

České vysoké učení technické v Praze

Fakulta elektrotechnická

**Katedra kybernetiky
Katedra počítačů**

IDA

**Intelligent Data Analysis
RESEARCH GROUP**



Vytěžování dat – cvičení VI

Klasifikační a regresní stromy

Petr Pošík: posik@labe.felk.cvut.cz

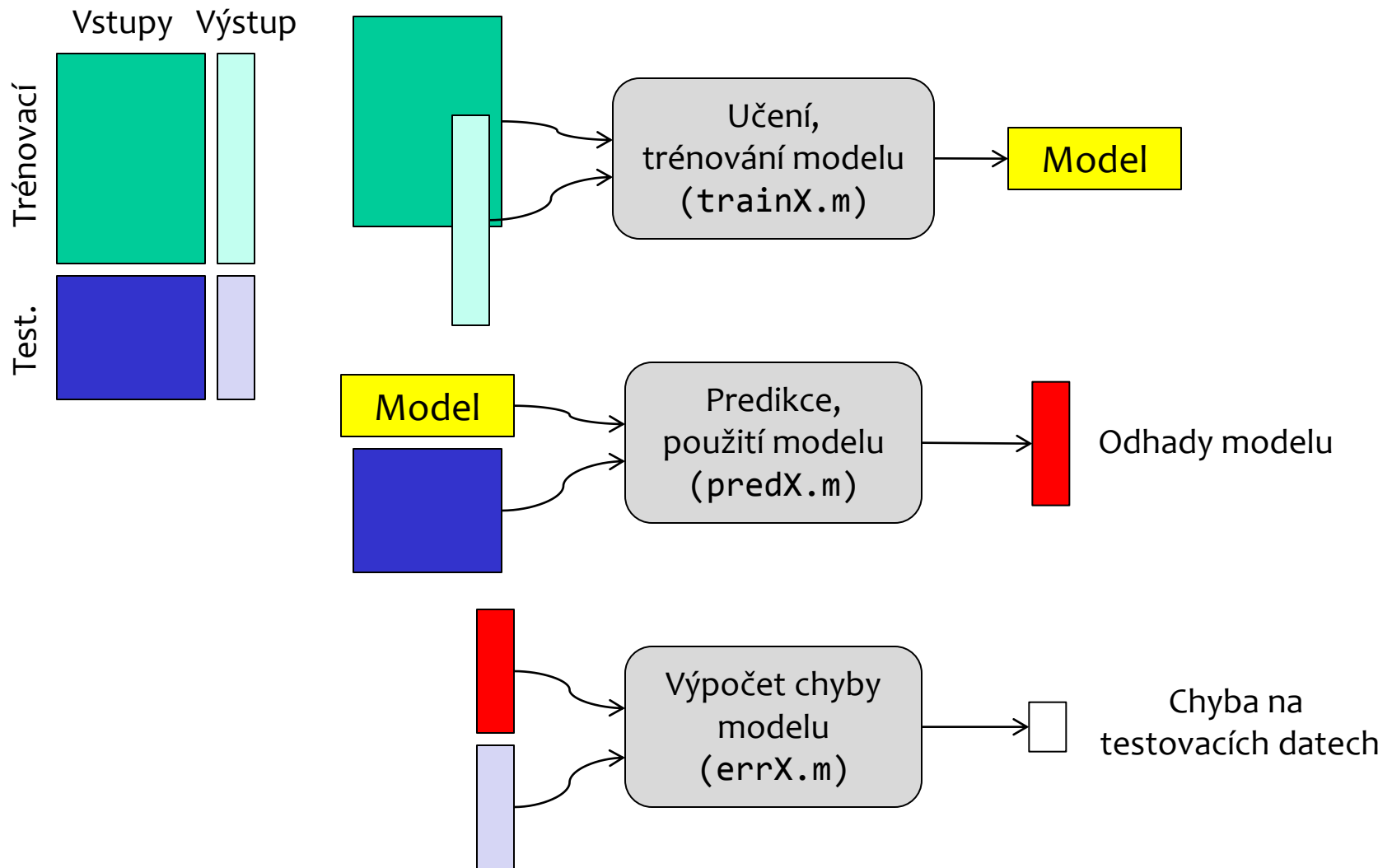
Pavel Kordík: kordikp@fel.cvut.cz

Program cvičení

- Klasifikační stromy
 - Vizualizace, chyby vs. „ohebnost“ modelu
- Regresní stromy
 - Vizualizace, chyby vs. „ohebnost“ modelu
- Křivky učení klas. a reg. stromů

