

Metódy strojového učenia a ich praktické použitie*

Martin Orlej

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

11. október 2021

Abstrakt

Strojové učenie sa v posledných rokoch čoraz viac spomína či už vo vedeckých prácach, alebo v rôznych článkoch, zameraných na technológie. Či už ide o niečo jednoduché, ako aplikácie na telefóny, alebo o vysoko pokročilé technológie ako autonómne jazdenie a počítačové videnie. V mojej práci by som sa chcel zamerať na rôzne modely strojového učenia, ako napr. lineárna regresia, Boltzmannove stroje alebo transformátory, na ich praktické a najefektívnejšie využitie, napríklad pri spracovávaní veľkého množstva údajov, spracovávaní jazyka (NLP) alebo počítačom videní a klasifikácii, a na slabé miesta a nevýhody týchto modelov. Rád by som taktiež jednoducho opísal aj už existujúci systém, ako algoritmus GPT-3 vyvinutý nadáciou OpenAI, a možnosti jeho využitia.

1 Úvod

Strojové učenie je typ umelej inteligencie, pri ktorom softvér predpovedá výsledok niektorej operácie presnejšie. Strojové učenie sa zameriava na aplikácie a systémy, ktoré sa učia z veľkého množstva údajov a vylepšujú vďaka nim po nejakom čase svoju presnosť, namiesto toho aby boli natvrdo naprogramované. Je veľmi nápomocný hlavne pri praktických úlohách, pri ktorých sa môže učiť zo získaných údajov. Typickým príkladom môže byť predpoveď počasia alebo analýza medicínskych snímok pri diagnostike rakoviny. Čím viac údajov sa do systému strojového učenia vloží, tým viac je daný systém presnejší. Strojové učenie je aplikované v dnešnej dobe takmer všade. Používa sa v robotike, virtuálnych asistentoch (Google, Siri, Alexa), hrách, spracovávaní jazyka a textu, odporúčania produktov v online obchodoch, analýza nákupov zákazníka, vyhľadávačoch (Google Search), detekcia podvodov, boti (Chatbot), filtrovanie spamu a taktiež aj v sociálnych sieťach. Existujú dve hlavné metódy strojového učenia, a to Supervised Learning (učenie s dohľadom niekoho, kto kontroluje výstupy systému) a Unsupervised Learning (stroj kontroluje výstupy sám). Každý z týchto systémov má svoje výhody a aj nevýhody. Týmto rozdeleniam sa budem venovať v nasledujúcich sekciách.

*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2021/22, vedenie: Ing. Fedor Lehocki

1.1 Motivácia

2 Základy strojového učenia

3 Regresia

3.1 Lineárna regresia

3.2 Decision tree

3.3 Random Forest

3.4 Neurónová sieť

4 Klasifikácia

4.1 Logická regresia

4.2 Support Vector Machine

4.3 Naive Bayes

4.4 Neurónová sieť

5 Praktické využitie strojového učenia

5.1 Medicína

5.2 Spracovávanie dát

5.3 Poľnohospodárstvo

5.4 Automatizácia

5.5 Autopilot

6 Model GPT-3

7 Diskusia

8 Záver