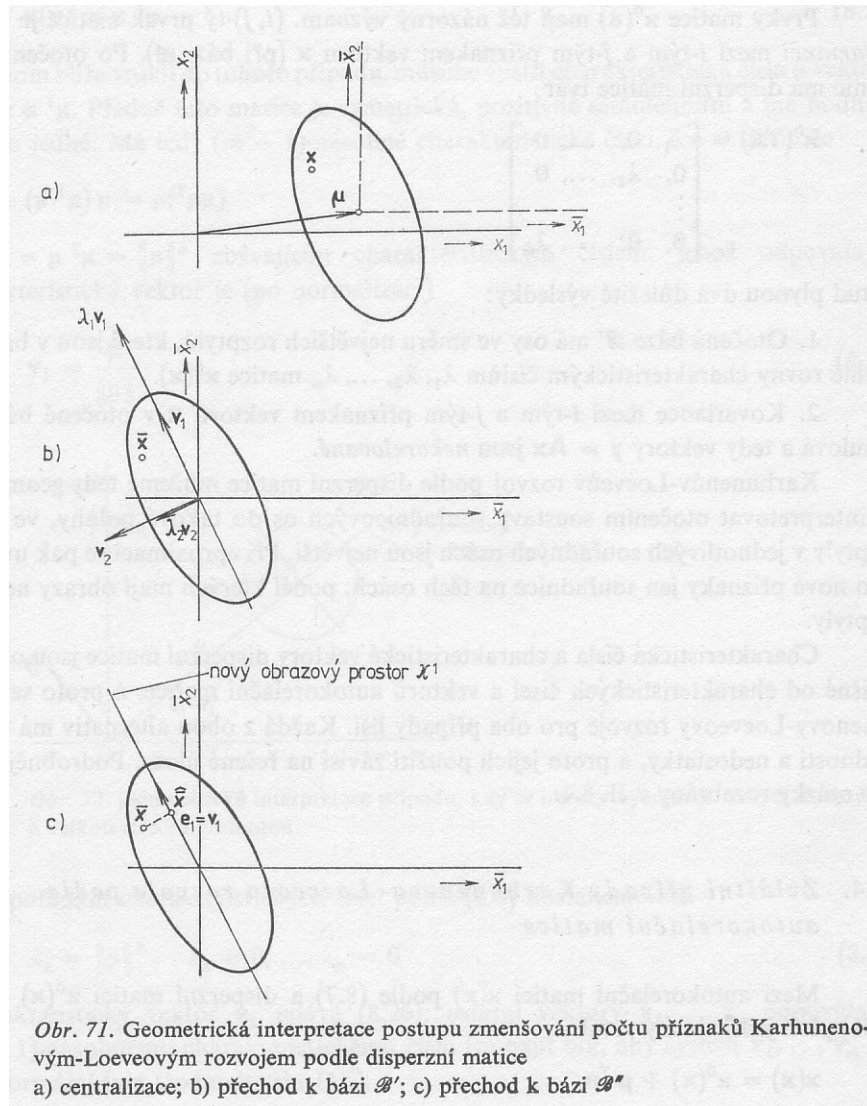


10 Techniky výberu príznakov

- kvalita rozhodovania (klasifikácie) je úzko spätá s množstvom informácie, ktoré má ten, kto rozhoduje – človek alebo počítač – k dispozícii. Inak povedané, pravdepodobnosť chyby u klasifikátora je nepriamo úmerná množstvu informácie, obsiahnutej v príznakových vektoroch (obrazcoch) klasifikovaných predmetov.
- To by znamenalo zvyšovať počet príznakov, ale to sa nedá ľubovoľne, lebo sme obmedzení technickou realizovateľnosťou klasifikátora a predlžuje sa čas, potrebný k vydaniu rozhodnutia. Výsledkom je nutne kompromis medzi veľkosťou tolerovanej chyby a časom na rozhodnutie. V tejto súvislosti sa vracia otázka *informatívnych* príznakov, ktoré sa podieľajú na správnosti klasifikácie.
- Voľba informatívnych príznakov nie je dostatočne formalizovaná, chýba nám teória na výber príznakov, na rozdiel od slušne formalizovanej teórie rozhodovania. Máme dve suboptimálne riešenia:
 - a) výber minimálneho počtu príznakov z vopred zvolenej množiny príznakov
 - b) usporiadanie príznakov vo vopred zvolenej množine príznakov
- Neurčuje sa, ako vybrať vopred zvolenú množinu príznakov a teda nemožno zaručiť, že obsahuje skutočne informatívne príznaky. Voľba tejto množiny je úplne empirická, kde veľkú úlohu hrá skúsenosť konštruktéra. Usporiadanie príznakov má zmysel u sekvenčných metód klasifikácie (napr. rozhodovanie lekára).