- Zadání první semestrální úlohy

Typický postup učení s učitelem

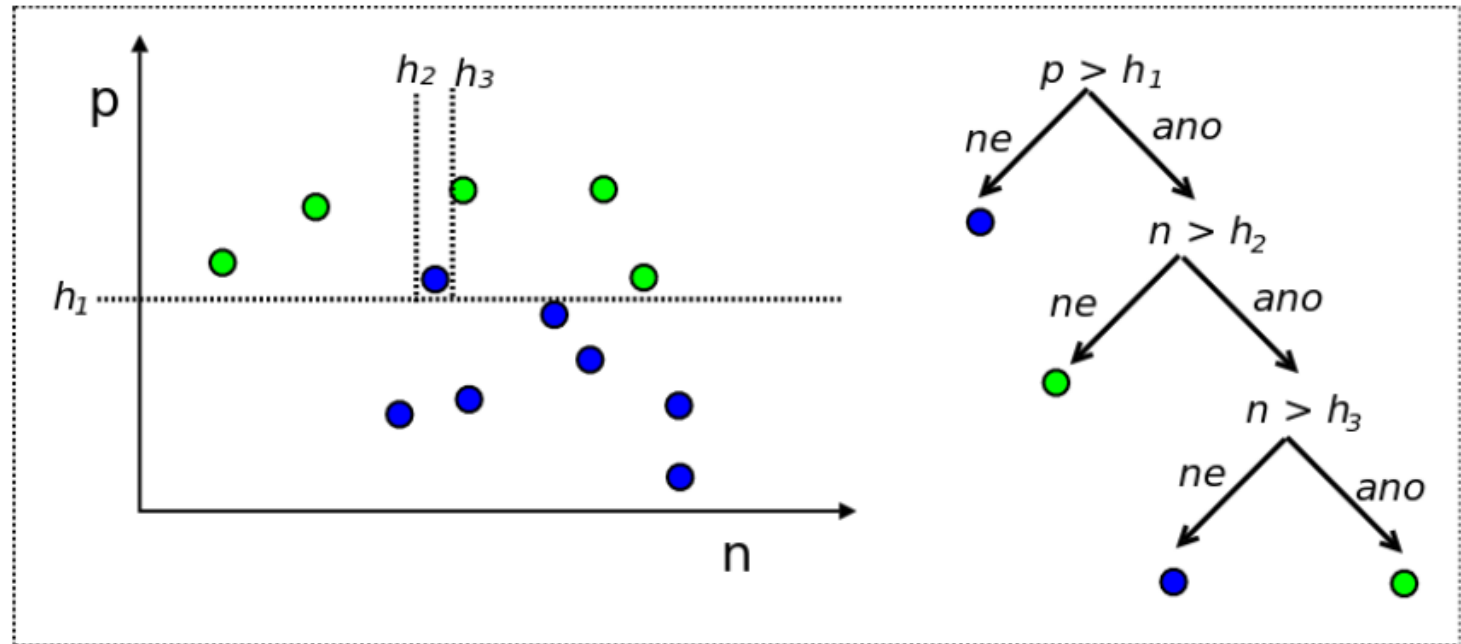


Rozhodovací stromy

Klasifikační rozhodovací stromy: Připomenutí

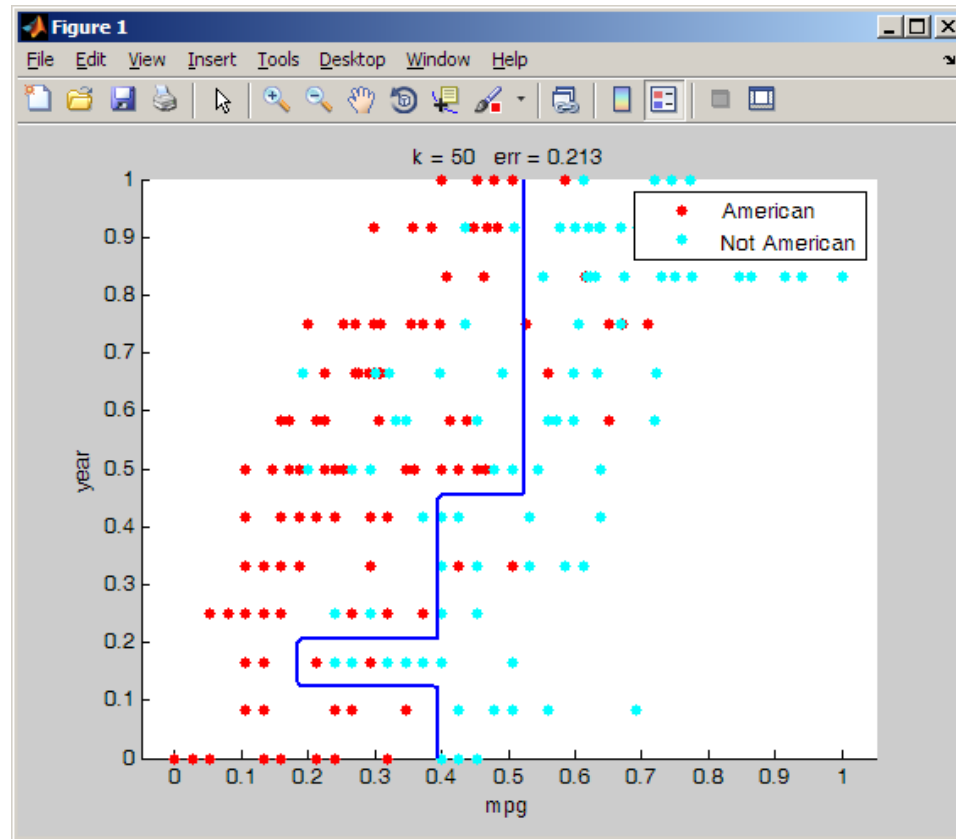
