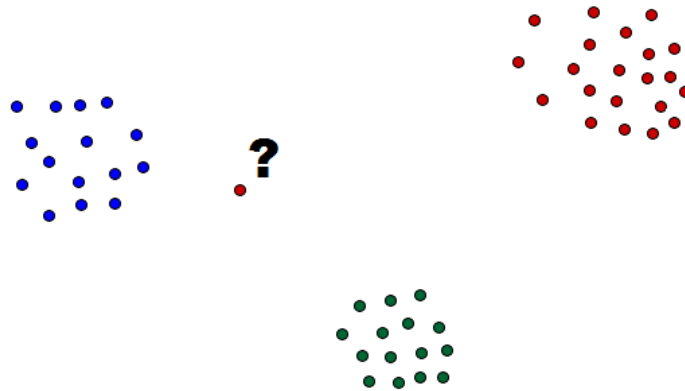


INTERPRETACE ODLEHLÝCH BODŮ



LEONA NEZVALOVÁ

JAK INTERPRETOVAT ODLEHLÉ BODY

DVĚ JEDNODIMENZIONÁLNÍ INTERPRETACE

instance	kladení vajec	zuby	třída
lev	ne	ano	savec
antilopa	ne	ano	savec
delfín	ne	ano	savec
sviňucha	ne	ano	savec
ptakopysk	ano	ne	savec
včela	ano	ne	hmyz
moucha	ano	ne	hmyz
vosa	ano	ne	hmyz

JEDNA DVOUDIMENZIONÁLNÍ INTERPRETACE

zvíře	srst	žije ve vodě	třída
lev	ano	ne	savec
antilopa	ano	ne	savec
delfín	ne	ano	savec
sviňucha	ne	ano	savec
sphinx	ne	ne	savec
včela	ne	ne	hmyz
moucha	ne	ne	hmyz
vosa	ne	ne	hmyz

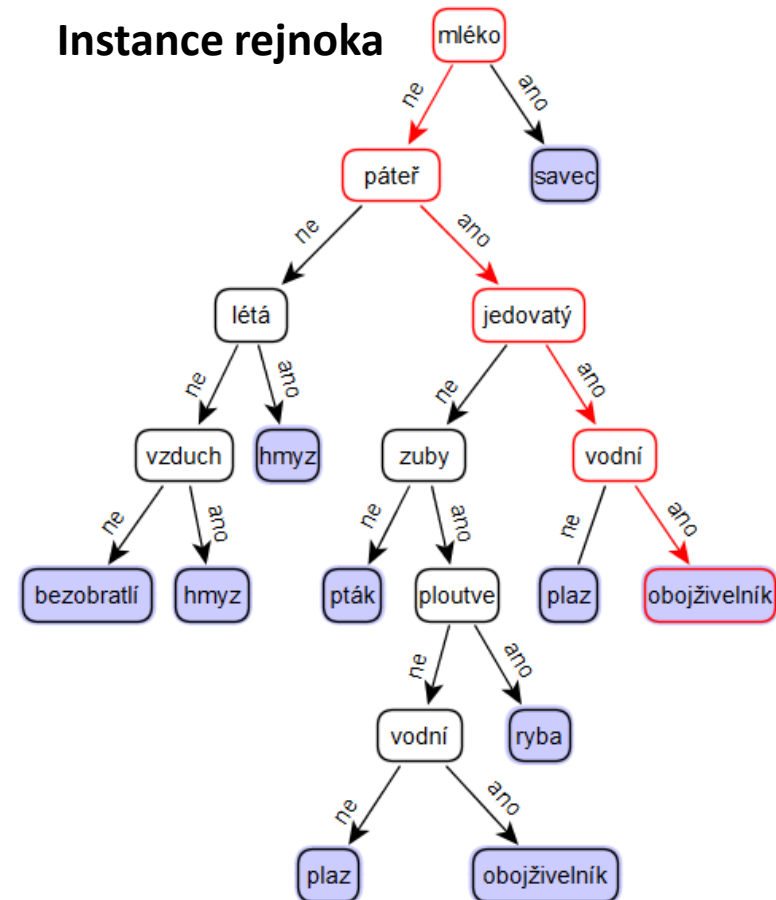
RF-OEX (RANDOM FOREST-OUTLIER EXPLANATION)

- Rozšiřuje systém weka-peka¹ => pracuje s klasifikovanými daty
- 2 metody interpretace
 - metoda založená na prořezávání náhodných stromů
 - metoda založená na častých vzorech
- Výsledná interpretace ve tvaru disjunktivní normální formy
- Každé konjunkci atributů interpretující odlehlost bodu je přiřazena váha, s jakou tyto atributy přispívají do celkové interpretace