- Používané metódy výberu príznakov prevádzajú pôvodnú úlohu na hľadanie vhodného zobrazenia $\mathbf{T} : X^m \rightarrow X^n$, ktorým sa pôvodný m -rozmerný príznakový priestor transformuje do nového n -rozmerného príznakového priestoru, pričom $n \leq m$, pričom toto zobrazenie má, pokiaľ možno, zachovať množstvo informácie, t.j. pravdepodobnosť chyby klasifikácie pri použití n príznakov by nemala byť omnoho väčšia než pri použití pôvodných m príznakov.
- Pretože klasická teória informácie nie je vhodná, ponúkajú sa dve riešenia:
 - a) zobrazenie \mathbf{T} sa volí tak, aby príznakové vektory z priestoru X^n boli čo najlepšou aproximáciou pôvodných príznakových vektorov z X^m v zmysle strednej kvadratickej odchýlky
 - b) zobrazenie \mathbf{T} sa volí tak, aby príznakové vektory z X^n minimalizovali odhad pravdepodobnosti chyby pre prípad, že použijeme kritérium minimálnej chyby.
- Z dôvodov výpočtu sa \mathbf{T} hľadá medzi lineárnymi zobrazeniami.
- Do prvej skupiny patria metódy vybudované na Karhunenovom-Loeveho rozvoji, do druhej skupiny výber príznakov podľa pomeru rozptylov.

Diskrétny Karhunenov-Loeveho rozvoj



- zvláště případy Karhunenovo-Loeveho rozvoja podľa autokorelačnej matice – medzi autokorelačnou maticou $\eta(\mathbf{x})$ a disperznou maticou $\eta^0(\mathbf{x})$ platí vzťah

$$\eta(\mathbf{x}) = \eta^0(\mathbf{x}) + \mu\mu^T$$

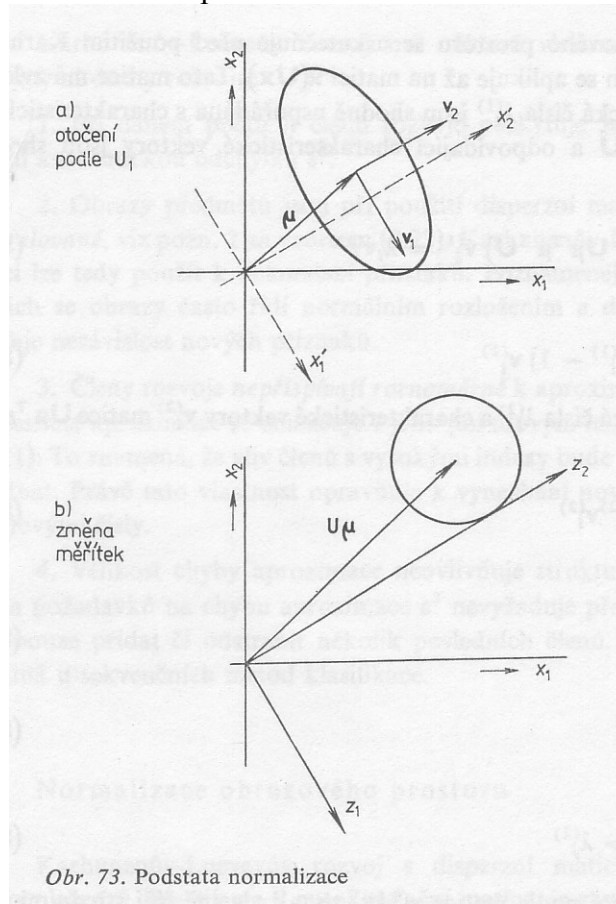
pri veľkých rozptyloch a malých stredných hodnotách sa autokorelačná a disperzná matica približne rovnajú, ako aj ich charakteristické čísla a vektory. Pri malých rozptyloch a veľkých stredných hodnotách naopak platí, že autokorelačná matica sa rovná približne druhému členu, ktorý má jediné nenulové vlastné číslo

