Umělá inteligence, rizika, příležitosti

prezentace je v pracovní verzi

Martina Šimůnková

Katedra matematiky, FP TUL

3. června 2025

Úvod

- Co je pro vás umělá inteligence? Co si pod tímto pojmem představujete?
- Co si představujete pod pojmem inteligence? Je ve vašich očích umělá inteligence opravdu inteligentní?
- Jaké nástroje umělé inteligence denně používáte?
- Jaké další nástroje umělé inteligence znáte, případně občas používáte?
- Bez kterých nástrojů si svůj dnešní život dokážete jen těžko představit?
- Které nástroje byste naopak nejraději viděli z vašeho světa zmizet?

Úvod

- Co je pro vás umělá inteligence? Co si pod tímto pojmem představujete?
- Co si představujete pod pojmem inteligence? Je ve vašich očích umělá inteligence opravdu inteligentní?
- Jaké nástroje umělé inteligence denně používáte?
- Jaké další nástroje umělé inteligence znáte, případně občas používáte?
- Bez kterých nástrojů si svůj dnešní život dokážete jen těžko představit?
- Které nástroje byste naopak nejraději viděli z vašeho světa zmizet?

Úvod

- Co je pro vás umělá inteligence? Co si pod tímto pojmem představujete?
- Co si představujete pod pojmem inteligence? Je ve vašich očích umělá inteligence opravdu inteligentní?
- Jaké nástroje umělé inteligence denně používáte?
- Jaké další nástroje umělé inteligence znáte, případně občas používáte?
- Bez kterých nástrojů si svůj dnešní život dokážete jen těžko představit?
- Které nástroje byste naopak nejraději viděli z vašeho světa zmizet?

Úvod

- Co je pro vás umělá inteligence? Co si pod tímto pojmem představujete?
- Co si představujete pod pojmem inteligence? Je ve vašich očích umělá inteligence opravdu inteligentní?
- Jaké nástroje umělé inteligence denně používáte?
- Jaké další nástroje umělé inteligence znáte, případně občas používáte?
- Bez kterých nástrojů si svůj dnešní život dokážete jen těžko představit?
- Které nástroje byste naopak nejraději viděli z vašeho světa zmizet?

Úvod

- Co je pro vás umělá inteligence? Co si pod tímto pojmem představujete?
- Co si představujete pod pojmem inteligence? Je ve vašich očích umělá inteligence opravdu inteligentní?
- Jaké nástroje umělé inteligence denně používáte?
- Jaké další nástroje umělé inteligence znáte, případně občas používáte?
- Bez kterých nástrojů si svůj dnešní život dokážete jen těžko představit?
- Které nástroje byste naopak nejraději viděli z vašeho světa zmizet?

Úvod

- Co je pro vás umělá inteligence? Co si pod tímto pojmem představujete?
- Co si představujete pod pojmem inteligence? Je ve vašich očích umělá inteligence opravdu inteligentní?
- Jaké nástroje umělé inteligence denně používáte?
- Jaké další nástroje umělé inteligence znáte, případně občas používáte?
- Bez kterých nástrojů si svůj dnešní život dokážete jen těžko představit?
- Které nástroje byste naopak nejraději viděli z vašeho světa zmizet?

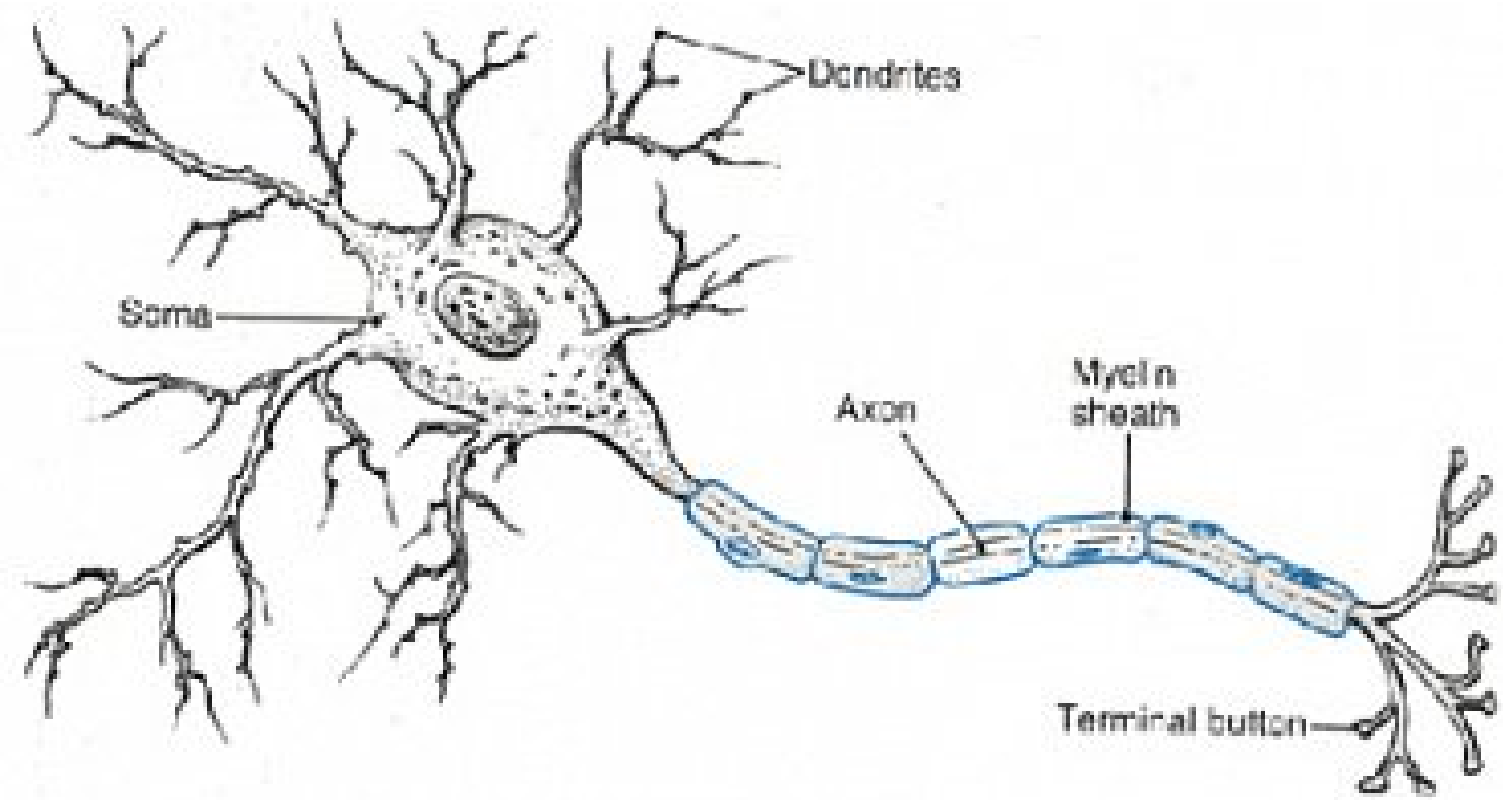
Program

1. Co jsou hluboké neuronové sítě.
2. Geoffrey Hinton, „otec“ hlubokých sítí a techno pesimista.
3. Kai-fu Lee, taiwanský vědec, vizionář, investor, techno realista.
4. Naše digitální stopa a její rizika.
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další zdroje.

Program

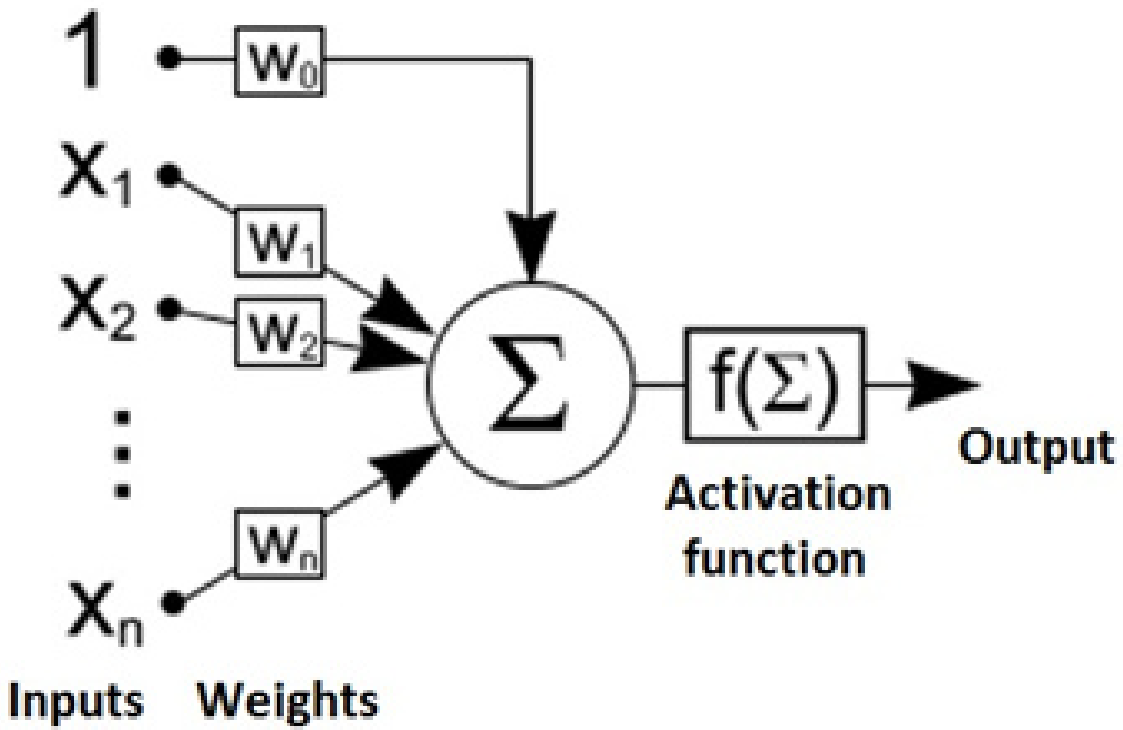
1. Co jsou hluboké neuronové sítě.
2. Geoffrey Hinton, „otec“ hlubokých sítí a techno pesimista.
3. Kai-fu Lee, taiwanský vědec, vizionář, investor, techno realista.
4. Naše digitální stopa a její rizika.
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další zdroje.

