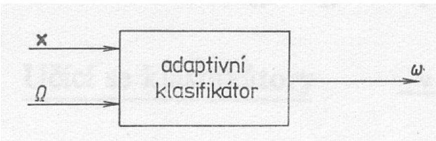
9 Učenie v rozpoznávaní obrazcov

- predpokladáme rozhodovacie pravidlo v tvare $\omega = d(\mathbf{x}, \mathbf{q})$, kde **q** vystupuje ako parameter, ktorým je možné zlepšovať kvalitu pravidla, teda minimalizovať chybu klasifikácie pri neseparabilných obrazoch. Pri učení predpokladáme informáciu učiteľa *Ω*, ktorá obvykle informuje klasifikátor o správnej klasifikácii.

- rozlišujeme tri druhy učenia:
  a) učenie s učiteľom – o tom budeme ďalej uvažovať
  b) učenie s nedokonalým učiteľom – informácia *Ω* nemusí byť vždy správna
  c) učenie bez učiteľa – vedie na metódy zhlukovej analýzy

- trénovacia množina – množina $\{[\mathbf{x}_1, \Omega_1], [\mathbf{x}_2, \Omega_2], ..., [\mathbf{x}_K, \Omega_K]\}$, okrem nej sa používa aj testovacia množina na experimentálne overenie správnosti nastavenia klasifikátora



Obr. 89. Adaptivní klasifikátor

- v činnosti učiaceho sa klasifikátora rozlišujeme dve fázy:
  a) fáza učenia – v ktorej sa mu predkladajú prvky trénovacej množiny a samočinne nastavuje hodnotu parametra **q** tak, aby sa postupne blížila k optimálnej hodnote $\mathbf{q}^*$. Skúsenosť sa ukladá v parametri **q**.
  b) fáza klasifikácie – v tej sa neposkytuje informácia učiteľa

- význam apriórnej informácie – ak by sme poznali apriórne pravdepodobnosti pri štatistickom rozpoznávaní – obvykle ich v tomto rozsahu nepoznáme, ale vieme napr., že obrazy majú normálne rozdelenie apod. Do apriórnej informácie zaraďujeme aj informáciu o separabilite. Trénovacia množina nepatrí do apriórnej znalosti.

- matematická formulácia učenia klasifikátora: ak informácia *Ω* poskytuje informáciu o triede, do ktorej sa má zaradiť neznámy obraz, potom stratu pre obraz **x** vyjadríme funkciou *Q* argumentu *Ω – ω,* t.j.

$$Q(\Omega - \omega) = Q(\omega_r - d(\mathbf{x}, \mathbf{q})),$$

  pričom často sa volí funkcia *Q* ako norma vektora.

- Učiace klasifikátory musia byť nastavené so zreteľom na všetky možné obrazy, preto pracujú s integrálnymi kritériami *strednej straty* v tvare

$$J(\mathbf{q}) = \int_{XxO_1} Q(\Omega - \omega)\, dF(\mathbf{x}, \Omega),$$

čo sa vzhľadom na že $Q(\Omega-\omega)$ závisí od Ω a cez rozhodovacie pravidlo aj od **x** a **q**, tak sa dá písať tento integrál ako

$$J(\mathbf{q}) = \int_{XxO_1} Q(\Omega, \mathbf{x}, \mathbf{q})\, dF(\mathbf{x}, \Omega),$$

kde O_1 je množina všetkých informácií Ω o triede, t.j. $O_1 = \{\omega_1, ..., \omega_R\}$ a $F(\mathbf{x}, \Omega)$ je distribučná funkcia združeného rozloženia náhodných veličín **x** a Ω.