$$\lambda_i = \mu\mu^T = \|\mu\|^2 \text{ a ostatných } (m - 1) \text{ vlastných čísel sa rovná nule.}$$

- vlastnosti Karhunenovo-Loeveho rozvoja – je výhodnejší oproti iným rozvojom, pretože:
 - a) pri danom počte n členov poskytuje zo všetkých rozvojev najmenšiu strednú kvadratickú odchýlku ε^2 .
 - b) príznaky pri použití disperznej matice sú po aproximácii nekorelované.

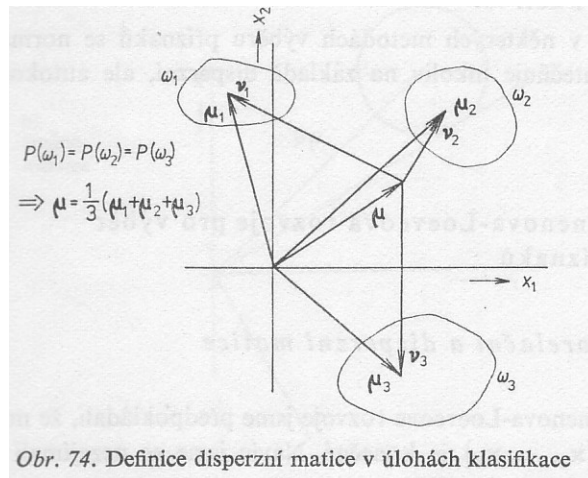
- c) členy rozvoja neprispievajú rovnomerne k aproximácii – vplyv každého člena sa znižuje s jeho poradovým indexom a preto môžeme niektoré členy vynechať.
- d) veľkosť chyby aproximácie neovplyvňuje štruktúru rozvoja, inými slovami zmena požiadaviek na ϵ^2 nevyžaduje prepočítavať celý rozvoj, ale stačí pridať alebo odstrániť niekoľko posledných členov. Toto je výhodné pri sekvenčných metódach klasifikácie.

- normalizácia obrazového priestoru

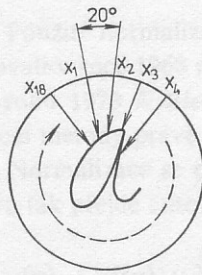


Aplikácia Karhunenovho-Loeveho rozvoja pre výber a usporiadanie príznakov

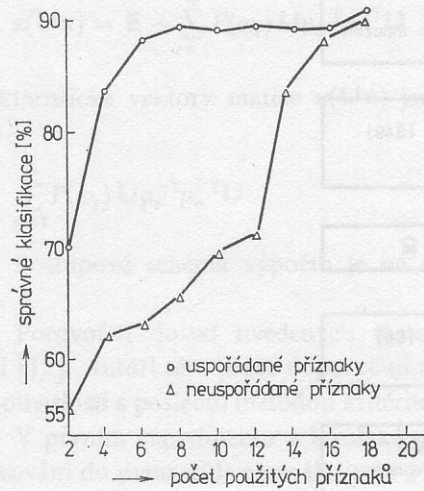
- výberová autokorelačná a disperzná matica – ako odhad neznámej autokorelačnej a disperznej matice



