

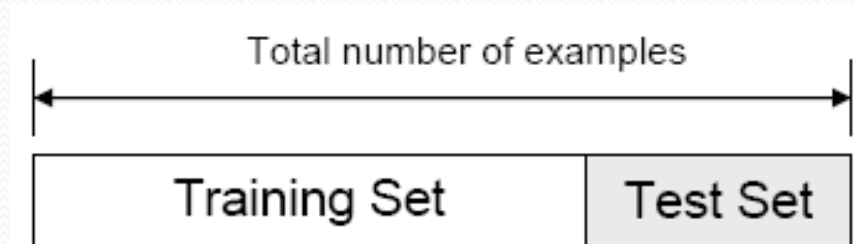
Rozpoznávanie obrazcov

Odhad chyby
Chybová funkcia klasifikátora
Matica zámen, ROC krivka

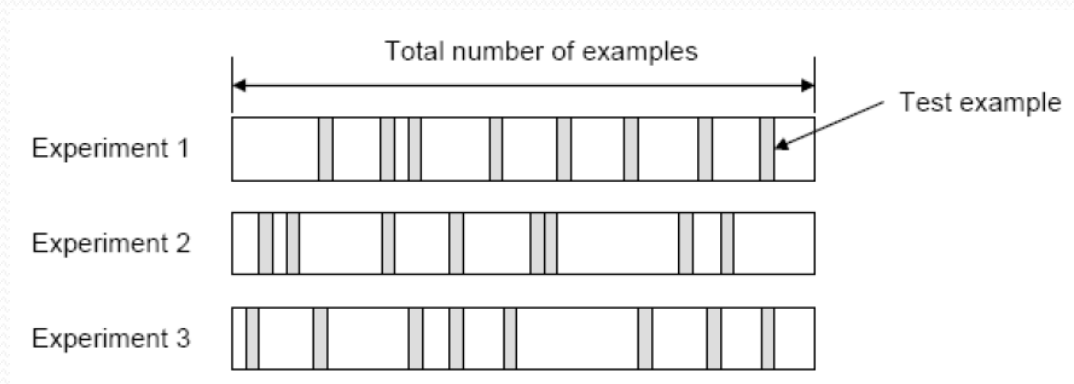
7.4.2014

Odhad chyby

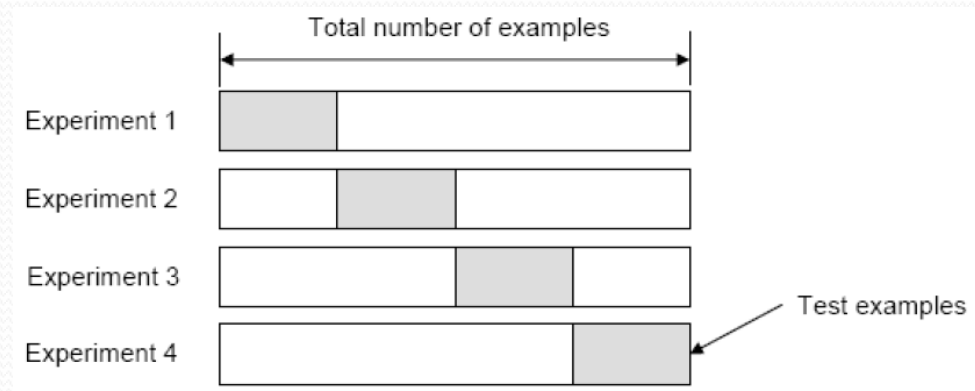
- Trénovacia množina
 - trénuje klasifikátor
- Testovacia množina
 - určená na odhad chyby klasifikátora
- Problémy:
 - riedka množina dát
 - “nešťastné” rozdelenie
- Riešenie (za cenu vyšších výpočtových nákladov):
 - Krížová validácia
 - Bootstrap