- Pretože pri konečnom počte R tried je rozloženie veličiny Ω na O_1 popísané frekvenčnou funkciou, môžeme strednú stratu upraviť na tvar

$$J(\mathbf{q}) = \sum_{r=1}^{R} P(\omega_r) \int_{XxO_1} Q(\mathbf{x}, \omega_r, \mathbf{q})\, dF(\mathbf{x}|\omega_r) =$$

$$\sum_{r=1}^{R} P(\omega_r) \int_{XxO_1} Q(\mathbf{x}, \omega_r, \mathbf{q})\, p(\mathbf{x}|\omega_r)\, d\mathbf{x},$$

čo interpretujeme tak, že v danej úlohe sa obrazy **x** a informácie $\Omega \in O_1$, t.j. triedy nevyskytujú rovnako často a preto pri výpočte strednej straty treba rešpektovať pravdepodobnosť ich výskytu.

- cieľom učenia je minimalizovať strednú stratu, t.j. nájsť taký parameter $\mathbf{q}^*$, pri ktorom stredná chyba nadobúda minimum. Ak $Q(\Omega, \mathbf{x}, \mathbf{q})$ zvolíme ako konvexnú funkciu, potom aj $J(\mathbf{q})$ je konvexná a nutnou a postačujúcou podmienkou na existenciu minima bude

$$\operatorname{grad}_{\mathbf{q}} J(\mathbf{q}^*) = 0.$$

- Takáto algebrická rovnica sa rieši buď analyticky alebo rekurentne. Analyticky je možné riešiť len veľmi jednoduché prípady. Ale obvykle dopredu nepoznáme distribučnú funkciu $F(\mathbf{x}, \Omega)$ a preto nemožno rovnicu bezprostredne riešiť.

- Z toho dôvodu ťažisko jednej skupiny metód učenia spočíva v odhadovaní podmienených hustôt pravdepodobností a apriórnych pravdepodobností na základe informácie z trénovacej množiny.

- Rekurentné riešenie vyžaduje prechod k novej hodnote **q**, ktorá sa približuje optimálnej hodnote $\mathbf{q}^*$. Vo všeobecnosti je možné použiť gradientné metódy, ale opäť narazíme na problém, že nepoznáme $F(\mathbf{x}, \Omega)$. Riešením býva stochastická aproximácia, ktorá nie je viazaná na znalosť distribučnej funkcie, tieto metódy učenia sú potom založené na priamej minimalizácii strát.

- induktívnosť a sekvenčnosť učenia – induktívnosť – z konečnej trénovacej množiny usudzujeme na nekonečnú množinu obrazov – sekvenčnosť – prvky trénovacej množiny sa predkladajú obvykle postupne.

- z konečnej trénovacej množiny nezískame optimálny parameter $\mathbf{q}^*$, ale parameter $\hat{\mathbf{q}}^*$, ktorý minimalizuje výberovú (empirickú) stratu´. Snažíme sa, aby pravdepodobnosť, že rozdiel $J(\mathbf{q}^*)$ a $J(\hat{\mathbf{q}}^*)$ je menší ako epsilon, bola menšia ako 1 – α.

Metódy učenia založené na odhadovaní hustôt pravdepodobností

- tieto metódy prevádzajú určenie minima algebrickej rovnice na stanovenie neznámej distribučnej funkcie $F(\mathbf{x}, \Omega)$ z trénovacej množiny.

- ukážeme to na klasifikácii založenej na kritériu minimálnej chyby. Tam potrebujeme určiť neznáme apriórne pravdepodobnosti, ktoré obvykle dopredu nepoznáme. Keďže hustoty pravdepodobnosti $p(\mathbf{x}|\omega_r)$ a $P(\omega_r)$ sú funkcie definované na celom obrazovom priestore, na základe trénovacej množiny určíme ich aproximácie $\hat{p}(\mathbf{x}|\omega_r)$ a $\hat{P}(\omega_r)$.

- požiadavky kladené na odhad hustôt pravdepodobnosti: nestrannosť, konzistentnosť (konverguje k správnej hodnote pri zväčšujúcom sa počte prvkov trénovacej množiny) a efektívnosť.

- dve skupiny odhadu podľa veľkosti apriórnej informácie: s nenulovou apriórnou informáciou a nulovou apriórnou informáciou. Obvykle sa vie, o tvare hustoty rozloženia obrazov triedy ω_r, pričom parameter $\mathbf{q}$ je dopredu neznámy. Najčastejšie sa uvažuje normálne rozdelenie. Metódy s apriórnou informáciou sú podstatne efektívnejšie a vyžadujú aj menšiu trénovaciu množinu.