- V praktických úlohách nepoznáme apriórne pravdepodobnosti, preto ich odhadujeme z trénovacej množiny a na jej základe stanovujeme výberovú autokorelačnú a výberovú disperznú maticu. Presnosť potom závisí na veľkosti trénovacej množiny.
- Príklad rozpoznávania rukou písaných písmen s 18 príznakmi. Skúmala sa úspešnosť rozpoznávania pri usporiadaní príznakov popísaným spôsobom, najprv pri dvoch, potom, štyroch, šiestich, až 18 príznakoch, ktorá je na obr. 77.



Obr. 76. Volba príznakov



Obr. 77. Vliv usporiadání príznakov při rozpoznávání písmen

- Karhunenov-Loeveho rozvoj umožňuje optimálne reprezentovať náhodné príznakové vektory v zmysle minima strednej kvadratickej odchýlky radom s pevne daným počtom členov. Umožňuje tak racionalizovať usporiadanie a výber príznakov, pričom za nové príznaky berie koeficienty pri jednotlivých členov radu.
- Ale pri klasifikácii sa neustilujeme o príznaky, ktoré by optimálne reprezentovali klasifikované obrazy, ale dôležité sú najmä príznaky, ktorými sa obrazy z jednotlivých tried čo najviac odlišujú. Môže sa stať, že príznaky vybrané Karhunenovým-Loeveho rozvojom budú reprezentovať predmety, ale budú pre všetky triedy rovnaké a teda pre klasifikáciu bezcenné, pretože majú nízku diskriminačnú schopnosť alebo sú málo informatívne.
- Toto rieši metóda Fukunagu-Koontza, ktorá je vhodná pre klasifikáciu do dvoch tried, pričom vyberá príznaky, ktoré majú vysokú diskriminačnú schopnosť. Jej podstatou normalizácia, ktorá sa uplatní nie na disperznú maticu (ako bežne), ale na autokorelačnú maticu, ktorá sa počíta výlučne z prvkov jednej triedy. Ukáže sa, že charakteristické čísla sú usporiadané tak, že tie, čo sú maximálne pre jednu triedu, sú minimálne pre druhú triedu, teda je možné vybrať také, ktoré najviac odlišujú obe triedy.

Výber a usporiadanie príznačov podľa pomerov rozptylov

- intuitívne sa zdá, že pravdepodobnosť chybného rozhodnutia bude tým menšia, čím menšie budú rozptyly vo vnútri jednotlivých tried a čím viac budú triedy od seba vzdialené. Veľkosť pomeru

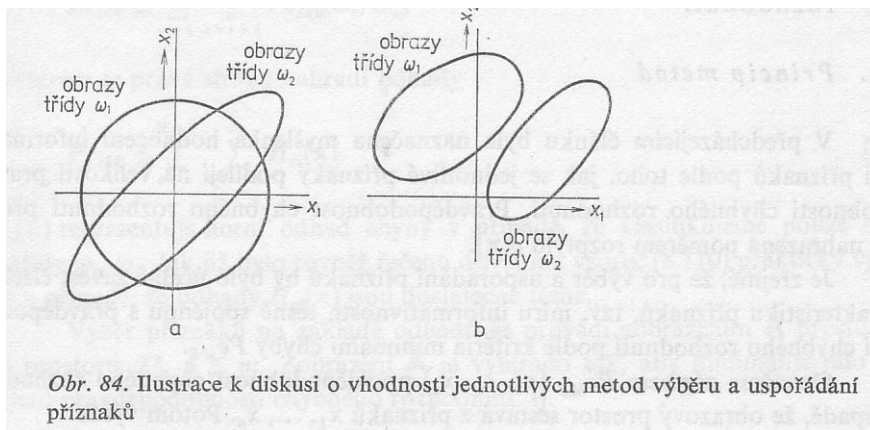
$$\frac{\text{rozptyl medzi triedami}}{\text{rozptyl vo vnútri triedy}}$$

je potom nepriamo úmerný pravdepodobnosti chybného rozhodnutia podľa kritéria minimálnej chyby.