Nervová buňka



https://www.root.cz/clanky/biologicke-algoritmy-4-neuronove-site

Umělý neuron (perceptron, 1957)

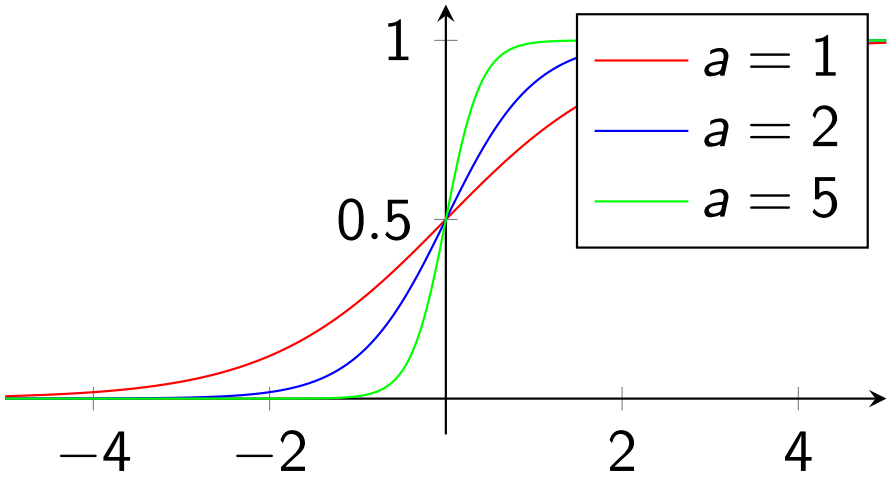


https://www.root.cz/clanky/biologicke-algoritmy-4-neuronove-site

$$\Sigma = w_0 + \sum_{k=1}^{n} w_k x_k$$

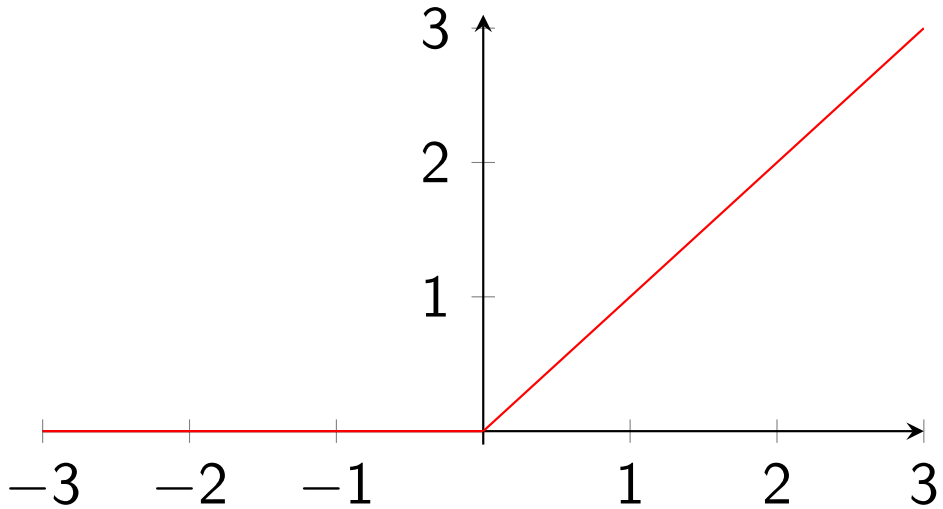
Příklady aktivačních funkcí

Sigmoid



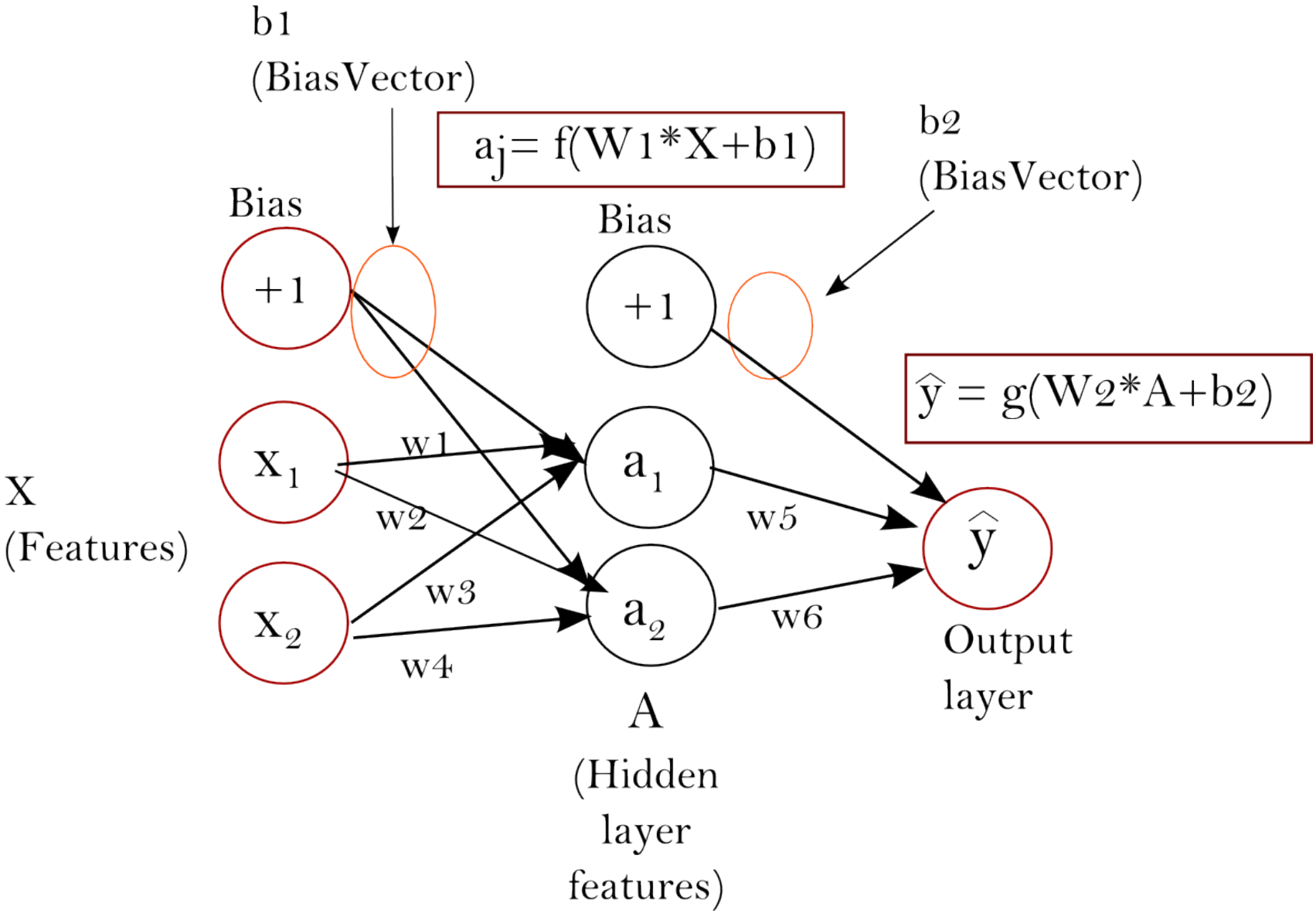
$$f(\Sigma) = \frac{1}{1 + \exp(-a\Sigma)}$$

ReLU (rectified linear unit)



$$f(\Sigma) = \max(0, \Sigma)$$

Hluboká neuronová síť (alespoň jedna skrytá vrstva)



https://i.sstatic.net/5SkqL.png

Vstup (input) a výstup (output) hluboké sítě

- Vstup: Informace o pixelech obrázku.
  Výstup: Pravděpodobnost, že na obrázku je ...
- Vstup: Text.
  Výstup: Pravděpodobnostní rozložení pokračujícího slova/tokenu textu.
- Vstup: Pozice v partii deskové hry (například šachy, go).
  Výstup: Pravděpodobnostní rozložení dalšího tahu, pravděpodobnost výhry.
- Vstup: Informace o žadateli o půjčku.
  Výstup: Pravděpodobnost, že žadatel bude půjčku řádně splácet.

Supervised learning

1. Sestavíme hlubokou síť a nastavíme náhodně její váhy w_k.
2. Sestavíme sadu dat. Ke každé položce v datech máme výstup v_0 nebo sadu výstupů (v_k).
3. Sadu rozdělíme na trénovací a testovací část.
4. Síti předložíme jednu položku trénovacích dat a podle výstupu sítě (y_k) upravíme váhy sítě w_i tak, aby se snížila hodnota odchylky

$$\sum_k (v_k - y_k)^2$$

   Váhy se přenastavují pomocí algoritmu zpětného šíření chyby (error backpropagation).
5. Opakujeme krok 4 pro další položky dat.
6. Kroky 4, 5 můžeme zopakovat několikrát pro celou datovou sadu.
7. Pomocí testovací sady dat natrénovanou síť otestujeme.