■ Jak strom funguje?

Separace v prostoru dvou reálných příznaků



Klasifikační stromy: Vizualizace

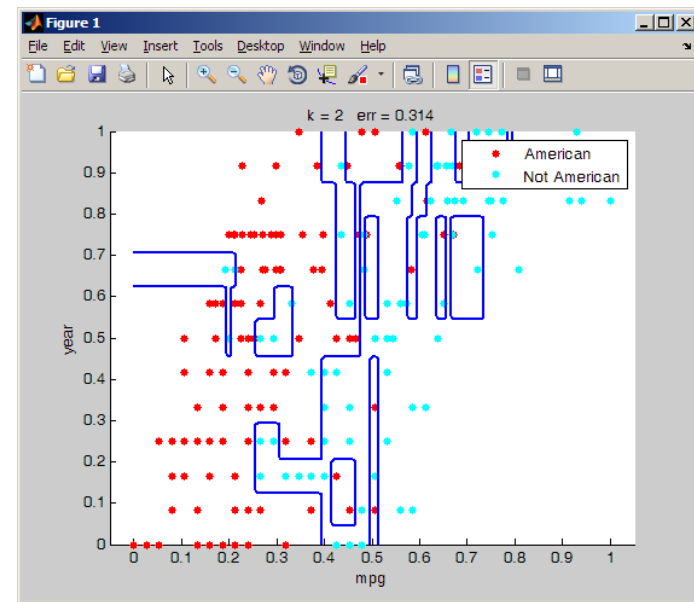
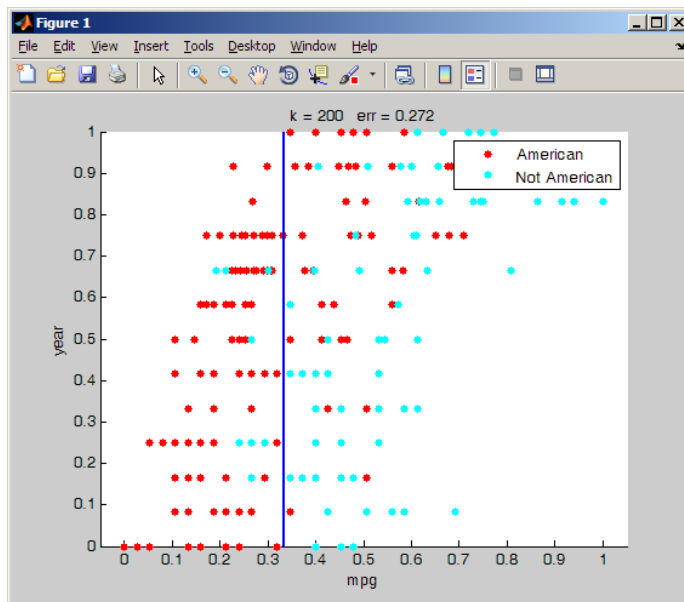
- viz `scrVizClassTree.m`



