

Datum: _____

Jméno: _____

1. (2 body) Necht' jsou data popsána reálnými příznaky x a y . Zaved'me další příznak z , definovaný jako $z = 2x + y$.

Jsou-li data lineárně separabilní v prostoru příznaků x, y , jsou lineárně separabilní i v prostoru příznaků x, y, z .

Jsou-li data lineárně separabilní v prostoru příznaků x, y, z , jsou lineárně separabilní i v prostoru příznaků x, y .

2. (4 body) Pro rozhodovací strom platí:

Klasifikace probíhá procházením uzlů od kořene až k listu. Výsledná třída je vypočtena z majoritního počtu trénovacích instancí, které padnou do příčného listu.

Existuje nekonečně mnoho stromů, které jsou z hlediska klasifikace ekvivalentní.

Jeho hloubka není přímo závislá na počtu vstupních atributů.

Počet listů je roven počtu tříd.

3. (3 body) Jaký je hlavní rozdíl mezi regresí a klasifikací?

4. (3 body) Provádíme-li rozšíření báze,

vybíráme podmnožinu vstupních proměnných, které mají na závislou veličinu největší vliv.

odvozujeme nové proměnné, přičemž doufáme, že alespoň některé z nich budou pro predikci vhodnější než původní proměnné.

zvyšujeme dimenzi vstupního prostoru a tedy i riziko přeučení modelu.

5. (3 body) Selektce proměnných typu *filter*:

Existuje 1 metoda filtrování proměnných, která dává nejlepší výsledky.

Často se využívá řazení proměnných podle jejich souvislosti se závislou proměnnou.

Dá se zařadit mezi metody předzpracování dat, protože není závislá na použitém (klasifikačním nebo regresním) modelu.

6. (2 body) Proveďte predikci spojitě závislé veličiny y bodu $x=5.2$ metodou 3 nejbližších sousedů. Trénovací *multimnožina* $\{(x,y)\}$ je:

$\{(2, 1.0), (3, 5.0), (3, 4.0), (4, 4.0), (5, 1.0), (6, 1.0), (6, 2.0), (7, 6.0)\}$

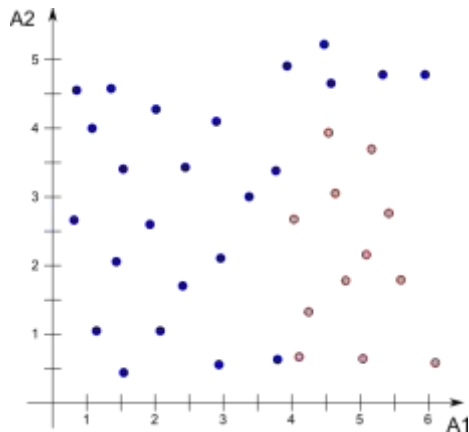
1

4/3

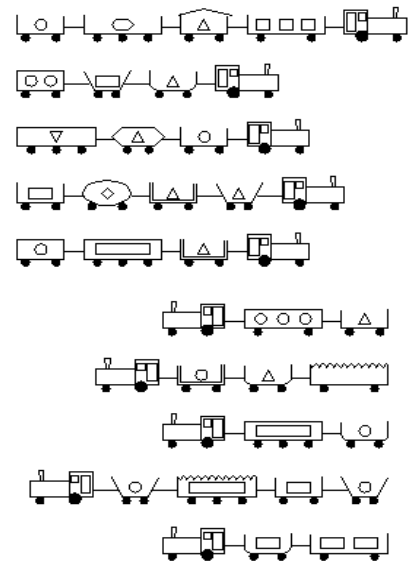
2

7/3

7. (4 body) Nakreslete rozhodovací strom, který data na obrázku bezchybně rozdělí. Do obrázku zakreslete rozdělení prostoru pro jednotlivé uzly stromu. Vysvětlete, proč jste jako první pro dělení vybrali proměnnou A1 nebo A2.



8. (3 body) Uvažujte 10 instancí dat na obrázku a asociační pravidlo:
vlak má tři vagóny => vlak jede doprava.
 Spočítejte podporu a spolehlivost tohoto pravidla



9. (4 body) Vysvětlete pojmy „true positives“ a „false negatives“.

10. (2 body) O boostingu platí:

- Vytvoří více slabých modelů, jejichž zkombinováním získá model přesnější.
- Jednotlivé modely na sobě nejsou závislé, může je vytvářet paralelně, prostě tak, že každému “předhodí” jinou část trénovacích dat.
- Jednotlivé modely jsou na sobě závislé, vytvářejí se postupně.
- Všechny trénovací případy mají po celou dobu učení stejnou váhu.