- Jednotlivé príznaky neprispievajú rovnomerne k veľkosti tohto pomeru. Vynechaním príznakov, ktorých príspevok je malý, nezväčšíme veľkosť pravdepodobnosti chyby. Hodnotu príspevku môžeme považovať za určitú mieru informatívnosti príznaku. Usporiadanie podľa klesajúcich hodnôt príspevkov zodpovedá usporiadaniu podľa miery informatívnosti v uvedenom zmysle. Hodnota príspevkov závisí od výberu súradnicového systému v príznakovom priestore X^m ; pričom je žiaduce dosiahnuť čo najväčšiu nerovnomernosť týchto príspevkov, ktorý umožní sústrediť čo najviac informácií do malého počtu príznakov. Zmena príznakového priestoru vhodným zobrazením je opäť podstatou tejto metódy.

Porovnanie a vlastnosti popísaných metód

- vo všeobecnosti môžeme povedať, že metódy založené na Karhunenovom-Loeveho rozvoji preferujú rozptyly (t.j. „tvar“ množín predstavujúcich triedy v príznakovom priestore), pričom metódy založené na pomere rozptylov kladú dôraz skôr na stredné hodnoty tried (t.j. na „rozmiestnenie“ množín predstavujúcich triedy v príznakovom priestore). Pre graficky znázornené úlohy na nasledujúcom obrázku je pre úlohu a) vhodná napr. metóda Fukunagu-Koontza a pre úlohu b) skôr metódy založené na pomere rozptylov.



Metódy založené na odhadoch pravdepodobnosti chybného rozhodnutia

- pri metódach založených na pomere rozptylov sa hodnotila informatívnosť podľa toho, ako prispievali k veľkosti pravdepodobnosti chybného rozhodnutia. Pravdepodobnosť chybného rozhodnutia bola nahradená pomerom rozptylov.
- alternatíva je zaviesť mieru informatívnosti, ktorá by bola tesne spojená s pravdepodobnosťou chybného rozhodnutia podľa kritéria minimálnej chyby Pe_{ME} . Taká miera sa nedá ručiť ľahko, preto sa robia dolné a horné odhady pravdepodobnosti chybného rozhodnutia Pe_{ME} . Používa sa zobrazenie z priestoru X^m do priestoru X^n , tak aby sa minimalizoval horný odhad pravdepodobnosti chybného rozhodnutia, pričom sa používajú obvykle lineárne zobrazenia.

Sekvenčné rozpoznávanie obrazcov

- Waldovo kritérium pre rozpoznávanie do dvoch tried:
- Označme $p(x_1, x_2, \dots, x_i | \omega_1)$, resp. $p(x_1, x_2, \dots, x_i | \omega_2)$, ako podmienené pravdepodobnosti výskytu príznakov z triedy ω_1 , resp. triedy ω_2 , s prvými i príznakmi. Pomer pravdepodobnosti má tvar:
$$\gamma_i = \frac{p(x_1, \dots, x_i | \omega_1)}{p(x_1, \dots, x_i | \omega_2)}$$
- Ďalej vyberieme dve konštanty A a B , také, že $0 < B < 1 < A < \infty$.
- Rozhodovacie pravidlo je definované takto: v i -tom kroku sa priberie k doterajším príznakom i -ty príznak, vyčíslia sa pomer pravdepodobnosti. Ak
 1. $\gamma_i \leq B$, potom sa vektor \mathbf{x} zaradí do triedy ω_2 a proces sa ukončí.
 2. $\gamma_i \geq A$, potom sa vektor \mathbf{x} zaradí do triedy ω_1 a proces sa ukončí.
 3. $B < \gamma_i < A$, potom sa priberie ďalší príznak x_{i+1} a proces sa opakuje.