Klasifikační stromy:

Chyba vs. „ohebnost“ modelu

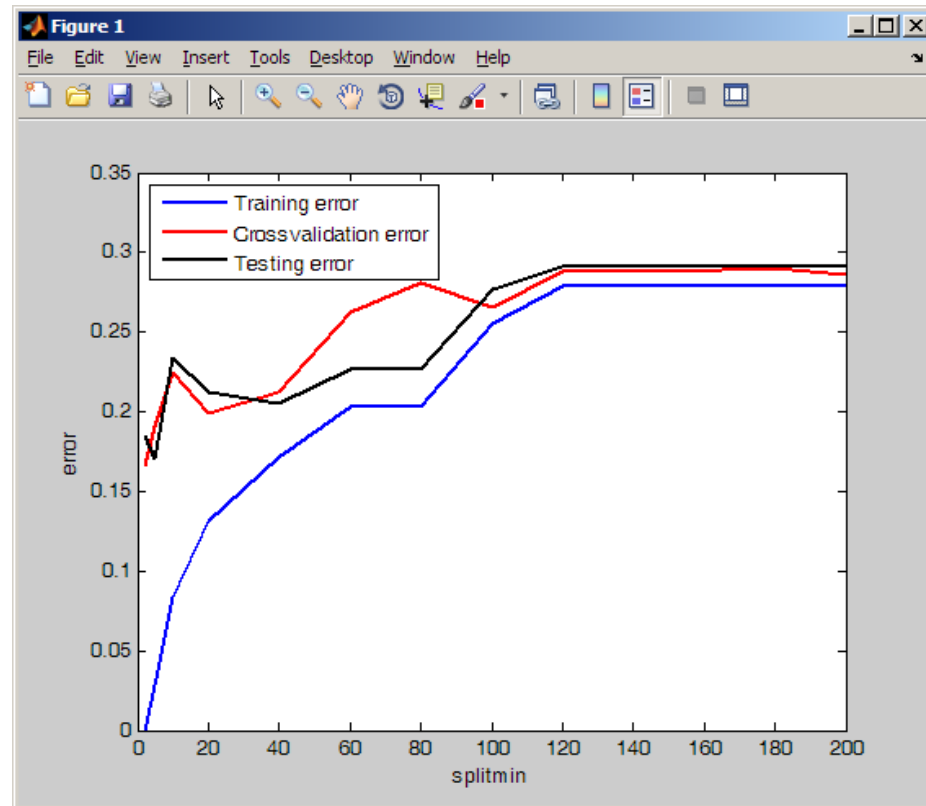
- Parametr *splitmin* určuje ohebnost modelu.
 - Je to minimální počet trénovacích případů v uzlu, aby jej algoritmus ještě mohl dále dělit
 - Jaký další parametr stromu by mohl hrát podobnou roli?



