

Vytěžování Dat

Cvičení 12 – Vyhodnocení modelů

Miroslav Čepek, Michael Anděl

16.12.2014



Evropský sociální fond
Praha & EU: Investujeme do vaší budoucnosti

Fakulta Elektrotechnická, ČVUT

Zadání domácího úkolu

- ▶ Cílem domácího úkolu je vytvořit několik klasifikačních modelů (rozhodovací strom, logistická regrese a naivní bayes) v Rapidmineru a vyhodnotit jeho přesnost.
- ▶ Pro učení modelu použijte data *Breast cancer*.
- ▶ Spočítejte:
 1. Klasifikační chybu jednotlivých modelů: rozhodovacího stromu, logistické regrese a naivní bayes pomocí 10 násobné cross-validace. Porovnejte, která modelovací metoda dává z pohledu přesnosti nejlepší výsledky.
 2. Pomocí uzlu *Compare ROCs* porovnejte vytvořte ROC křivky pro jednotlivé modely a porovnejte je. Srovnajte výkon klasifikátorů s náhodným klasifikátorem.
 3. Pomocí rozhodovacího stromu spočítejte " míru důvěry" (confidence) zařazení jednotlivých instancí do třídy benign.
 4. Pomocí míry důvěry vytvořte tabulku pro zobrazení ROC křivky pro třídu Benign. Vytvořte graf z Vaší tabulky a porovnejte s grafem z Rapidmineru.

- ▶ Obsahem zprávy bude:
 - ▶ zhodnocení přesnosti klasifikátorů s jen na základě jejich přesnosti.
 - ▶ zhodnocení přesnosti klasifikátorů s přihlédnutím k ROC křivkám. Jsou klasifikátory lepší než náhodný klasifikátor?
 - ▶ tabulka a výsledný graf pro zobrazení ROC křivky.
 - ▶ Který klasifikátor byste použili, pokud máte možnost léčit jen 1% ze všech pacientů s Maligním nádorem (třída Malignant)? Změní se vaše volba, pokud máte možnost léčit 5% pacientů?