

Analýza a klasifikace dat – přednáška 5 – doplnění

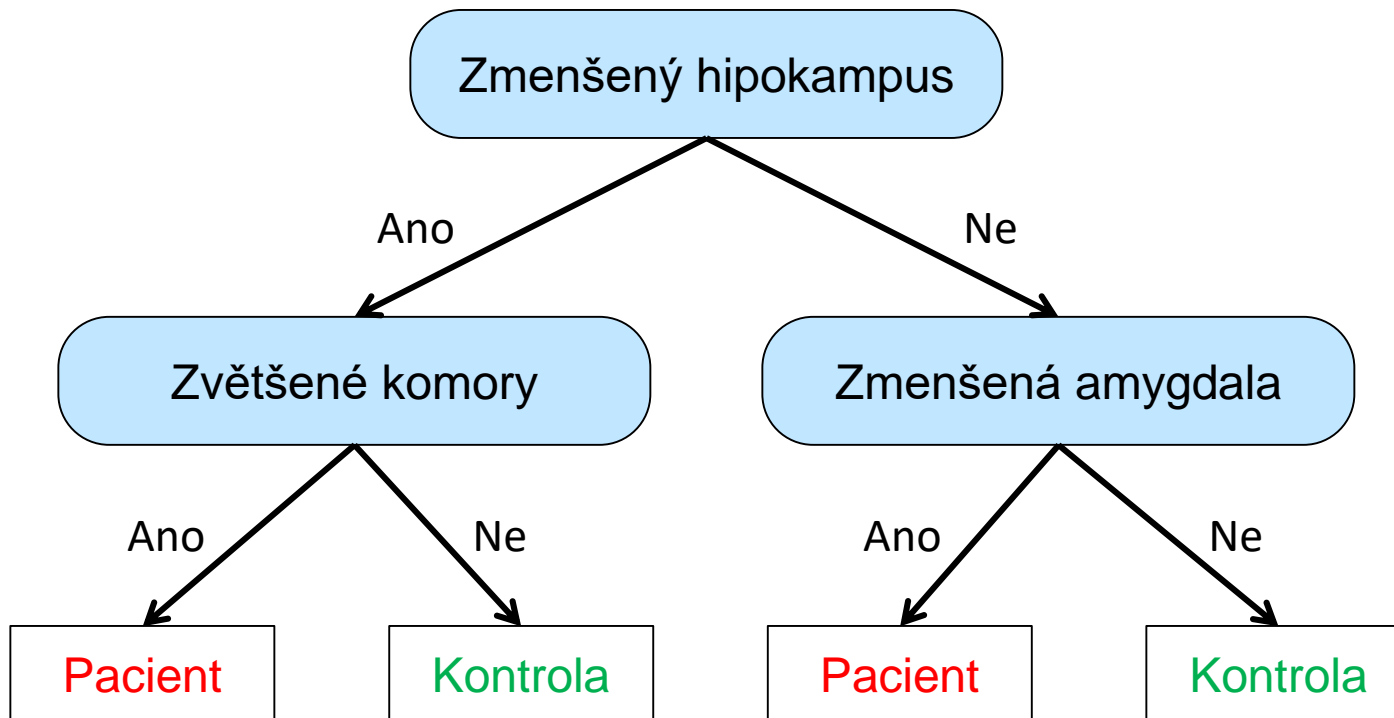


RNDr. Eva Koriťáková, Ph.D.

Podzim 2017

Klasifikační (rozhodovací) stromy

Princip: Postupné rozdělování datasetu do skupin podle hodnot jednotlivých proměnných.

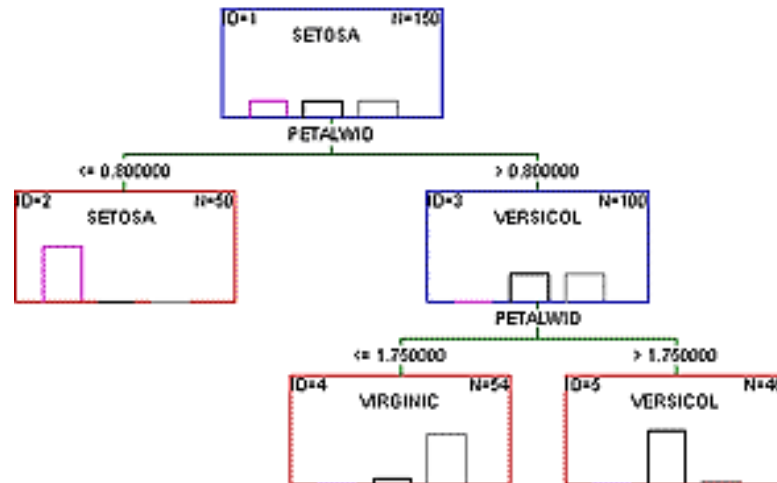


Patří mezi metody sekvenční klasifikace.