Klasifikační stromy:

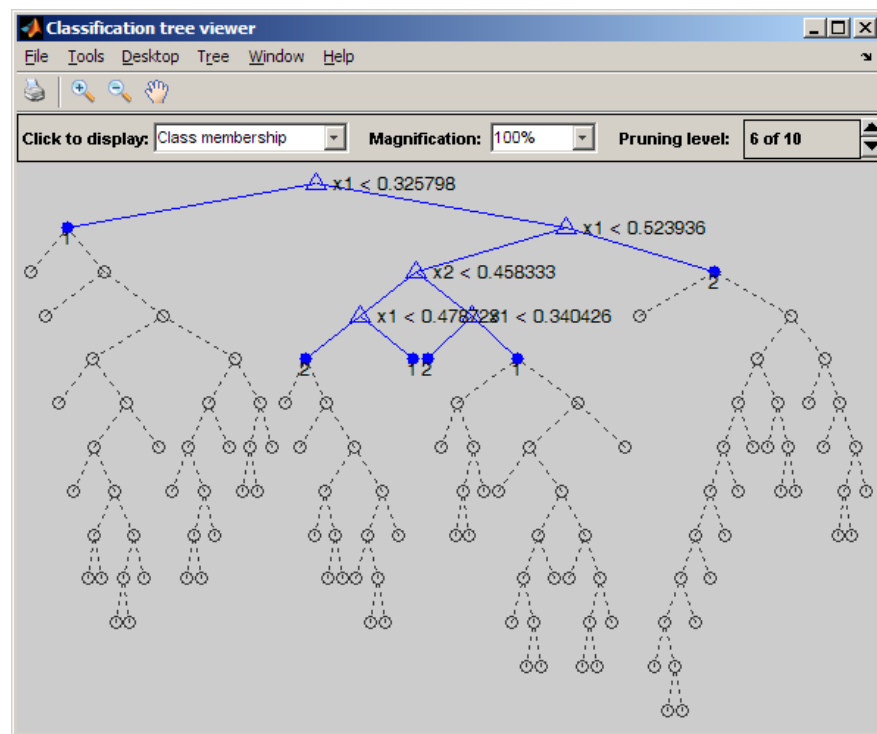
Chyba vs. „ohebnost“ modelu II

- Závislost chyby stromu na parametru *splitmin*
 - viz `scrClassTTErrorTree.m`
 - Jak to, že trénovací chyba pro *splitmin* = 2 je nulová?



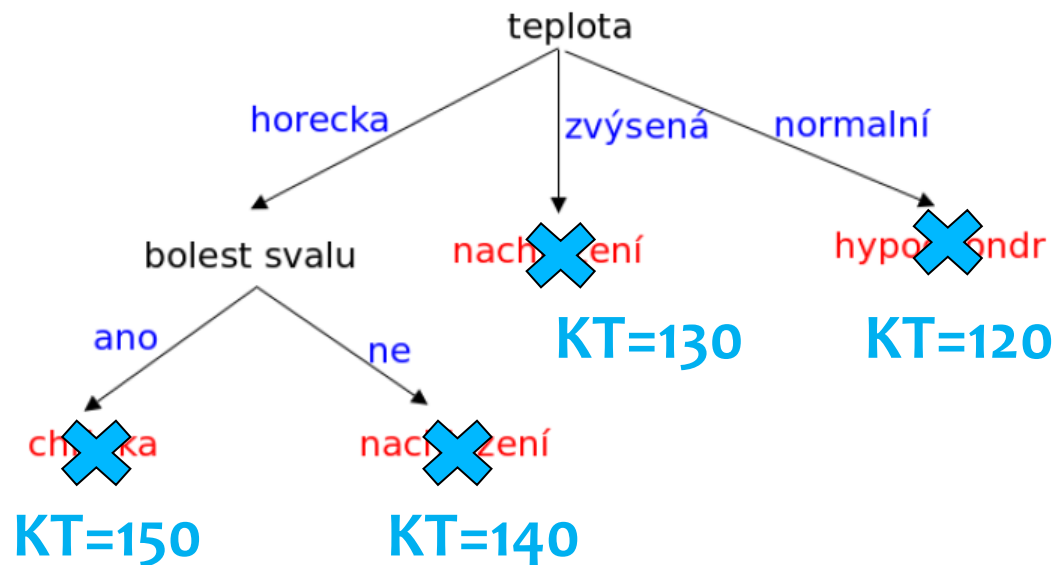
Stromy: Prořezávání

- Snižuje složitost modelu, odtraňuje „zbytečné“ větve, zlepšuje generalizaci
- Po spuštění `scrVizClassTree.m` byste měli mít ve workspace proměnnou `model`
- Zavolejte `model=prune(model)`
- `view(model)` a hrajte si s úrovní prořezání



