

Jméno, příjmení:

E-mail:

## Zkouška z předmětu Strojové učení – 12. 1. 2005

Své odpovědi vždy stručně zdůvodněte, aby bylo zřejmé, proč tak odpovídáte. Maximální čas k dispozici: 120 minut. Stupnice hodnocení:  $A \geq 90\%$ ,  $90\% > B \geq 80\%$ ,  $80\% > C \geq 70\%$ ,  $70\% > D \geq 60\%$ ,  $60\% > E \geq 50\%$ ,  $50\% > F$ .

1. V tabulce jsou dány tréninkové příklady. Cílovým atributem je *HrátTennis* jehož možné hodnoty jsou *ano* nebo *ne*. Indukujte pomocí metody ID3 rozhodovací strom, který bude schopný predikovat hodnotu cílového atributu na základě hodnot ostatních atributů. Indukci stromu dokumentujte výpočtem. Kam indukovaný strom zařadí příklad: *zataženo, chladno, vysoká, slabý*? **Max. 17%.**

<i>den</i>	<i>předpověď počasí</i>	<i>teplota</i>	<i>vlhkost</i>	<i>vitr</i>	<i>HrátTennis</i>
D1	slunečno	teplo	vysoká	slabý	<b>ne</b>
D2	slunečno	teplo	vysoká	silný	<b>ne</b>
D3	zataženo	teplo	vysoká	slabý	<b>ano</b>
D4	deštivo	mírně	vysoká	slabý	<b>ano</b>
D5	deštivo	chladno	normální	slabý	<b>ano</b>
D6	deštivo	chladno	normální	silný	<b>ne</b>
D7	zataženo	chladno	normální	silný	<b>ano</b>
D8	slunečno	mírně	vysoká	slabý	<b>ne</b>
D9	slunečno	chladno	normální	slabý	<b>ano</b>
D10	deštivo	mírně	normální	slabý	<b>ano</b>
D11	slunečno	mírně	normální	silný	<b>ano</b>
D12	zataženo	mírně	vysoká	silný	<b>ano</b>
D13	zataženo	teplo	normální	slabý	<b>ano</b>
D14	deštivo	mírně	vysoká	silný	<b>ne</b>

2. Vysvětlete stručně, co je a) otevřená doména, b) uzavřená doména, c) zašuměná data (jaké typy šumu mohou existovat?). Uveďte k a), b) a c) příklad odlišný od příkladu zmíněného na přednášce. **Max. 4%.**

3. Jak by algoritmus *c5/See5* řešil případ, když by v tabulce z otázky č. 1 v některých řádcích chyběly hodnoty některých atributů? Vysvětlete, na čem je řešení používané v *c5/See5* založeno. Pokud by klasifikátor *c5/See5* vygenerovaný z tréninkových dat v otázce č. 1 měl klasifikovat např. instanci *zataženo, chladno, ?, slabý*, jak by v principu klasifikace mohla dopadnout? (Není nutno provést výpočet, je zapotřebí ukázat principiálně, jak by daný případ mohl být řešen.) **Max. 18%.**

4. Čím je charakterizováno učení a) on-line, b) of-line, c) inkrementální, d) dávkové. Je možné mezi sebou kombinovat typy a) a b) s typy c) a d)? Na čem závisí možné kombinace typů učení? Uveďte příklady. **Max. 5%.**

5. Co je to AdaBoost, na čem je metoda založena, jak funguje, proč a kdy je vhodná k použití? Za jakých okolností např. z hlediska vlastností dat by AdaBoost nemusela pozitivně působit na řešení? **Max. 18%.**

6. V čem se u genetických algoritmů liší standardní a seřazovací metoda, jaký je přínos seřazovací metody oproti metodě standardní? Na čem jsou obě metody založeny? Může standardní metoda ovlivnit možnou degeneraci pozitivně nebo negativně, a proč? Jak to ovlivní seřazovací metoda? Jak lze principiálně u genetických algoritmů omezit vliv degenerace? **Max. 17%.**

7. Na čem je založen Gibbsův algoritmus, jaké má vlastnosti, výhody a nevýhody? Kdy byste ho použili a proč? **Max. 6%.**

8. Na čem je založena metoda *IB3* ze skupiny metod *IBL*? Jaký přínos mají její specifické vlastnosti oproti *IB1* a *IB2*? Co je typické pro *IB1* a *IB2*? K čemu a proč se specifické vlastnosti *IB3* hodí? **Max. 15%.**