Podrobnější informace: <https://www.iba.muni.cz/res/file/ucebnice/komprdova-rozhodovaci-stromy-lesy.pdf>

Klasifikační (rozhodovací) stromy - doplnění

- kategoriální proměnné – rozdělení se provede podle kategorií (viz obrázek na předchozím slidu)
- spojité proměnné – nalezne se nejlepší dělicí hodnota a pak dojde k rozdělení; proměnná se může použít i vícekrát s různými dělicími hodnotami



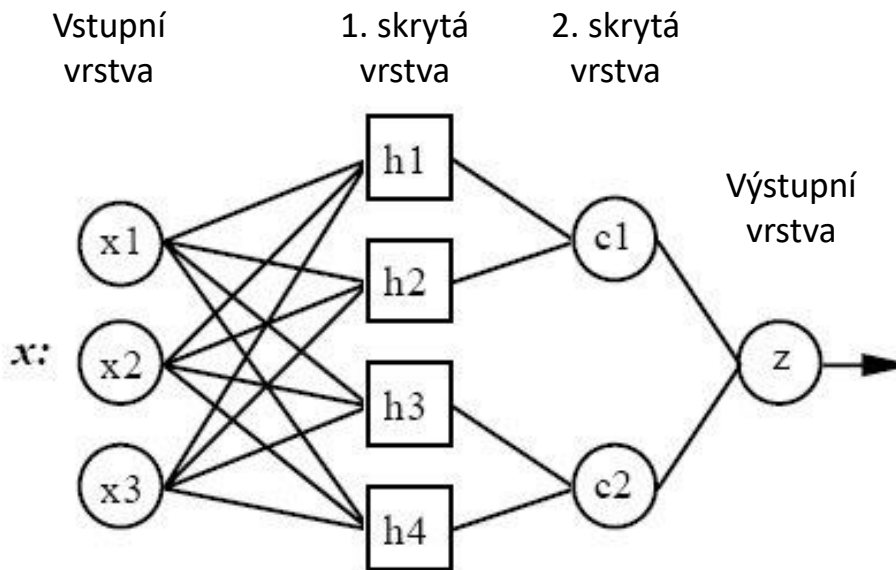
- postup vytvoření stromu:
 - nejprve se vytvoří strom, kde v „listech“ (terminálních uzlech) jsou vždy jen subjekty z jedné skupiny
 - následuje tzv. „prořezávání stromu“ – odstraní se ty uzly, které jsou nejvíce zbytečné (příliš velký strom je totiž zpravidla přeučeny a funguje špatně na testovacích datech)

Klasifikační (rozhodovací) lesy

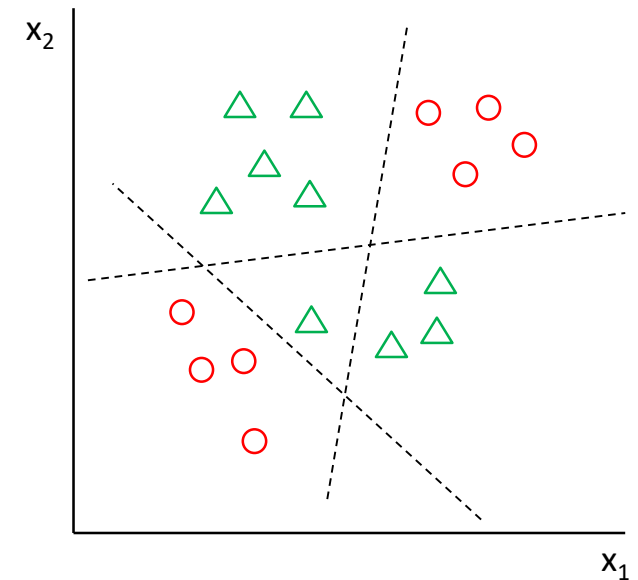
- klasifikační lesy – použití více klasifikačních stromů ke klasifikaci
- použije se zpravidla jen část dat na vytvoření (tzn. naučení) jednotlivých stromů:
 - náhodně vybrané subjekty
 - náhodně vybrané proměnné
- finální klasifikace testovacích dat se provede „hlasováním“ výsledků z klasifikace pomocí jednotlivých stromů

Neuronové sítě

Princip: Postupné učení neuronové sítě (tzn. postupné nastavování vah u jednotlivých neuronů), aby byla chyba klasifikace trénovací množiny minimální. Umožňuje i nelineární klasifikaci.



Nelineární klasifikace



○ pacienti
△ kontroly

Více typů neuronových sítí – např.:

- Vícevrstvé neuronové sítě typu perceptron
- RBF (Radial Basis Function) sítě
- LVQ (Learning Vector Quantization) sítě

Typy neuronových sítí

- Vícevrstvý perceptron (Multilayer Perceptron):
 - lze si představit jako modely logistické regrese seřazené do sítě (obrázek viz předchozí slide)
 - aktivační funkce: hyperbolický tangens, logistická funkce nebo případně další
- RBF (Radial Basis Function) sítě:
 - kombinace učení s učitelem a bez učitele
 - aktivační funkce: radiální bázové funkce
- LVQ (Learning Vector Quantization) sítě:
 - založené na adaptivním shlukování (obdoba k-means)