Reinforcement learning

Neuronová síť řídí chování agenta v prostředí, reaguje na změny v prostředí.

Chování agenta je po určitém počtu akcí ohodnoceno.

Na základě ohodnocení jsou přenastaveny váhy sítě.

Příklady:
Samořiditelná auta: počet kilometrů bez nehody.
Odpověď chatbota: ohodnocení uživatele (například počtem hvězdiček).
Konverzace chatbota s uživatelem: $+-$ ohodnotí člověk v roli hodnotitele (anotátora).
Desková hra: výhra, prohra.
Pohyb robota: zda a jak rychle splní úkol.

Metoda nejmenších čtverců – formulace problému

Popis problému:
Daty $\{(x_i, y_i)\}_{i=1}^n$ prokládáme přímku $y = ax + b$. Zvolíme koeficienty a, b, pro které má minimální hodnotu cenová funkce c

$$c(a, b) = \sum_{i=1}^{n} (ax_i + b - y_i)^2$$

Metoda nejmenších čtverců – výpočet rovnice přímky

Přímý výpočet a, b**:** Z podmínky $\nabla c = 0$ dostaneme soustavu

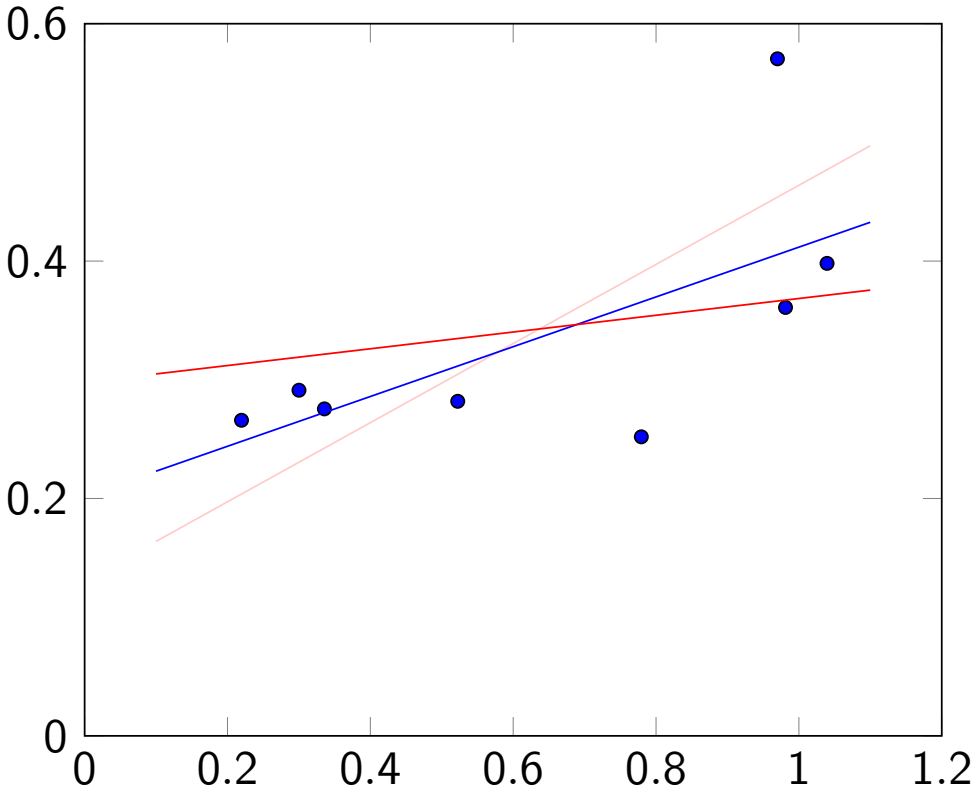
$$\begin{aligned} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i + nb &= \sum_{i=1}^{n} y_i \end{aligned}$$

Výpočet strojovým učením (machine learning): Zvolíme náhodně koeficienty a, b a postupně je pomocí dat aktualizujeme metodou gradient descent, α je vhodně zvolené malé kladné číslo

$$(a, b)_{new} = (a, b) - \alpha\nabla c$$

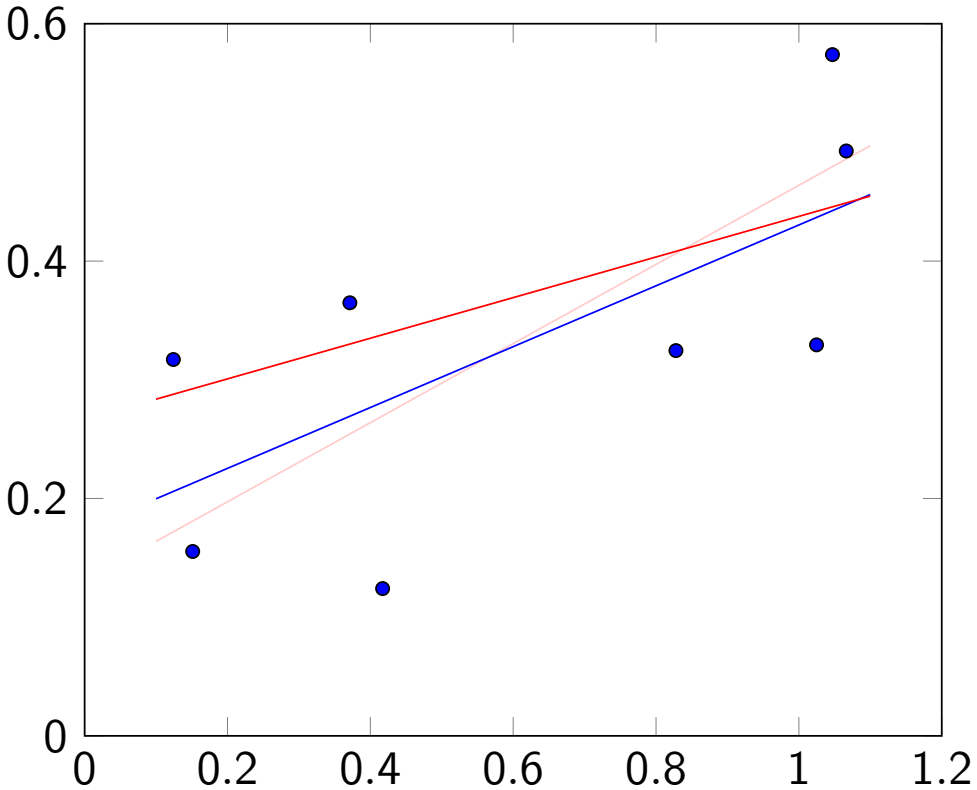
Metoda nejmenších čtverců – gradient descent method

Ilustrace strojového učení bez neuronové sítě:
Sytě červená přímka je zvolena náhodně a postupně je daty upravovaná metodou gradient descent.
Slabě červená přímka je proložená všemi daty. Modrá aktuálními daty.



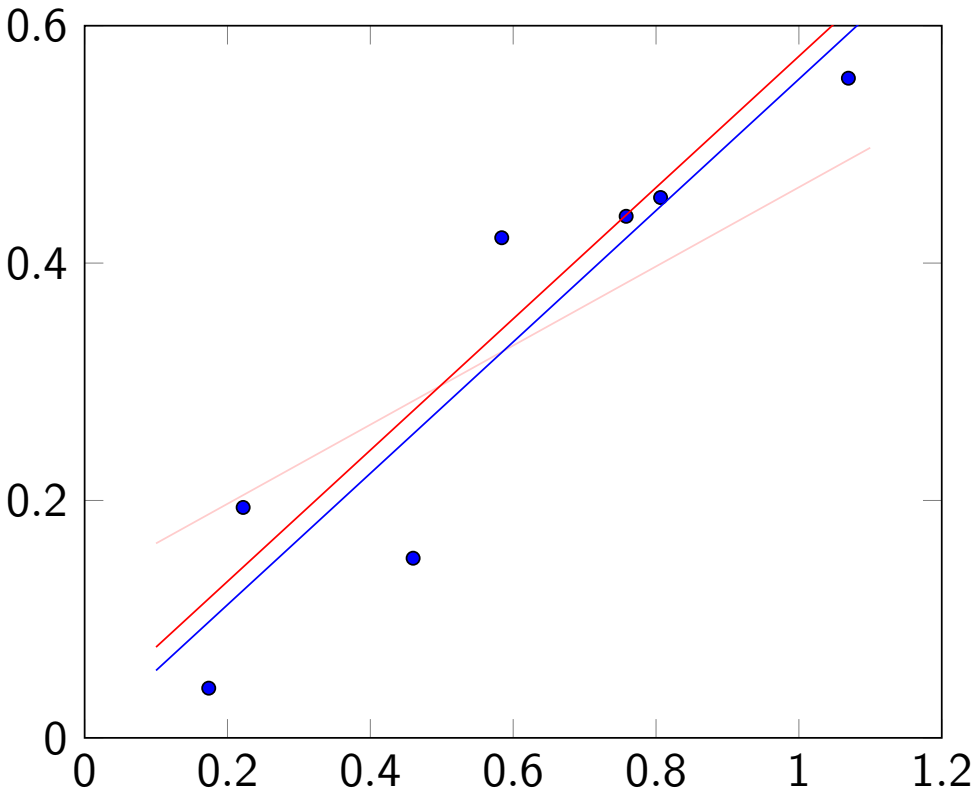
Metoda nejmenších čtverců – gradient descent method

Ilustrace strojového učení bez neuronové sítě:
Sytě červená přímka je zvolena náhodně a postupně je daty upravovaná metodou gradient descent.
Slabě červená přímka je proložená všemi daty. Modrá aktuálními daty.



