

Strojové učení pro středoškoláky – Závěrečná zpráva

Jindřich Libovický, Tomáš Musil, Rudolf Rosa

14. prosince 2015

V rámci projektu jsme připravili pětítýdenní seminář o strojovém učení pro studenty středních škol. Cílem projektu bylo popularizovat mezi středoškoláky strojové učení a tím ukázat, že studium informatiky nemusí být jen o programování složitých systémů, ale že existují i další zajímavá témata, kterými se studenti na MFF UK mohou zabývat.

Na strojovém učení jsou založeny technologie, se kterými se všichni dennodenně setkáváme. Jsou mezi nimi jak ty, které mají stále trochu futuristický nádech, jako je automatický překlad nebo rozpoznávání mluvené řeči, tak i technologie, které už zcela samozřejmě přijímáme, jako je rozpoznávání obličejů na fotkách na Facebooku, doporučování videí na YouTube, vyhledávání obrázků na Googlu nebo detekce spamu.

Přednášky jsme začali připravovat v létě 2015 i v souvislosti se zamýšleným popularizačním webem o strojovém učení. Přípravy se účastnili: *Jindřich Libovický* (3. ročník doktorského studia, Matematická lingvistika), *Tomáš Musil* (2. ročník navazujícího magisterského studia, Umělá inteligence) a *Rudolf Rosa* (3. ročník doktorského studia, Matematická lingvistika). Rudolf Rosa je od září 2015 na šestiměsíční stáži ve společnosti Google a dalších příprav se už nezúčastnil, vystoupil formou videohovoru na druhé přednášce. Přípravu kursu jsme konzultovali s *Ondřejem Bojarem* z Ústavu formální a aplikované lingvistiky MFF UK.

Seminář proběhl v listopadu 2015 a účastnilo se ho 9 středoškoláků, z nichž 5 přišlo více než jednou.

1 Propagace akce

S propagací semináře jsme začali na konci září 2015. Na webu ÚFALu jsme vytvořili jednoduchou webovou stránku se základními informacemi o semináři. Potenciální posluchače jsme oslovili následujícími kanály:

- rozdávání letáček na akci *Jeden den s informatikou*
- oslovením přátel a známých, kteří pracují na gymnáziích (o to jsme požádali i kolegy z ÚFALu prostřednictvím mailové konference) a studentů učitelství informatiky na PedF (prostřednictvím členů jejich studentského spolku)
- krátkou zprávou na webu OVVP
- oficiální emailovou komunikací OVVP
- krátkou zprávou na webu Matfyz.cz
- krátkou zprávou na webu ÚFALu
- sdílením na osobním Facebookovém profilu
- sdílením na Facebookovém profilu ÚFALu

Celkem se přihlásilo 22 potenciálních účastníků, z toho 16 středoškoláků (14 z pražských gymnázií, 2 mimo Prahu), 3 studenti MFF UK a 3 další lidé.

Od přihlášených jsme v krátkém dotazníku zjistili jejich časové preference, do jakého ročníku chodí a úroveň jejich programátorských dovedností. Na základě sesbíraných odpovědí jsme seminář pořádali vždy v pátek od 16:00 do 17:30 v učebně S9.



Obrázek 1: Foto z přednášky 13. 11., kde prostřednictvím videokonference promluvil o strojovém učení v Googlu Rudolf Rosa, který tam byl tou dobou na stáži.

2 Průběh semináře

V přednáškách jsme studenty seznámili se základními metodami strojového učení a předvedli některé aplikace:

1. **přednáška** – co je strojové učení, metoda nejbližších sousedů (data: klasifikace druhů kosatců), lineární regrese
2. **přednáška** – strojové učení v Googlu (videohovor s R. Rosou), klastrování metodou K-means
3. **přednáška** – úvod do pravděpodobnosti, naivní Bayesovský klasifikátor, perceptron (data: klasifikace emailů)
4. **přednáška** – rozhodovací stromy a lesy (data: přeživší z Titaniku), deep learning (použití neuronových sítí pro NLP a strojové vidění)
5. **přednáška** – rekapitulace metod na příkladu odhadování recidivy rakoviny prsu, diskuse o budoucnosti strojového učení