Regresní rozhodovací stromy

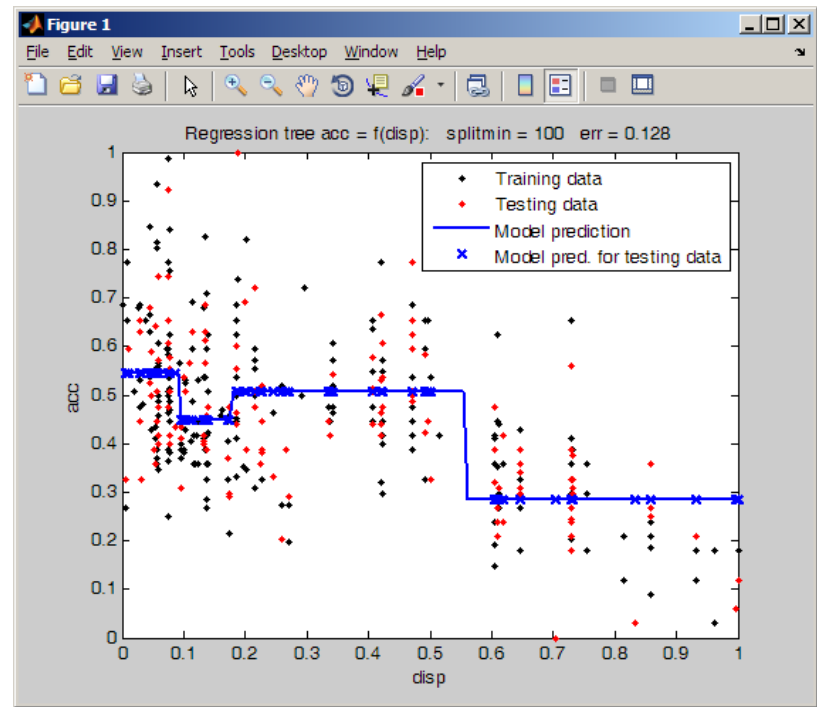
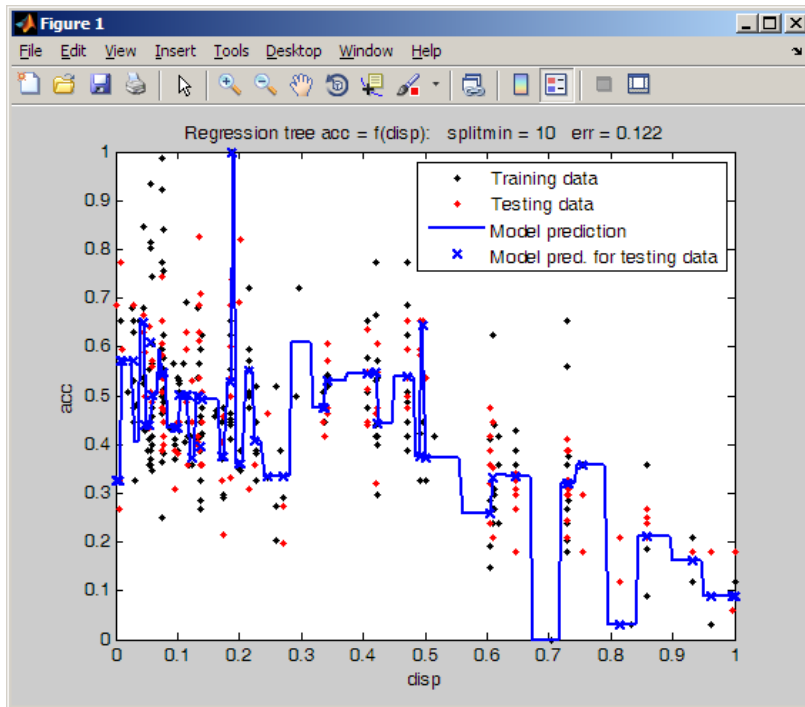
Regresní stromy: Jak se liší od klasifikačních?



- Místo nominální veličiny (chřipka, nachl., hypoch.) modelují spojitou veličinu, např. krevní tlak (KT).
- Hodnota bývá průměr příslušných tr. případů.