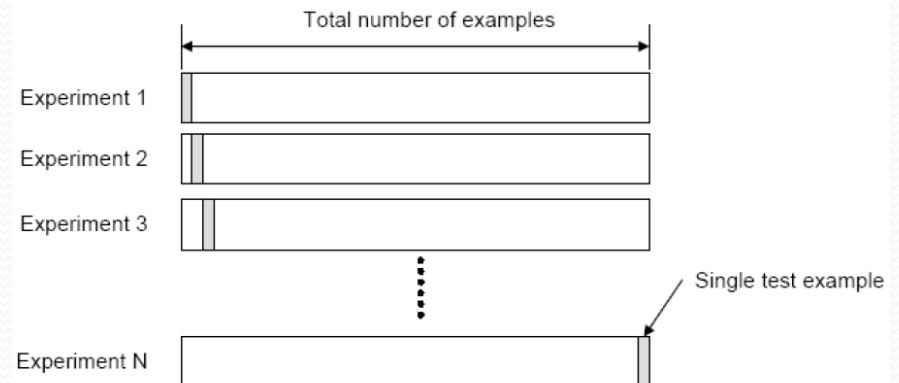
- Náhodné prevzorkovanie



- K-násobná krížová validácia



- Vrstvená K-násobná krížová validácia
 - rozdelenie množiny X na podmnožiny nie je náhodné
 - jednotlivé podmnožiny zachovávajú distribúciu tried
 - podiel príkladov z jednotlivých tried je rovnaký tak v pôvodnej množine X , ako aj v každej z jej podmnožín X_1, \dots, X_K
- Krížová validácia s vynechaním jednej vzorky
 - K je rovné počtu vzoriek



K

- závisí na veľkosti množiny
- Pre veľmi veľké množiny
 - $K=3$ postačuje
- Pre veľmi malé množiny
 - validácia s vynechaním jednej vzorky
- Najviac sa používa
 - $K=10$

Chybová funkcia klasifikátora

- Niektoré chyby klasifikácie môžu byť drahšie (majú horšie následky) ako iné
- Príklad:
 - neodhalenie smrteľnej choroby, ktorá by sa dala ľahko liečiť liekmi, ktoré nemajú vedľajšie účinky. Teda neklasifikovanie choroby v prípade chorého človeka má horšie následky ako nasadenie liekov u zdravého človeka.

| Skutočnosť \ Rozhodnutie | Neliečiť | Liečiť |
|-----------------------------|------------------|--------------------------|
| Zdravý | Správne | Malé vedľajšie účinky |
| Chorý | Smrteľná choroba | Správne |

- rôzne typy chýb:
 - Falošne pozitívna klasifikácia
 - Falošne negatívna klasifikácia