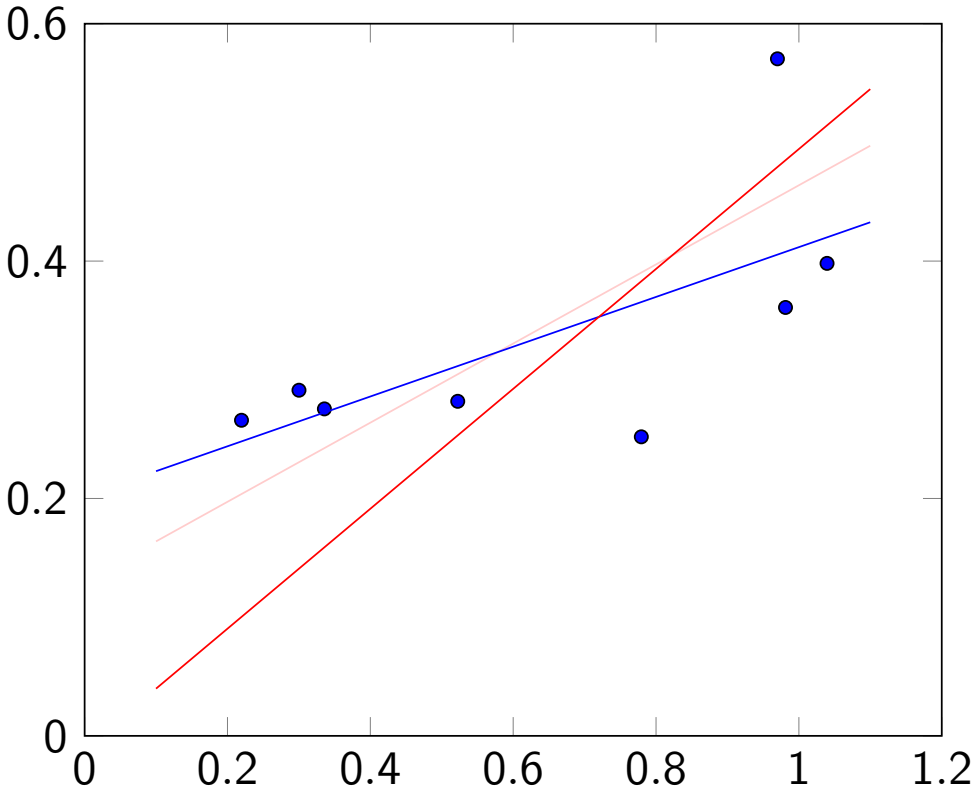
Metoda nejmenších čtverců – gradient descent method

Ilustrace strojového učení bez neuronové sítě:
Sytě červená přímka je zvolena náhodně a postupně je daty upravovaná metodou gradient descent.
Slabě červená přímka je proložená všemi daty. Modrá aktuálními daty.



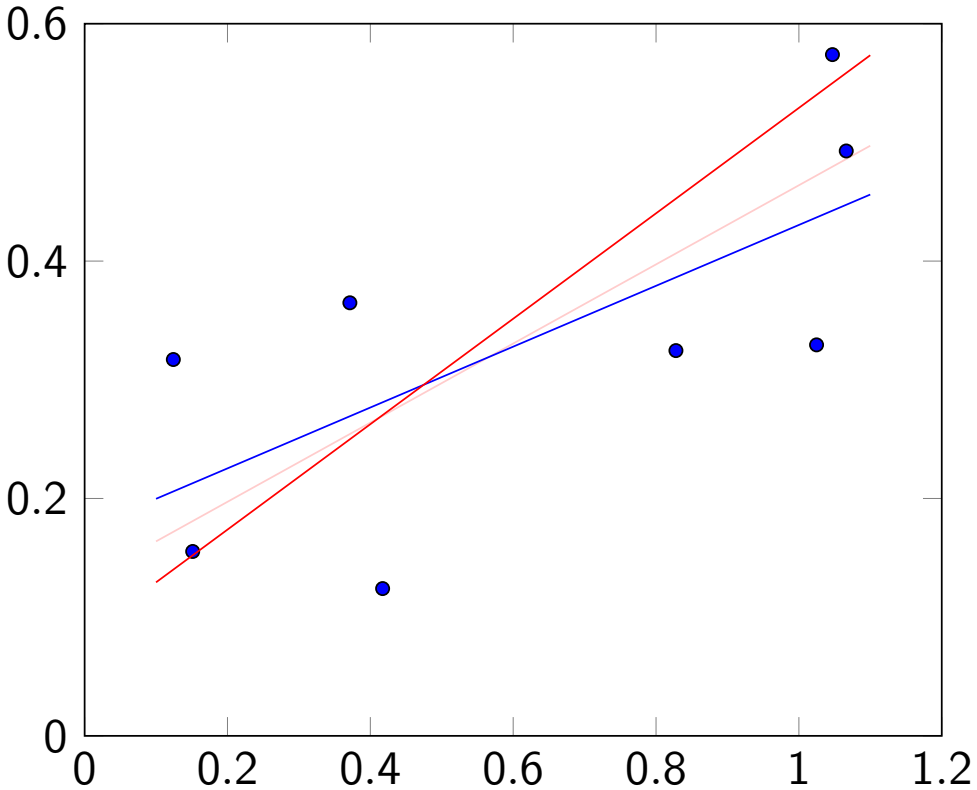
Metoda nejmenších čtverců – gradient descent method

Ilustrace strojového učení bez neuronové sítě:
Sytě červená přímka je zvolena náhodně a postupně je daty upravovaná metodou gradient descent.
Slabě červená přímka je proložená všemi daty. Modrá aktuálními daty.



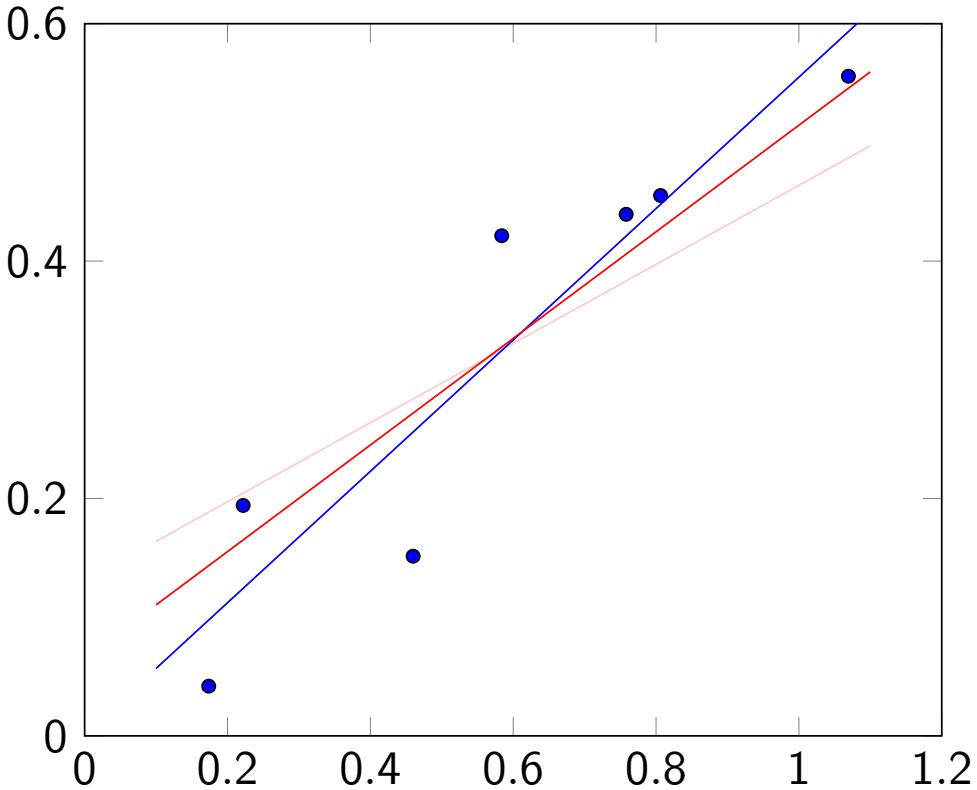
Metoda nejmenších čtverců – gradient descent method

Ilustrace strojového učení bez neuronové sítě:
Sytě červená přímka je zvolena náhodně a postupně je daty upravovaná metodou gradient descent.
Slabě červená přímka je proložená všemi daty. Modrá aktuálními daty.



