

Pokyny k vypracovaniu projektov
2-INF-150: Strojové učenie, Zima 2009
Vyučujúci: Tomáš Vinar

Základnou úlohou v rámci projektu je:

- vybrať si nejaký problém, na ktorý sa dajú aplikovať metódy strojového učenia (môže byť takmer čokoľvek, napr. aj niečo čo potrebujete urobiť do diplomovej práce)
- naimplementovať jednu alebo viacero metód alebo variantov riešenia problému
- otestovať implementáciu na skutočných alebo simulovaných dátach
- vyhodnotiť úspešnosť zameraním výkonnosti na testovacích dátach a porovnaním buď s inou voľne dostupnou implementáciou, triviálnym riešením nevyužívajúcim pokročilé metódy (napr. ľahko implementovateľnú heuristiku), alebo porovnať viacero variant medzi sebou (napr. rôzne kernely pre SVM, rôzne rozvoje báz a pod.)

Návrh projektu je max. jedna strana textu, ktorá obsahuje popis problému, ktorý sa chystáte riešiť, dáta na ktorých sa chystáte svoj projekt trénovať a testovať a popis spôsobu vyhodnocovania experimentov. Nie je potrebné do návrhu písať, aké metódy sa chystáte použiť na riešenie projektu. Návrh projektu je potrebné odovzdať do **30.11.2009**.

Samotný **projekt** treba odovzdať najneskôr **72 pracovných hodín pred konaním skúšky**. V projekte sa hodnotí iba to, čo je **napísané a odovzdané**. Odovzdávajú sa dve časti: **report o projekte** v papierovej forme (pod dvere M163) a **zdrojové texty a dáta** elektronicky ako jeden .zip súbor e-mailom na adresu vinar@fmph.uniba.sk

Požiadavky na report: Report má byť v rozsahu 5-15 strán a je hlavným podkladom pre hodnotenie projektu. Report musí obsahovať najmenej:

- Nadpis a meno autora projektu
- Abstrakt (cca 1-2 odstavce) sumarizujúci cieľ, použité metódy a výsledky projektu
- Úvod do problematiky: popis úlohy, odkiaľ ste získali dáta a ich stručnú charakteristiku (napr. veľkosť trénovacej / testovacej množiny, koľko pozitívnych / negatívnych príkladov, priemernú veľkosť inštancie, popis atribútov, priemery a štandardné odchýlky zaujímavých atribútov)
- Popis a zdôvodnenie použitých metód
- Stručný popis technických detailov implementácie (v akom jazyku, aké knižnice ste použili a prečo, s akými technickými problémami ste sa stretli pri riešení)
- Experimentálnu evaluáciu: akým spôsobom vyhodnocujete úspešnosť? Porovnanie s niečím iným, tabuľky, grafy, ich interpretácia a závery.
- Záver: myslíte si, že bol projekt úspešný alebo neúspešný? Prečo? Keby ste projekt robili znovu, čo by ste urobili ináč? Aké iné problémy by sa dali riešiť podobným spôsobom? **Aj dobre spracovaný neúspešný projekt môže získať plný počet bodov**, ak obsahuje dobrú analýzu príčin neúspechu.

Inšpirovať sa môžete aj na nasledujúcej stránke (aj keď nejde o presne rovnaké zadanie):
<http://www.stanford.edu/class/cs229/projects2006.html>

Požiadavky na elektronickú časť:

- Komentované zdrojové texty vlastných programov a skriptov
- Dáta s ktorými ste pracovali (ak ide o verejne prístupné dáta z internetu, alebo sú dáta veľké a výsledný .zip súbor by presiahol 5MB, môžete napísať linku do súboru readme.txt)
- Súbor readme.txt popisujúci jednotlivé súbory, ako ich skompilovať a spustiť

Orientačné hodnotenie:

- Výber problému: 5%
- Výber a správne použitie metód: 30%
- Implementácia: 20%
- Dizajn experimentu: 25%
- Report: 20%

Kvalita reportu vplýva aj na ostatné časti. Ak napríklad z reportu nemožno pochopiť, o čo sa vlastne snažíte, tak to značne zníži známku aj v ostatných častiach.

Opisovanie. Tak ako vo zvyšku kurzu, opisovanie (od spolužiakov, alebo z internetu) bud prísne postihnuté – vid'. formálne upozornenie na stránke predmetu. Keďže ide o veľkú časť známky, v prípade zistenia opisovania nemožno rátať s úspešným ukončením predmetu.