

SU4SS: web o strojovém učení pro středoškoláky

V rámci projektu SU4SS vytvoříme webové stránky určené především studentům středních škol, které budou srozumitelné a zajímavé (ale přesto věcně správně a do hloubky) seznamovat čtenáře se strojovým učením.

Jsme přesvědčeni, že podobný informační zdroj v češtině dosud neexistuje, což způsobuje nízké povědomí o strojovém učení a nízký zájem o něj. To pak má za následek nedostatek odborníků na strojové učení, který na Matfyzu pocítujeme jak při svém studiu, tak při samostatné výzkumné činnosti. Založením webu chceme studenty středních škol motivovat k zájmu o strojové učení a ke studiu Matfyzu.

Domníváme se, že drtivá většina studentů středních škol si pod studiem informatiky představí zejména programování (pokud si tedy vůbec něco představí). To je škoda, protože informatika toho zahrnuje mnohem více, a právě strojové učení považujeme za oblast velice zajímavou a zároveň užitečnou. Částečně neprávem je strojové učení považované za něco velice složitého a nepřístupného, a běžně se s ním nesetká dokonce ani bakalářský student Matfyzu; význam tohoto oboru přitom ukazuje fakt, že figuruje v názvech dvou magisterských studijních zaměření.

Přitom například *naivní Bayesův klasifikátor* je myšlenkově triviální, pro pochopení vyžaduje pouze znalost podmíněné pravděpodobnosti, která je běžnou součástí středoškolských osnov, a lze jej snadno implementovat i v tabulkovém procesoru, nebo dokonce manuálně s pomocí papíru a tužky. Strojové učení umožňuje řešení mnoha praktických úloh, které jsou jinými způsoby jen obtížně řešitelné. Studium strojového učení zároveň umožňuje porozumět technologiím, které se stávají běžnou součástí našeho života, jako je například rozpoznávání řeči, spam-filtry, řazení zobrazených příspěvků na Facebooku, ...

V období **duben až srpen 2015** plánujeme na webových stránkách, které založíme, důkladně probrat tři algoritmy pro klasifikační úlohy strojového učení: *perceptron*, *logistickou regresi* a *naivního Bayese* (případně i další algoritmy, pokud nám to dovolí naše časové možnosti – zejména *Support vector machines*, *rozhodovací stromy* a *neuronové sítě*). Tato témata vyžadují určité předchozí znalosti a zároveň logicky vedou na další úzce související problémy, předběžná navrhovaná struktura je tedy následující:

- ukázat a vysvětlit, co je strojové učení
- *perceptron*, *množina rysů*, geometrická intuice (*separační nadrovina*)
- úvod do statistiky a pravděpodobnosti
- *logistická regrese*
- *naivní Bayes*
- *generativní* versus *diskriminativní* model
- *přeučení*, *normalizace* hodnot rysů
- případná pokročilejší témata (*SVM*, *kernely*, *DT*, *NN*, *testování hypotéz*)

Jednotlivé texty vždy budou zahrnovat přístupně podané informace nutné pro základní porozumění tématu, a dále rozšiřující podrobnější informace (odvození, důkazy, pokročilejší varianty apod.). Součástí budou i praktické úlohy, které čtenář bude moci samostatně vyřešit. Web tak poskytne všem studentům základní seznámení se strojovým učení, a zároveň zvědavějším studentům umožní hlubší pochopení problematiky.

Na projekt žádáme částku **5000 Kč** na stipendia pro účastníky projektu (Mgr. Rudolf Rosa, Mgr. Jindřich Libovický, Bc. Tomáš Musil), kterou rozdělíme v závislosti na množství vytvořených materiálů.