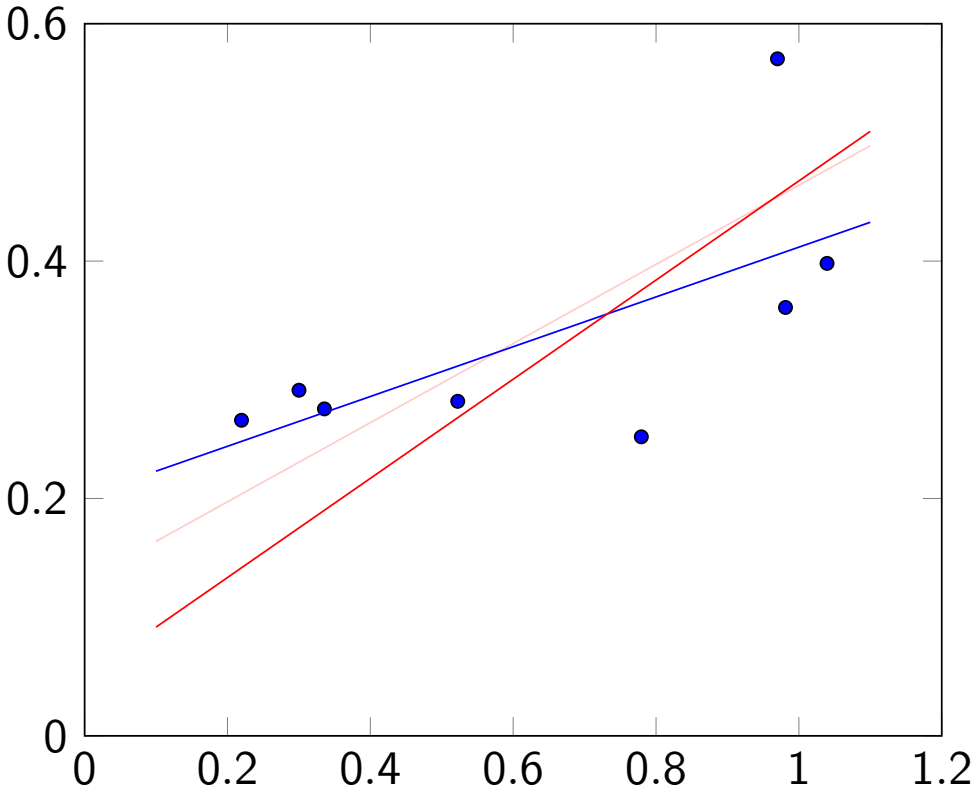
Metoda nejmenších čtverců – gradient descent method

Ilustrace strojového učení bez neuronové sítě:
Sytě červená přímka je zvolena náhodně a postupně je daty upravovaná metodou gradient descent.
Slabě červená přímka je proložená všemi daty. Modrá aktuálními daty.



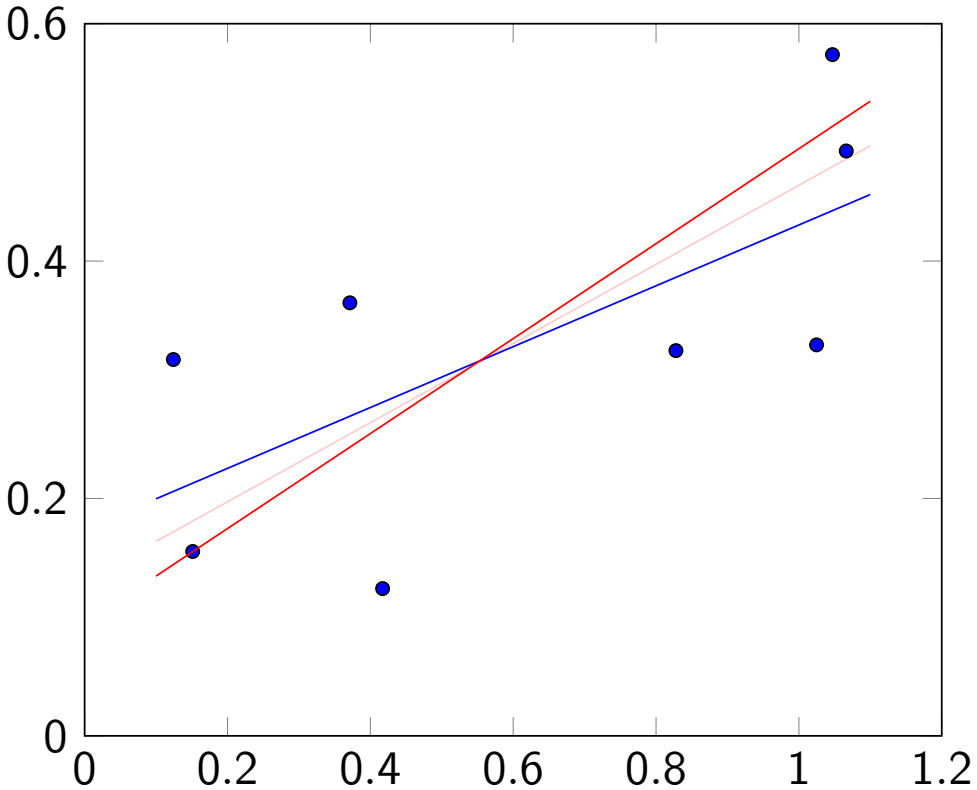
Metoda nejmenších čtverců – gradient descent method

Ilustrace strojového učení bez neuronové sítě:
Sytě červená přímka je zvolena náhodně a postupně je daty upravovaná metodou gradient descent.
Slabě červená přímka je proložená všemi daty. Modrá aktuálními daty.



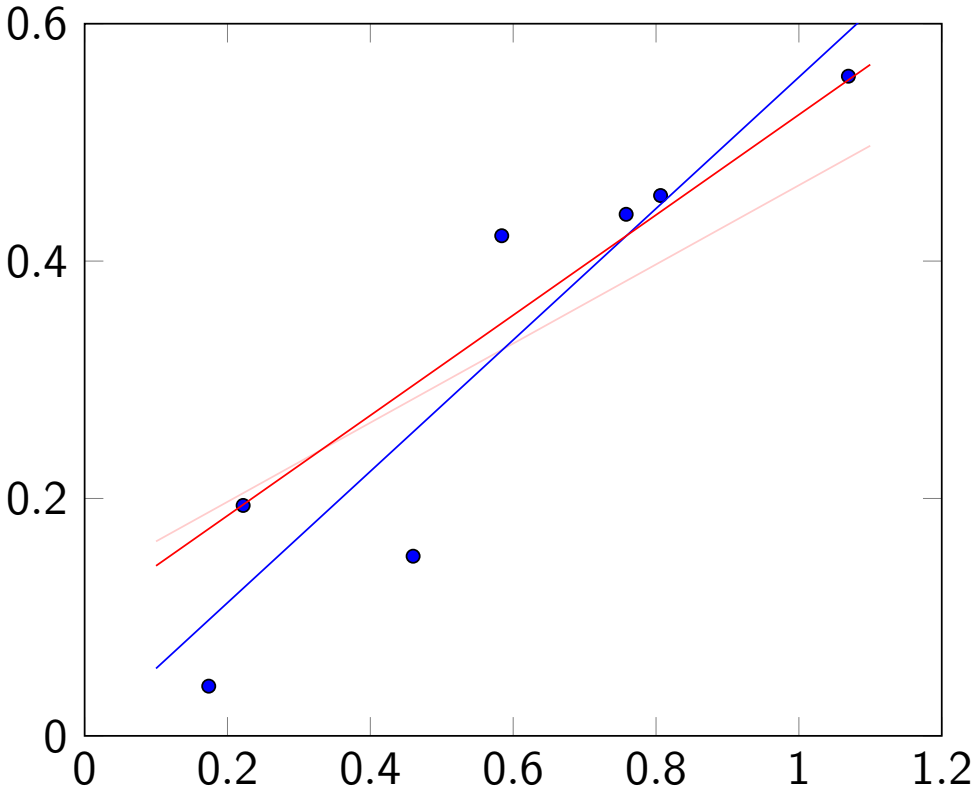
Metoda nejmenších čtverců – gradient descent method

Ilustrace strojového učení bez neuronové sítě:
Sytě červená přímka je zvolena náhodně a postupně je daty upravovaná metodou gradient descent.
Slabě červená přímka je proložená všemi daty. Modrá aktuálními daty.



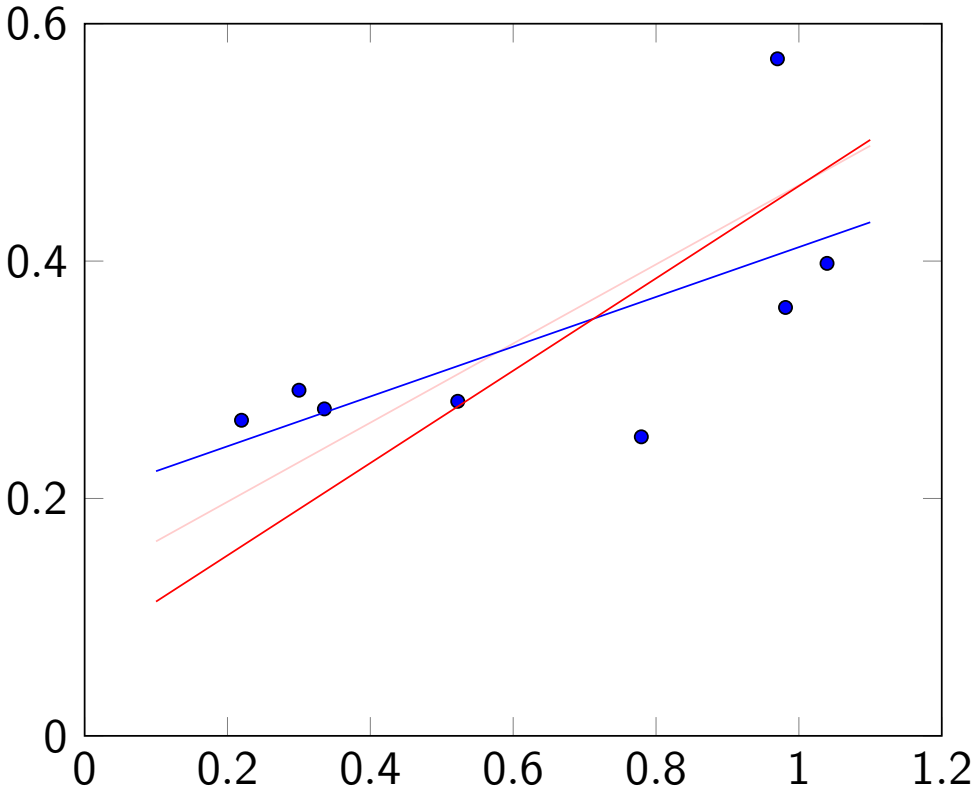
Metoda nejmenších čtverců – gradient descent method

Ilustrace strojového učení bez neuronové sítě:
Sytě červená přímka je zvolena náhodně a postupně je daty upravovaná metodou gradient descent.
Slabě červená přímka je proložená všemi daty. Modrá aktuálními daty.