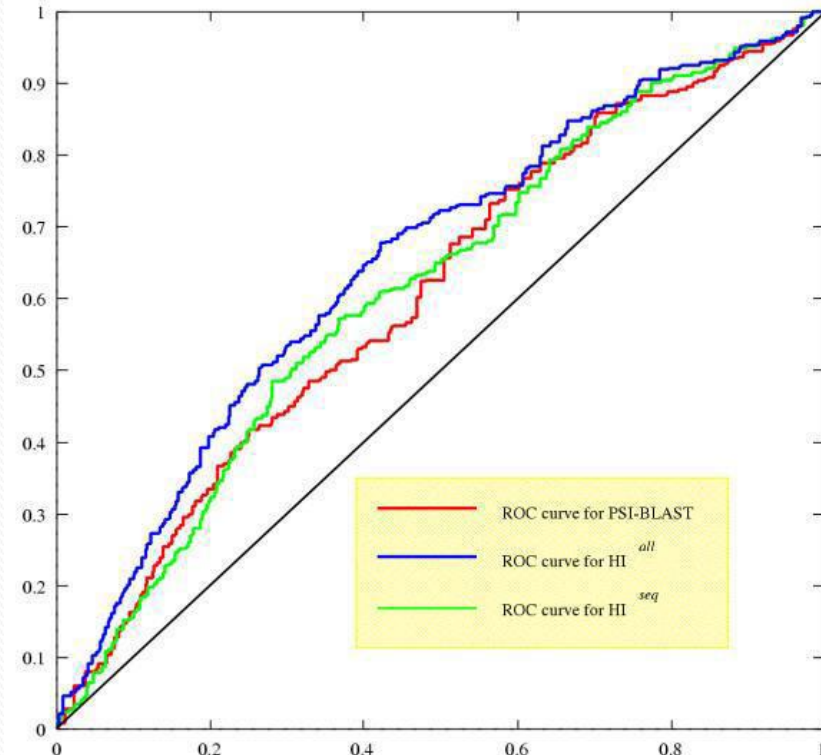
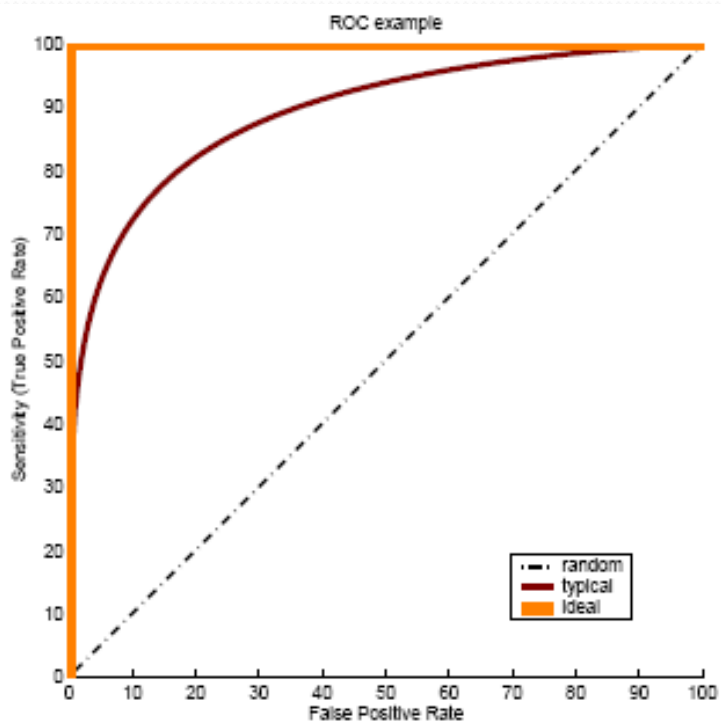
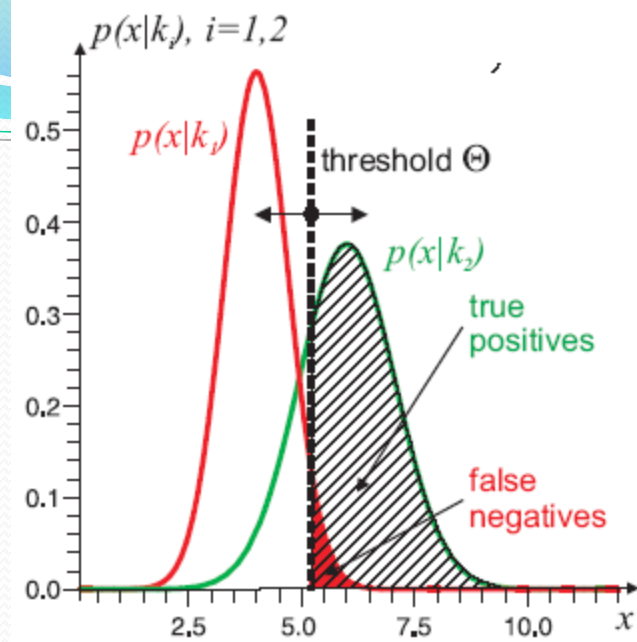
Matica zámen

| true class i | class j predicted by a classifier | | | | | | | | | | |
|----------------|-------------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9' | 'R' |
| '0' | 97 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| '1' | 0 | 98 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| '2' | 0 | 0 | 96 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |
| '3' | 0 | 0 | 2 | 95 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| '4' | 0 | 0 | 0 | 0 | 98 | 0 | 0 | 0 | 0 | 2 | 0 |
| '5' | 0 | 0 | 0 | 1 | 0 | 97 | 0 | 0 | 0 | 0 | 2 |
| '6' | 1 | 0 | 0 | 0 | 0 | 1 | 98 | 0 | 0 | 0 | 0 |
| '7' | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 98 | 0 | 0 | 1 |
| '8' | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 96 | 1 | 1 |
| '9' | 1 | 0 | 0 | 0 | 3 | 1 | 0 | 0 | 0 | 95 | 0 |

- Špeciálna trieda R – zamietnutie každej triedy

ROC krivky

- <http://www.mathworks.com/matlabcentral/fileexchange/19950-roc-curve>



McNemarov test

- Pre oba klasifikátory (K_1 , K_2) učíme nasledovné hodnoty:
 - n_{00} : počet vzoriek správne klasifikovaných oboma klasifikátormi
 - n_{01} : počet vzoriek správne klasifikovaných K_1 ale nesprávne K_2
 - n_{10} : počet vzoriek nesprávne klasifikovaných K_1 ale správne K_2
 - n_{11} : počet vzoriek nesprávne klasifikovaných oboma klasifikátormi