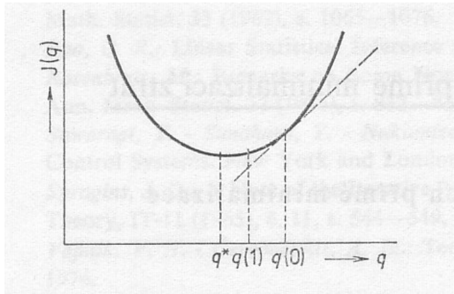
- Trénovacia množina ako *n* - rozmerný náhodný výber, ktorý má svoju strednú hodnotu aj smerodajnú odchýlku. Výberový priemer je konzistentným odhadom pre túto strednú hodnotu, pričom výberový priemer sa počíta ako lineárny odhad.

- okrem lineárnych odhadov sa používajú aj ďalšie metódy, založené na pravdepodobnosti.

Metódy učenia založené na priamej minimalizácii strát

- rekurentné riešenie – spočíva v tom, že k optimálnemu parametru $\mathbf{q}^*$ sa dopracujeme postupným vypočítavaním jeho odhadov. Treba určiť funkciu, ako sa zo starého odhadu vypočíta nový odhad. Takýmto rekurentným spôsobom vieme riešiť aj úlohy, ktoré sa analyticky nedajú vyjadriť.

- Príkladom sú gradientné metódy, ktoré tvoria odhady pomocou dotyčníc (derivácií) funkcie, ale prakticky sú pre učenie nepoužiteľné.

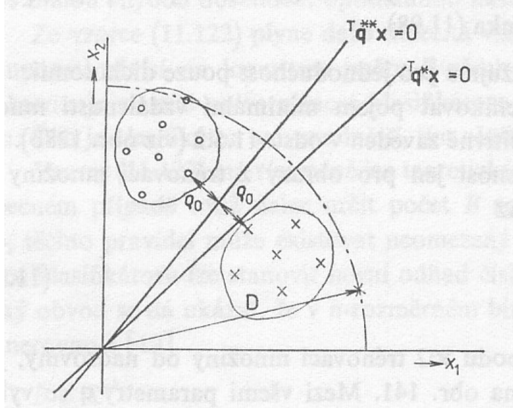


Obr. 115. Princip gradientních metod

- základný tvar algoritmu priamej minimalizácie – použijeme metódu stochastickej aproximácie

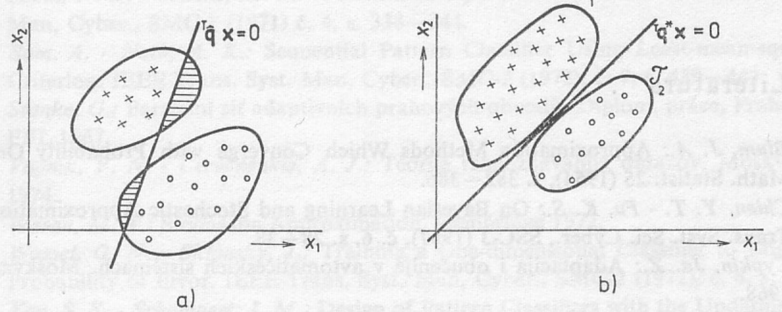
- riešenie pre neseparabilné triedy -

- riešenie pre separabilné triedy – Novikoffova veta, že v konečnom počte krokov zostrojíme rozdeľujúce nadroviny – pričom korekcie robíme pri nesprávnej klasifikácii – hľadáme parameter $\mathbf{q}^{**}$, ktorý je optimálny a zároveň najlepší v tom zmysle, že polí vzdialenosť medzi najbližšími dvoma obrazmi z trénovacej množiny



Obr. 142. Výběr parametru **q****

- dôležitý je výber trénovacej množiny



Obr. 144. Výběr trénovací množiny
a) nevhodný; b) vhodný