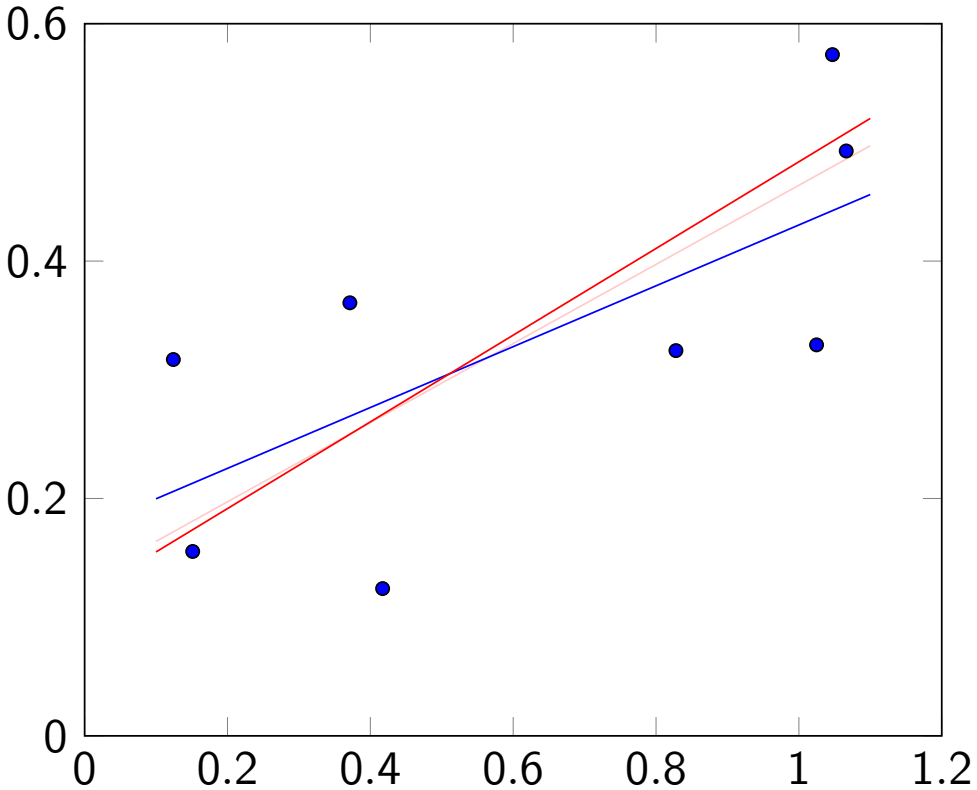
Metoda nejmenších čtverců – gradient descent method

Ilustrace strojového učení bez neuronové sítě:
Sytě červená přímka je zvolena náhodně a postupně je daty upravovaná metodou gradient descent.
Slabě červená přímka je proložená všemi daty. Modrá aktuálními daty.



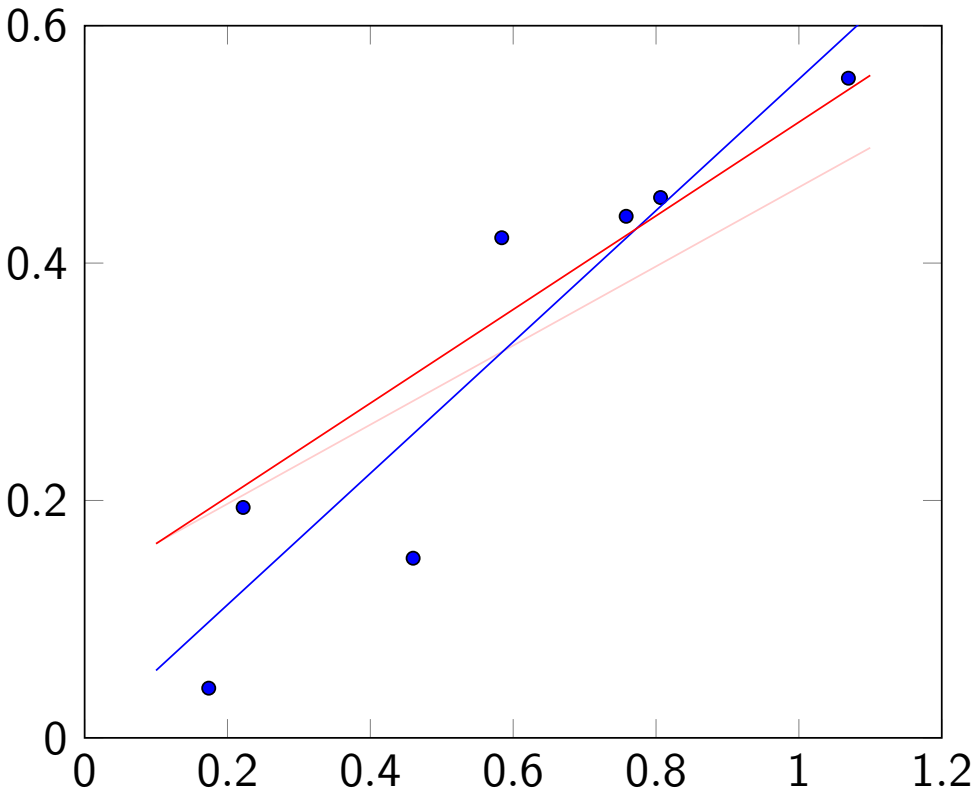
Metoda nejmenších čtverců – gradient descent method

Ilustrace strojového učení bez neuronové sítě:
Sytě červená přímka je zvolena náhodně a postupně je daty upravovaná metodou gradient descent.
Slabě červená přímka je proložená všemi daty. Modrá aktuálními daty.



Metoda nejmenších čtverců – gradient descent method

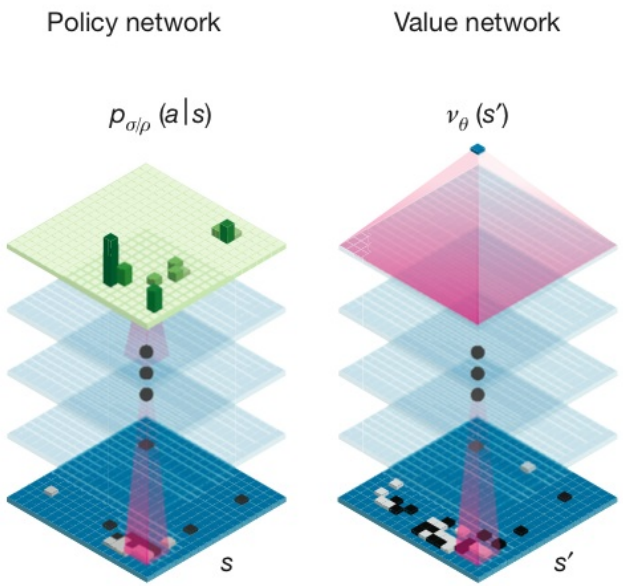
Ilustrace strojového učení bez neuronové sítě:
Sytě červená přímka je zvolena náhodně a postupně je daty upravovaná metodou gradient descent.
Slabě červená přímka je proložená všemi daty. Modrá aktuálními daty.



Deskové hry

DeepBlue
V roce 1997 porazilo Garriho Kasparova.
Uvažuje a propočítává všechny možné tahy do určité hloubky.

AlphaGo
V roce 2016 porazilo I Se-tola.
Uvažuje jen tahy vybrané neuronovou sítí, která je natrénovaná primárně na partiích lidí a následně na hře proti sobě.



AlphaZero
Systém, který se učí výhradně hraním sám proti sobě, na vstupu má jen pravidla. Otestován na hrách šachy, go, šógi.

Program

1. Co jsou hluboké neuronové sítě.
2. Geoffrey Hinton, „otec“ hlubokých sítí a techno pesimista.
3. Kai-fu Lee, taiwanský vědec, vizionář, investor, techno realista.
4. Naše digitální stopa a její rizika.
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další zdroje.

Geoffrey Hinton

* 1947, Londýn

1970s na universitě v Edinburgu studuje lidskou mysl a k tomu používá jako nástroj simulaci počítačem
`https://www.youtube.com/watch?v=qrvK_KuIeJk&t=100s`

2012 průlom v rozpoznávání obrazů na soutěži ImageNet

2013 – 2023 člen týmu Google Brain

2018 Turingova cena za „konceptuální a technické průlomy, které z hlubokých neuronových sítí učinily kritickou komponentu výpočetní techniky“ spolu s Yoshuou Bengiem a Yannem LeCunem (formulace z wikipedie)

2024 Nobelova cena za fyziku za „zásadní objevy a vynálezy, které umožňují strojové učení s umělými neuronovými sítěmi“ spolu s Johnem Hopfieldem (formulace z wikipedie)

ImageNet

ImageNet je soutěž v rozpoznávání obrazů.
Dopředu je určeno tisíc kategorií objektů a při soutěži je softwaru předloženo padesát tisíc obrázků, kterým je přiřazena jedna z těchto tisíců kategorií. Software každému obrázku přiřadí až pět těchto kategorií a pokud se jedním z nich trefí, je považováno vyhodnocení za správné.