RF-OEX

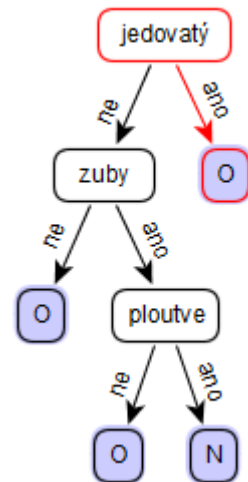
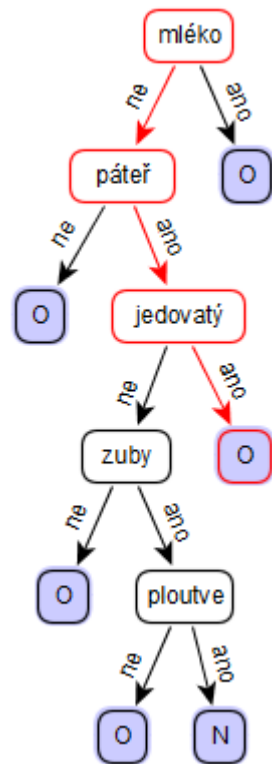
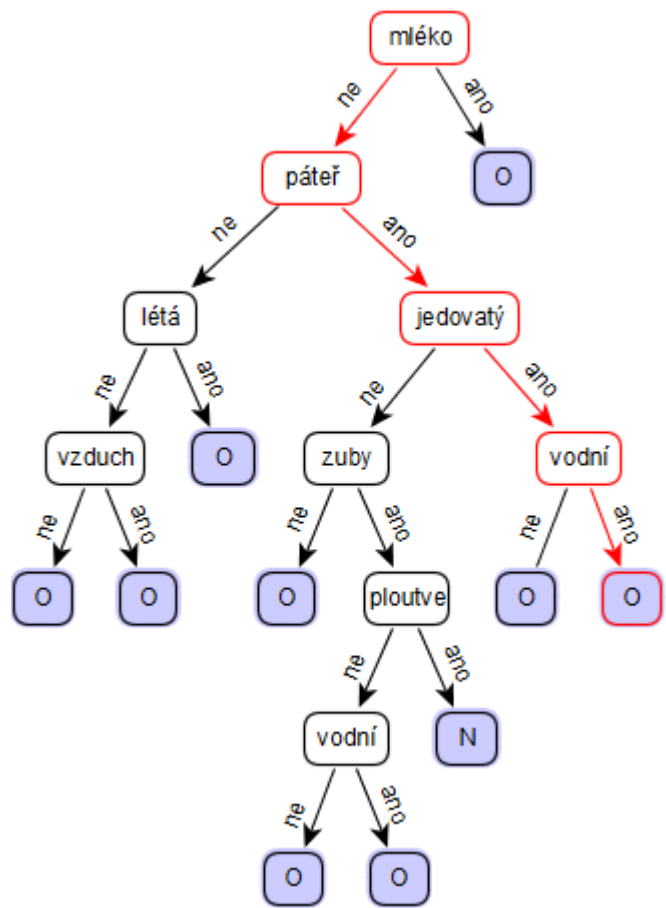
- Metody pracují s množinou těchto stromů, které klasifikovaly zkoumaný bod do **jiné třídy, než kterou je bod popsán**
- Každý takový strom musí obsahovat alespoň v jednom uzlu atribut, který dělá zkoumaný bod více **podobný instancím z jiné třídy**
- **Otázka: které uzly to jsou?**

Instance rejnoka



INTERPRETACE PROŘEZÁVÁNÍM NÁHODNÝCH STROMŮ

1. Nahrazení původních tříd dvěma novými
 - **N** označuje třídu, kterou je bod popsán
 - **O** označuje všechny třídy ostatní
2. Odstranění irelevantních podstromů
3. Zkoumání vlivu na klasifikaci při postupné změně atributů v uzlech, kterými klasifikace prochází



INTERPRETACE PROŘEZÁVÁNÍM NÁHODNÝCH STROMŮ

- Klasifikací zkoumaného bodu na upraveném stromě získáme množinu atributů, jejich operátorů a hodnot
- Konjunkce prvků této množiny je součástí interpretace zkoumaného bodu
- Váha, s jakou konjunkce do celkové interpretace přispívá, je dána četností jejího výskytu ve všech upravených stromech

NUMERICKÉ ATRIBUTY

- Zatímco nominální atribut je vždy s hodnotou svázán operátorem '=', hodnota takového atributu v interpretaci nabývá vždy hodnoty, která je dána zkoumanou instancí
- Dělení uzlu s numerickým atributem **att** a hodnotou **y** generuje vždy dvě větve :
 - větev pro hodnoty atributu **att < y**
 - Větev pro hodnoty atributu **att ≥ y**
- Hodnota numerického atributu v interpretační konjunkci je tvořena z průměru hodnot tohoto atributu, které se vyskytovali ve stromech obsahující danou interpretaci a daný operátor

