti

$$m_1 = \begin{bmatrix} 1 \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}, \quad m_2 = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$$

$$C_1 = C_2 = \frac{1}{9} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

$$C^{-1} = \begin{bmatrix} 8 & 4 & 4 \\ 4 & 8 & 4 \\ 4 & 4 & 8 \end{bmatrix}$$



## Bayesovský klasifikátor – rozhodovací funkce

- Pro stejné kovarianční matice platí

$$d_j(x) = x^T C^{-1} m_j - \frac{1}{2} m_j^T C^{-1} m_j$$

$$d_1(x) = 4x_1 - 1.5$$

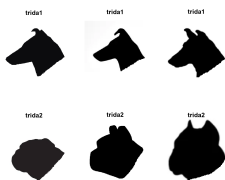
$$d_2(x) = -4x_1 + 8x_2 + 8x_3 - 5.5$$

$$d_1(x) - d_2(x) = 8x_1 - 8x_2 - 8x_3 + 4 = 0$$

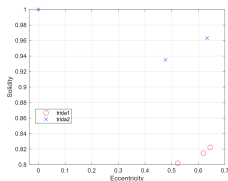
- Vidíme, že
  - rozhodovací hranice je **rovina**
  - na obrázku je její část v jednotkové krychli
  - třídy jsou odděleny efektivně

# Bayesovský klasifikátor – Klasifikace siluet psích hlav

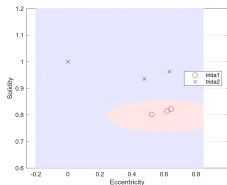
- Vstup: binární siluety hlav psů (2 třídy)
- Extrakce příznaků:
  - Excentricita
  - Solidita
- Každý obrázek  $\rightarrow$  bod v  $\mathbb{R}^2$
- Model: Naivní Bayes (gaussovské rozdělení příznaků)
- Pro každou třídu odhadujeme střední hodnotu a rozptyl



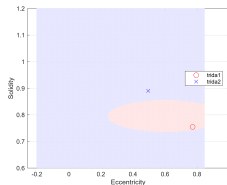
Trénovací data



Příznaky



Model



Testovací data