V průběhu přednášek jsme promítali interaktivní vizualizace probíhajícího učení. Zdrojový kód bylo možné průběžně upravovat a ukazovat důsledky jeho změn na výsledcích učení. Ukázalo se však, že je obtížné najít data, která by byla pro tento způsob prezentace vhodná. Bylo by totiž potřeba, aby byla dostatečně přehledná na to, aby se na nich dala daná metoda strojového učení dobře vysvětlit a současně dost komplexní, aby se na nich dalo ukazovat různé chování při různých nastaveních učicího algoritmu. Proto jsme v pozdějších přednáškách přešli k vysvětlování na jednoduchých příkladech u tabule. Při tom se nám dařilo studenty lépe zapojit a nechat je vymýšlet, jak by algoritmy mohly fungovat a nacházet protipříklady k vlastním nápadům a nápadům spolužáků. Z tohoto pohledu se jsou velice názorné rozhodovací stromy a klastrování.

V průběhu kursu jsme nakonec došli k závěru, že nejvhodnější je rozdělit přednášku na dvě části. První část má charakter semináře, ve kterém vykládáme některou metodu strojového učení, příklady kreslíme na tabuli a snažíme se studenty co nejvíce zapojit. V druhé části necháme studenty odpočinout a promítáme jim prezentaci, ve které ukazujeme reálné použití vyložených metod a snažíme se probranou látku zasadit do širšího kontextu.

Materiály, které jsme studentům v průběhu přednášek promítali, jsou dostupné na adrese <https://github.com/tomasmcz/su4ss>.

	6. 11.	13. 11.	20. 11.	27. 11.	4. 12.
středoškoláci	7	6	2	3	2
ostatní	4	3	1	1	0

Tabulka 1: Počty účastníků na jednotlivých přednáškách

Účast na semináři postupně klesala (viz tabulka 1). Pokles účasti si vysvětlujeme tím, že část studentů strojové učení nezaujalo (ti nejspíš přišli pouze na první přednášku). K našemu překvapení byli ale také studenti, byli při přednáškách aktivní, vypadali, že je obsah zajímá, dokonce se vyjadřovali k plánu na další přednášky, a přesto už příště nepřišli. Za překvapivé považujeme také to, že velká část přihlášených nepřišla ani jednou (viz tabulka 2), a to přesto, že kromě jednoho všichni v dotazníku vyplnili, že jim páteční termín vyhovuje. O důvodech můžeme pochopitelně jen spekulovat, je ovšem možné že to, že to, že se seminář konal v pátek v podvečer a pro studenty v cizím prostředí, kam museli dojíždět, vedlo k tomu, že se jim nakonec nechtělo a dali na páteční večer přednost jinému programu.

četnost návštěv	ani jednou	jednou	dvakrát	třikrát	čtyřikrát	pětkrát
počet studentů	7	4	2	1	1	1

Tabulka 2: Histogram četnosti návštěv studentů (tedy kolik přihlášených studentů přišlo na kolik přednášek)

3 Závěr

V souladu s deklarovanými cíli projektu jsme studenty seznámili s tím, co je to strojové učení, jaké je jeho využití v praxi a proč by se něčím takovým mohli chtít zabývat.

Studenti, kteří přišli, víceméně věděli, že chtějí studovat informatiku, ale kromě MFF UK zvažovali i jiné školy. Po přednáškách několikrát padaly dotazy ohledně studia informatiky na MFF UK a jak se liší do studia informatiky na jiných vysokých školách.

Pro nás samotné bylo prospěšné zjistit, co je na středoškolské úrovni přístupné. Ověřili jsme, jak studenti intuitivně chápou objekty z lineární algebry a pravděpodobnosti (aniž bychom je hlouběji vysvětlovali). Zjistili jsme, které metody výkladu fungují lépe a které hůře. To vše považujeme za cenné zkušenosti, které bychom rádi dále využili při popularizaci strojového učení.