

# Vytěžování Dat

## Cvičení 12 – Vyhodnocení modelů

Miroslav Čepek

Fakulta Elektrotechnická, ČVUT

13.12.2011

# Zadání domácího úkolu

- V sw Rapidminer vytvořte rozhodovací strom z dodaných dat a vyhodnoťte jeho přesnost.
- Použijte data (*Breast cancer*) ze stránek předmětu, tato data nahrajte do SW Rapidminer.
- Spočítejte:
  - 1 Chybu rozhodovacího stromu, logistické regrese a bayesovské sítě<sup>1</sup> pomocí 10 násobné cross-validace. Porovnejte, která modelovací metoda dává nejlepší výsledky.
  - 2 Pomocí rozhodovacího stromu spočítejte "míru důvěry" (confidence) zařazení jednotlivých instancí do třídy benign.
  - 3 Pomocí míry důvěry vytvořte tabulku pro ROC a Lift pro třídu Benign. Výpočet demonstруйте!
  - 4 Hodnoty z tabulky zanešte do grafu a učiňte závěr o schopnosti rozhodovacího stromu klasifikovat *Breast cancer* data.

---

<sup>1</sup>vše z prostředí Rapidmineru

- Obsahem zprávy bude:
  - ▶ Popis výsledků jednotlivých modelovacích metod a závěr o tom, která metoda je nejlepší,
  - ▶ tabulka obsahující hodnoty pro ROC a Lift (alespoň 10 hodnot) a ukázkou výpočtu,
  - ▶ graf ROC a Lift křivky,
  - ▶ váš komentář k předchozím bodům.