Regresní stromy: Predikce

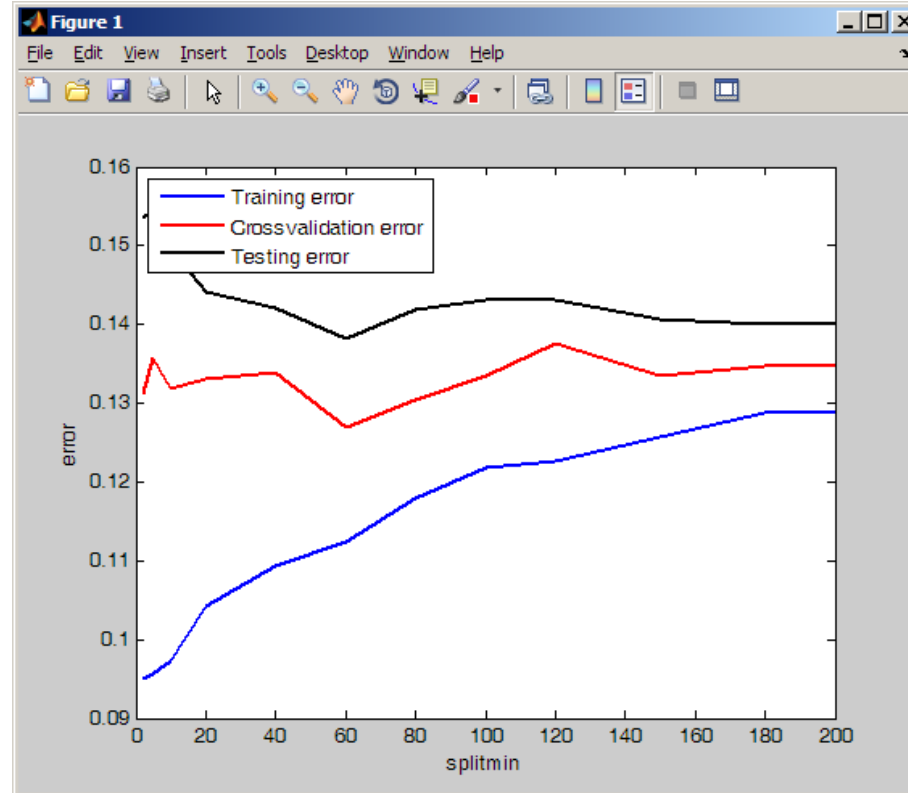
- viz `scrVizRegrTree.m`
- $acc = f(displ)$
- Experimentujte s parametrem *splitmin*



Regresní stromy

Chyba vs. „ohebnost“ modelu

- Závislost chyby stromu na parametru *splitmin*
 - viz `scrRegrTTErrorTree.m`

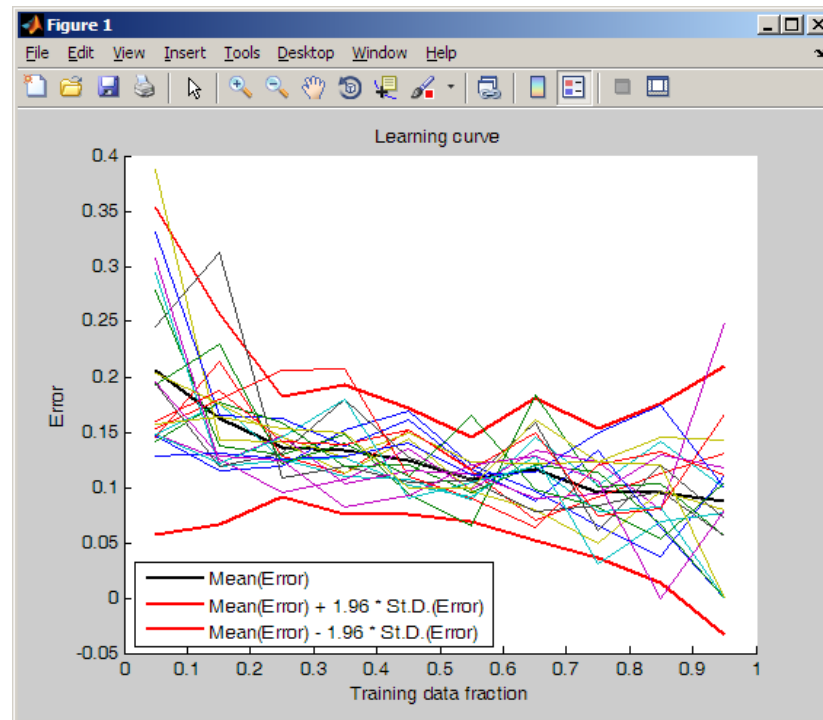


KŘIVKY UČENÍ

Klasifikační stromy:

Křivka učení

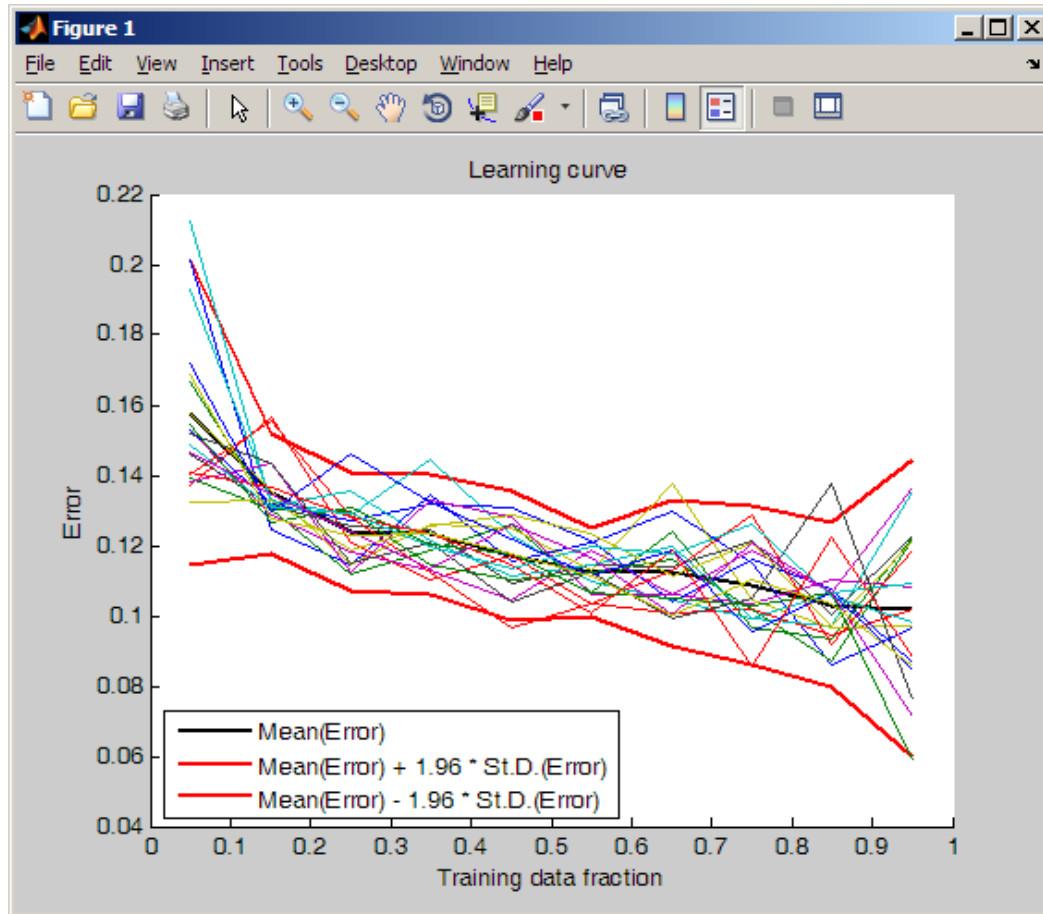
- Opakování z přednášky:
 - přesnost (chyba) modelu (na testovacích datech) v závislosti na velikosti trénovacích dat
 - viz `scrClassLearningCurveTree.m`



Regresní stromy

Křivka učení

- viz `scrRegrLearningCurveTree.m`



Zadání první semestrální práce

Viz CourseWare na

<http://cw.felk.cvut.cz/doku.php/courses/y336vd>