INTERPRETACE NA ZÁKLADĚ HLEDÁNÍ ČASTÝCH VZORŮ

- Pokud by stromy byly generovány náhodně, nalezené časté konjunkce atributů ve stromech bychom mohli považovat za výslednou interpretaci

- Při stavbě stromů v systému weka-peka je však zohledněn informační zisk atributu, čím vyšší informační zisk, tím pravděpodobněji bude atribut do uzlu stromu vybrán



- Ve stromech, se kterými metody RF-OEX pracují, se opakovaně vyskytují ty atributy, které mají vysoký informační zisk

- **Otázka: Které nalezené konjunkce atributů jsou opravdu součástí interpretace?**

INTERPRETACE NA ZÁKLADĚ HLEDÁNÍ ČASTÝCH VZORŮ

- Relevantní konjunkce atributů zjistíme znovuvyhodnocením faktoru odlehlosti na instancích popsaných pouze zkoumanými atributy
- Výsledky ze všech častých vzorů standardizujeme a porovnáme
- Faktor odlehlosti zjišťovaný na libovolné množině atributů je zpravidla větší nebo roven faktoru odlehlosti zjišťovanému na libovolné menší podmnožině těchto atributů
 - Příklad: Faktor odlehlosti instance rejnoka na atributu *jed* je 7,21. Na attributech *jed* a *páteř* je 7,47.
- Váha interpretace je měřena jako rozdíl faktoru odlehlosti na dané množině atributů a faktoru odlehlosti na takové podmnožině těchto atributů, která dosahuje maximálního faktoru odlehlosti ze všech podmnožin menších o jeden prvek.
 - Příklad – pokračování: doplňme že faktor odlehlosti rejnoka na atributu *páteř* je 1,04. Protože $7,21 > 1,04$, váha interpretace (*jed* \wedge *páteř*) je $7,47 - 7,21 = 0,26$. Můžeme říci že rejnoka interpretuje samostatný atribut *jed*. Kombinace (*jed* \wedge *páteř*) má pro interpretaci minimální význam.

VÝSLEDKY - ZOO

instance	metoda proř. stromů	metoda častých vzorů	ILP ²
ptakopysk (savec)	vejce = ano (0,59) zuby = ne (0,41)	vejce = ano (0,46) zuby = ne (0,54)	vejce = ano
rejnok (ryba)	jedovatý = ano (0,97)	jedovatý = ano (0,66)	-----
škorpión (bezobratlý)	vzduch = ano (0,29)	vzduch = ano (0,27) ocas = ano (0,32)	-----
želva (plaz)	zuby = ne (0,79)	zuby = ne (0,22) páteř = ano (0,23)	-----
mořský had (plaz)	vodní = ano (0,38) vzduch = ne (0,35)	vodní = ano (0,36)	-----
slimák (bezobratlý)	predátor = ne (0,4) vzduch = ano (0,33)	predátor = ne (0,27) vzduch = ano (0,39)	-----
...			
čolek (obojživelník)	ocas = ano (0,99)	ocas = ano (0,21)	ocas = ano
slunéčko sedmítečné (hmyz)	predátor = ano (0,62) srst = ne (0,28)	predátor = true (0,58)	predátor = ano

VÝSLEDKY – REZOLUČNÍ STROMY

instance	metoda proř. stromů	metoda častých vzorů	metoda mělkých stromů ³
270	p1 = ne (0,88)	p1 = ne (1)	p1 = ne (0,95)
396	----- -----	----- -----	----- -----
389	p8 = ne \wedge p16 = ne \wedge p17 = ne (0,26) p8 = ne (0,24) p8 = ne \wedge p17 = ne (0,2)	p8 = ne (0,39)	p1 = ne (1)
187	p14 = ano (0,91)	p14 = ano (0,54) p14 = ano \wedge p18 = ne \wedge p17 = ne (0,24) p14 = ano \wedge p18 = ne \wedge p16 = ne (0,22)	p14 = ano (1)
236	p14 = ano (0,91)	p14 = ano (0,54) p14 = ano \wedge p18 = ne \wedge p17 = ne (0,24) p14 = ano \wedge p18 = ne \wedge p16 = ne (0,22)	p14 = ano (1)
438	p8 = ne \wedge p17 = ne (0,31) p1 = ano \wedge p8 = ne \wedge p17 = ne (0,2)	p8 = ne (0,48)	p16 = ano (1)
82	p2 = ano (0,76)	p2 = ano (1)	

RF-OEX: SROVNÁNÍ

INTERPRETACE PROŘEZÁVÁNÍM NÁHODNÝCH STROMŮ

- + rychlost
- + minimálnost
- + podrobnější informace o numerických atributech
- interpretace nemusí být úplná

INTERPRETACE NA ZÁKLADĚ HLEDÁNÍ ČASTÝCH VZORŮ

- + úplnost
- časová náročnost
- interpretace může obsahovat irelevantní atributy