Odkazy:
`www.pinecone.io/learn/series/image-search/imagenet/`
Chybovost v ročníku 2012.
Graf chybovosti v ročnících 2010, 2011, 2012.
Chybovost v ročníku 2013.

Názory Geoffrey Hintona

"You believe they can understand." "Yes." ...

https://www.youtube.com/watch?v=qrvK_KuIeJk&t=64s

(1:04 – 6:11)

Chatbot understands

https://www.youtube.com/watch?v=qrvK_KuIeJk&t=497s

(8:17 – 10:32)

Příležitosti v medicíně

https://www.youtube.com/watch?v=qrvK_KuIeJk&t=630s

(10:30 – 10:55)

Hrozby: fake news, pracovní trh, autonomní zbraně

https://www.youtube.com/watch?v=qrvK_KuIeJk&t=668s

(10:55 – 11:22)

Můžeme se hrozbám vyhnout?

https://www.youtube.com/watch?v=qrvK_KuIeJk&t=682s

(11:22 –)

Program

1. Co jsou hluboké neuronové sítě.
2. Geoffrey Hinton, „otec“ hlubokých sítí a techno pesimista.
3. Kai-fu Lee, taiwanský vědec, vizionář, investor, techno realista.
4. Naše digitální stopa a její rizika.
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další zdroje.

Aplikace strojového vidění

V pětiminutovém videu zopakujeme princip strojového učení.

Dále uvidíte aplikaci strojového vidění. O důsledcích, rizicích a příležitostech použití aplikace budeme následně diskutovat.

`https://www.youtube.com/watch?v=aZ5EsdnpLMI&t=116s`

Kai-fu Lee

* 1961 Taiwan
1973 stěhuje se do USA za bratrem
1986 PhD na Carnegie Mellon University, rozpoznávání řeči
2009 zakládá v Číně Sinovation Ventures (fond pro rizikové investice)
2013 mu byla diagnostikována lymfómie, přežil a změnil svůj vztah k práci, nejbližším, technologiím

„I think most people have no idea and many people have wrong idea.“
. . . „more than electricity.“
`https://www.youtube.com/watch?v=aZ5EsdnpLMI&t=69s`

Program

1. Co jsou hluboké neuronové sítě.
2. Geoffrey Hinton, „otec“ hlubokých sítí a techno pesimista.
3. Kai-fu Lee, taiwanský vědec, vizionář, investor, techno realista.
4. Naše digitální stopa a její rizika.
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další zdroje.

Christopher Wylie, Mindf*ck. Inside Cambridge Analytica's Plot to Break the World, 2019

Zpočátku to byl ten nejnezajímavější start projektu v celé historii. Nic se nedělo. Uplynulo pět, deset, patnáct minut a lidi začali být netrpěliví. „Co to kurva má být?“ ozval se Nix. „Proč se na tohle máme koukat?“ Ale já věděl, že to bude chvíli trvat, než si lidé na MTurk dotazníku všimnou, vyplní ho a nainstalují si aplikaci, aby dostali zaplaceno. Chvilku potom, co se Nix začal rozčilovat, dorazil první výsledek.

A pak nastala lavina. Dorazila první dávka dat, pak dvě další, pak jich přišlo dvacet, sto, tisíc – všechno během pár vteřin.

⋮

Počty údajů v tabulkách začaly narůstat exponenciálně, jak se do databáze načítaly profily facebookových přátel. Byl to úžasný zážitek pro všechny, ale pro datové vědce mezi námi to byl přímo adrenalin.

Christopher Wylie, Mindf*ck. Inside Cambridge Analytica's Plot to Break the World, 2019

Bannon začal jezdit do Londýna častěji, aby si ověřoval, jak pokračujeme. Jedna z jeho návštěv se uskutečnila krátce po tom, co jsme naši aplikaci spustili. Všichni jsme se vypravili do zasedačky, v jejímž čele byla obrovská obrazovka. Jucikas přednesl krátkou prezentaci, načež se obrátil na Bannona.

„Řekněte nějaké příjmení."

Bannon vypadal pobaveně a jedno řekl.

„Dobře. A teď stát."

„Já nevím, třeba Nebraska."

Jucikas zadal údaje a na obrazovce se objevil dlouhý seznam linků, v Nebrasce měla toto příjmení řada lidí. Klikl na jedno s ženským křestním jménem a na obrazovce se o dotyčné objevilo úplně všechno. Její fotografie, kde pracuje, kde bydlí. Kdo jsou její děti a kam chodí do školy. Jaké řídí auto. V roce 2012 volila Mitta Romneyho, miluje Katy Perry, jezdí v Audi, není to žádná intelektuálka ...

Christopher Wylie, Mindf*ck. Inside Cambridge Analytica's Plot to Break the World, 2019

„Máme na ně i telefonní čísla?“ zeptal se. Řekl jsem, že ano. Načež, v jednom z těch brilantních okamžiků, které čas od času míval, vzal do ruky mikrofon telefonu s hlasitým odposlechem a požádal o číslo. Vyťukával číslo podle toho, jak mu je Jucikas předříkával.
Po několika zazvonění se ozval ženský hlas. „Haló?“ Načež Nix tím svým nejsnobštějším britským akcentem řekl „Dobrý den, madam. Velice se omlouvám, že vás obtěžuji. Volám z Cambridgeské univerzity, provádíme průzkum. Mohu mluvit s paní Jenny Smithovou, prosím?“ Žena potvrdila, že je Jenny a Nix jí začal klást otázky, založené na tom, co věděl z dat na obrazovce.
Paní Smithová, zajímal by mě váš názor na televizní seriál *Hra o trůny*.
Jenny se začala rozplývat nadšením – přesně tak, jako na Facebooku . . .

Christopher Wylie, Mindf*ck. Inside Cambridge Analytica's Plot to Break the World, 2019

Když si to zpětně vybavuji, připadá mi naprosto šílené, že Bannon – který byl v té době nikdo a teprve za více než rok nechvalně proslul jako Trumpův poradce – seděl v naší kanceláři, volal náhodným Američanům, dával jim osobní otázky a ti lidé byli úplně šťastni, že na ně mohou odpovídat.
Dokázali jsme to. Zrekonstruovali jsme desítky miliónů životů v paměti počítače a potenciálně další stovky milionů mohly následovat. Byla to epochální chvíle. Byl jsem hrdý na to, že jsme vytvořili tak mocný nástroj. Byl jsem si jistý, že jsme vytvořili cosi, o čem budou lidé mluvit desítky let.

⋮

Mercer investoval do CA desítky milionů dolarů rřív, než firma vůbec získala nějaká data nebo vyvinula nějaký software pro Ameriku. Každý investor by to musel považovat za velice riskantní investici. Na druhé straně jsme věděli, že Mercer není ani hloupý ani neopatrný a jistě si riziko

Christopher Wylie, Mindf*ck. Inside Cambridge Analytica's Plot to Break the World, 2019

dobře spočítal. Celou dobu si mnoho z nás myslelo, že když Mercer podstoupil takové riziko, musel očekávat, že díky našim výsledkům vydělá moře peněz pomocí svého hedgeového fondu. Jinými slovy, mysleli jsme si, že firma nevznikla kvůli nějakým revolučním politickým plánům alt-right, ale aby Mercerovi vydělala peníze. Nixův vřelý vztah k penězům tento pocit ještě zesiloval.
Dneska už samozřejmě víme, že to tak vůbec nebylo. Nevím, co víc k tomu říct, než že jsem byl naivnější, než jsem si tenkrát připouštěl. I když jsem měl na svůj věk už dost zkušeností, bylo mi dvacet čtyři a zjevně jsem se ještě musel o životě hodně naučit. Když jsem k SCL nastoupil, představoval jsem si, že pomůžu firmě vyvíjet program boje proti radikalizaci a pomůžu tím chránit Británii, Ameriku a jejich spojence před hrozbami šířenými na internetu. Časem jsem si zvykl na podivné prostředí takové práce, které bralo jako normální řadu věcí, jež by náhodnému pozorovateli připadaly podivné.

Christopher Wylie, Mindf*ck. Inside Cambridge Analytica's Plot to Break the World, 2019

Sociální média také používají techniky, které aktivují „hravou smyčku“ a „nepravidelný rozvrh posilování“ v našem mozku. Jsou to zážitky častých, ale nepravidelných odměn, jež vytvářejí jisté očekávání, ale finální odměna je tak nepředvídatelná, že ji nelze plánovat. Tím se vytváří cyklus nejistoty, očekávání a zpětné vazby, který se stále posiluje. Náhodný režim hracího automatu znemožňuje hráči vytvořit nějakou vítěznou strategii nebo plán, takže jediná cesta k odměně je hrát pořád dál. Četnost odměn je naprogramovaná tak, aby hráče stimulovala v okamžiku, kdy už ztrácí energii pokračovat a udržovala ho ve hře. V případě hazardních hráčů vydělává kasino podle toho, kolika kol se hráč účastní. V případě sociálních médií vydělává platforma podle toho, kolikrát uživatel na něco klikne. Proto existuje v newsfeedu nekonečné skrolování – mezi uživatelem, který znovu a znovu posunuje text na obrazovce a gamblerem, který znovu a znovu tiskne tlačítko hracího automatu, je jen nepatrný rozdíl.

