

Kvíz
BI-ZNS Cvičení 10

Autor kvízu: Mgr. Ing. Ladislava Smítková Janků, Ph.D.

Jméno studenta: JAN ŠMERDA
Studijní skupina: F

1. Co je to rozhodovací strom?
2. Vysvětlete rozdíl mezi klasifikačními stromy a regresními stromy.
3. Rozhodněte, která tvrzení jsou pravdivá:
(a) stromy typu C&RT jsou binární stromy **ANO**
(b) kritériální statistika Entropie má obecně vyšší hodnotu než kritériální statistika Chyba klasifikace, ale nižší než Giniho index **NE**
(c) kritériální statistika Entropie má obecně vyšší hodnotu než kritériální statistika Chyba klasifikace i než Giniho index **ANO**
(d) stromy typu CHAID jsou binární stromy **NE**
4. V bodech popište postup konstrukce C&RT stromů.
5. Co je to prořezávání stromu na základě klasifikační chyby?
6. Napište vzorec pro výpočet Giniho indexu v kořenovém uzlu klasifikačního stromu typu C&RT.
7. Vyjmenujte zastavovací pravidla růstu stromu.
8. Popište podstatu konstrukce stromů typu PRIM.
9. Na jakém principu je založena metoda Random Forest?
10. Je v úloze klasifikace přesnější jeden rozhodovací strom nebo les rozhodovacích stromů? **les**
11. Co je to boosting?
12. Jak se konstruuje trénovací a testovací množina ze souboru dat (pozorování)?
13. Vysvětlete pojem přeučení klasifikačního stromu.

1) Skladba, která nám umožní rozdělit hodnot do tří (predice)
podle nainstalované referenční sady dat.

2) Klasifikace - přeje pro vstup nainstalované do výstupu předem známých kategorií
regrese - položí se hodnotami proti k užší a omezené spíše výstup

3) 1) rozhodne, jaké řešení skromu poradijeme

2) Každá část dělíme na třídy pomocí rozhodovacích stromů,
abychom v každém kroku dosáhli co nejvyšší
číslo (greedy algorithms)

3) Po dokonalém rozdělení končíme

5) Strom se projevuje tak, aby se do všech tří skupin
stejná chyba klasifikace

$$6) G = \sum_{i=1}^K \frac{N_i}{N} G(i)$$

7) - projevování na základě klasifikace chyb
- projevování na základě odchylek
- přímé rozhodnutí růstu stromu

8) Získá se minimum porovnání, kde se obno
tvoří podle jedné hodnoty predikce.
Po dosažení minima se obno opět rozšiřuje a pomocí
projevování se získá optimální velikost obno.
Obno se, dokud není dosaženo konečného počtu obno.

1) Učím jednoduchých rozhodovacích stromů, které se
podílejí na klasování přírodních kategorií (klasifikace), nebo
se jejich výstupy využívají.

11) Generování více modelů a jejich složení na základě
vše, tak aby z nejlepšího pohledu vznikl 1 model.

12) Je třeba množením se hodnoty v přírodních poměrech
(např. 80% - živost, 20% - lesnost) podle přírodních měřítek.

13) Uvažujeme se dosahem přírodních čísel tříd přírodních
vše včetně. Uvažujeme se pak načasování a čas v měsích
délce a dosahem přírodních výsledků - jaké na klasování množství.