

## VIKMB03 *Strojové učení jaro–2007* Přehled učiva ke kolokviu/zápočtu

Je zapotřebí znát principy a významy níže uvedených pojmů, jejich aplikovatelnost na automatizované získávání znalostí, výhody a nevýhody různých probraných typů algoritmů, jak se algoritmy učí a testují. Nevyžaduje se matematický základ probrané látky. Otázky kolokvia jsou založeny výhradně na látce probrané na přednáškách. Výukové materiály (soubory \*.PDF) poskytnuté prostřednictvím IS-MU obsahují řadu kapitol a částí, které v předmětu na přednáškách probírány nebyly a nebudou zkoušeny (zejména matematické popisy a dále detaily na přednáškách nezmiňované nebo věci, na něž bylo na přednáškách upozorněno). Některé přednášky byly doplněny demonstračními ukázkami, které nejsou v poskytnutých \*.PDF materiálech zahrnuty (např. systém WEKA, apod.).

Kolokvium/zápočet je formou písemných odpovědí na zadané otázky. Pro splnění kolokvia je nutno získat minimálně 67 %, pro zápočet minimálně 57 % ze správných odpovědí.

Otázky mají stejnou váhu vzhledem k hodnocení: 1 otázka = max. 20 %, celkem je 5 otázek. Zcela nezodpovězena (0 %) může být nanejvýš 1 otázka. Maximální čas na vypracování odpovědí je 60 minut. Odpovědi mají být stručné a musí vystihovat podstatu včetně vysvětlení.

Příklad možné otázky v písemce: *Jakými způsoby (a proč zrovna tak) lze reprezentovat textové dokumenty pro kategorizaci strojovým učení, v čem se projeví volba různé reprezentace, jak se uvedenými reprezentacemi natrénuje a otestuje metoda naivního Bayese, a jakou formu zde má trénováním získaná znalost?*

---

### **Přehled témat (probráno na přednáškách – obsaženo v \*.PDF v IS-MU/Studijní materiály):**

#### **1. Strojové učení**

indukce, dedukce, abdukce; generalisace; data, informace, znalost, vzájemný hierarchický vztah; klasifikace; typy reprezentace znalostí, výhody a nevýhody; reprezentace dat, možnost vzájemného převodu; trénování a testování; Occamova (Ockhamova) břitva; relevantní a nerelevantní atributy trénovacích příkladů; množství trénovacích příkladů; přeučení a generalisace; vliv různých vlastností a parametrů na učení

#### **2. Rozhodovací stromy**

entropie, generování stromů z příkladů, převod stromů na pravidla

#### **3. Metoda nejbližšího souseda**

podobnost; vzdálenost a její stanovení; porovnávání; 1-NN,  $k$ -NN; princip učení a klasifikace

#### **4. Evoluční výpočty, genetické algoritmy**

simulace darwinovské evoluce; populace, generace, křížení, mutace, diversita, funkce přizpůsobení; kódování a dekódování genů a chromosomů; optimalisace; výběr jedinců pro vytvoření další generace

#### **5. Naivní Bayes, bayesovská kategorizace textů**

apriorní a aposteriorní pravděpodobnosti; hypotézy; podmíněné pravděpodobnosti; výhoda a nevýhoda "naivity"; princip aplikace na klasifikaci a filtraci textových dokumentů, přednosti a nedostatky algoritmu

#### **6. Perceptron, umělé neuronové sítě**

princip umělého lineárního perceptronu, váhy, vstupy, výstup, trénování; reprezentace binárních logických a booleovských funkcí, vliv linearity přenosové funkce na nemožnost reprezentace určitých funkcí (XOR); důvod spojování lineárních perceptronů do sítě; sigmoida, nelineární perceptron, důvody jeho zavedení; spojování nelineárních perceptronů do sítě, architektura dopředné sítě (vrstvy); vlastnosti sítí z nelineárních jednotek; princip trénování (vliv chyby na výstupu); aplikační možnosti (reprezentace funkcí, klasifikace)