Christopher Wylie, Mindf*ck. Inside Cambridge Analytica's Plot to Break the World, 2019

V létě roku 2014 začala Cambridge Analytica na Facebooku a dalších platformách vytvářet fiktivní stránky, které vypadaly jako skutečná diskusní fóra, skupiny a zdroje zpráv. To byla běžná a dobře vyzkoušená taktika, kterou mateřská firma SCL používala v operacích proti extremistům v jiných částech světa. Není mi jasné, kdo ve firmě vydával příkazy k provádění těchto dezinformačních operací, ale pro mnoho příslušníků staré gardy, kteří na takových operacích po celém světě roky pracovali, bylo toto všechno zcela normální. K americkému obyvatelstvu prostě přistupovali úplně stejně jako k Pákistáncům nebo Jemencům v projektech, které si Američané nebo Britové objednali.

Christopher Wylie, Mindf*ck. Inside Cambridge Analytica's Plot to Break the World, 2019

Interní testy také prokázaly, že digitální a sociální obsah reklamy, který Cambridge Analitica spustila, efektivně zvyšoval počet online zapojení. Osoby, které byly online cíleny testovacími reklamami, měly sociální profil sladěný se svými volebnímy záznamy, takže firma znala jejich jména a identitu v „reálném světě“. Firma pak začala používat čísla udávající míru zapojení (*engagement rates*), aby zjistila potenciální dopad reklam na účast ve volbách. Jedno interní memorandum rozebíralo výsledky experimentu, jehož se zúčastnili zaregistrovaní voliči, kteří poslední dvoje volby vynechali. Cambridge Analytica odhadovala, že kdyby se 25% těchto nevoličů, kteří začali klikat na materiály zasílané firmou, rozhodlo jít k volbám, mohl by se zvýšit počet voličů republikánské strany v klíčových státech přibližně o 1%, což je zhruba většina, kterou při vyrovnaných volbách vítězná strana získá.

Center for Human Technology

Center for Human Technology
https://www.humanetech.com/who-we-are

The Social Dilemma, film z roku 2020, trailer
https://www.imdb.com/video/vi303154457/?ref_=ttvg_vi_1

Vyhlížíme okamžik, kdy technologie překonají lidstvo . . . technologie dříve využijí lidské slabosti.
https://www.youtube.com/watch?v=iYVVgGWUKKg&t=315s

Jsou lidé tak zlí?
https://www.youtube.com/watch?v=iYVVgGWUKKg&t=977s

Sociální sítě jako bysnysový nástroj

Sociální sítě jsou bysnysový nástroj na cílení reklamy.

...

používají negativní zkreslení

...

naštvanost si společnost z virtuálního světa odnese do světa reálného.

Poslechneme si 15 až 20 minut trvající analýzu, pak o ní budeme diskutovat.

`https://www.youtube.com/watch?app=desktop&v=3ING7N_otRI&t=6m41s`

Negativní externalita (je zahrnuto v předchozím)

`https://www.youtube.com/watch?app=desktop&v=3ING7N_otRI&t=13m26s`

Program

1. Co jsou hluboké neuronové sítě.
2. Geoffrey Hinton, „otec“ hlubokých sítí a techno pesimista.
3. Kai-fu Lee, taiwanský vědec, vizionář, investor, techno realista.
4. Naše digitální stopa a její rizika.
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další zdroje.

Jak používat generativní roboty?

Směrnice rektora o využití umělé inteligence při výuce a tvůrčí práci (`https://doc.tul.cz/12546`)

- Četli jste tuto směrnici?
- Pracujete při výuce s touto směrnicí? Jakým způsobem?
- Kontrolujete její dodržování? Jakým způsobem?
- Zajímáte se, jakým způsobem vaši studenti s generativními roboty pracují?
  Byli byste tak laskavi a sdíleli s námi svá zjištění?
- Dáváte studentům rady/instrukce nad rámec směrnice? Jaké?
- Jakými zásadami se sami při používání generativních robotů řídíte?

Jak používat generativní roboty?

Směrnice rektora o využití umělé inteligence při výuce a tvůrčí práci (`https://doc.tul.cz/12546`)

- Četli jste tuto směrnici?
- Pracujete při výuce s touto směrnicí? Jakým způsobem?
- Kontrolujete její dodržování? Jakým způsobem?
- Zajímáte se, jakým způsobem vaši studenti s generativními roboty pracují?
  Byli byste tak laskavi a sdíleli s námi svá zjištění?
- Dáváte studentům rady/instrukce nad rámec směrnice? Jaké?
- Jakými zásadami se sami při používání generativních robotů řídíte?

Jak používat generativní roboty?

Směrnice rektora o využití umělé inteligence při výuce a tvůrčí práci (`https://doc.tul.cz/12546`)

- Četli jste tuto směrnici?
- Pracujete při výuce s touto směrnicí? Jakým způsobem?
- Kontrolujete její dodržování? Jakým způsobem?
- Zajímáte se, jakým způsobem vaši studenti s generativními roboty pracují?
  Byli byste tak laskavi a sdíleli s námi svá zjištění?
- Dáváte studentům rady/instrukce nad rámec směrnice? Jaké?
- Jakými zásadami se sami při používání generativních robotů řídíte?

Jak používat generativní roboty?

Směrnice rektora o využití umělé inteligence při výuce a tvůrčí práci (`https://doc.tul.cz/12546`)

- Četli jste tuto směrnici?
- Pracujete při výuce s touto směrnicí? Jakým způsobem?
- Kontrolujete její dodržování? Jakým způsobem?
- Zajímáte se, jakým způsobem vaši studenti s generativními roboty pracují?
  Byli byste tak laskavi a sdíleli s námi svá zjištění?
- Dáváte studentům rady/instrukce nad rámec směrnice? Jaké?
- Jakými zásadami se sami při používání generativních robotů řídíte?

Jak používat generativní roboty?

Směrnice rektora o využití umělé inteligence při výuce a tvůrčí práci (`https://doc.tul.cz/12546`)

- Četli jste tuto směrnici?
- Pracujete při výuce s touto směrnicí? Jakým způsobem?
- Kontrolujete její dodržování? Jakým způsobem?
- Zajímáte se, jakým způsobem vaši studenti s generativními roboty pracují?
  Byli byste tak laskavi a sdíleli s námi svá zjištění?
- Dáváte studentům rady/instrukce nad rámec směrnice? Jaké?
- Jakými zásadami se sami při používání generativních robotů řídíte?

Jak používat generativní roboty?

Směrnice rektora o využití umělé inteligence při výuce a tvůrčí práci (`https://doc.tul.cz/12546`)

- Četli jste tuto směrnici?
- Pracujete při výuce s touto směrnicí? Jakým způsobem?
- Kontrolujete její dodržování? Jakým způsobem?
- Zajímáte se, jakým způsobem vaši studenti s generativními roboty pracují?
  Byli byste tak laskavi a sdíleli s námi svá zjištění?
- Dáváte studentům rady/instrukce nad rámec směrnice? Jaké?
- Jakými zásadami se sami při používání generativních robotů řídíte?

Generativní modely vydělávají?

Chatboti zpravidla firmám snižují náklady a tím vydělávají.

Ne vždy to tak je:
https://www.svetandroida.cz/cursor-ai-bot-halucinace/

Diskuze pod článkem:
Je možné chatbota naučit nehalucinovat?
Naučit ho říct prosté nevím?

Generativní modely vydělávají?

Chatboti zpravidla firmám snižují náklady a tím vydělávají.

Ne vždy to tak je:
`https://www.svetandroida.cz/cursor-ai-bot-halucinace/`

Diskuze pod článkem:
Je možné chatbota naučit nehalucinovat?
Naučit ho říct prosté nevím?

Generativní modely vydělávají?

Chatboti zpravidla firmám snižují náklady a tím vydělávají.

Ne vždy to tak je:
`https://www.svetandroida.cz/cursor-ai-bot-halucinace/`

Diskuze pod článkem:
Je možné chatbota naučit nehalucinovat?
Naučit ho říct prosté nevím?

Word2vec

Word2vec je je číselná reprezentace slov pomocí vektorů velké dimenze.

Slovo jako soubor konceptů:
ptáci = živý organismus + množné číslo + podstatné jméno + létají + . . .
+ koncovky, pády . . . naučí se z dat (data driven).

Zahrnuje sémantické významy, například

$$\text{král - muž} + \text{žena} \doteq \text{královna}$$

Tomáš Mikolov o gramatice českého jazyka v jazykových modelech
`https://www.youtube.com/watch?v=AD1Qz1bBtWQ&t=961s`
(necelé dvě minuty)

Tomáš Mikolov, techno optimista

Z rozhovoru Tomáše Mikolova s Jakubem Horákem (z 16. 1. 2025, za paywallem na herohero):

Jazykový model reflektuje soubor trénovacích dat. Není tam nějaká mysl, která by něco plánovala, někam směřovala. je to papoušek, co opakuje, co bylo v trénovacích datech.
Kdo ovládá, co je v trénovacích datech, ovládá, co bude chatbot říkat.

Monetizace: jednak propagovat cokoliv, kdo si to zaplatí, druhak – dříve se používaly triky na zviditelnění v google vyhledávání – nyní hrozí zaplavení internetu texty ve snaze "probublat" reklamu do trénovacích dat (není snadné algoritmicky rozpoznat text psaný lidmi a text vygenerovaný, T.M. a spol. na to mají článek).

Vznikne vědomí? Určitě jo, jsem optimista, dožijeme se toho.

Tomáš Mikolov, techno optimista

Rekurentní sítě: byly nestabilní při trénování. Zatímco někteří psali články, že tato nestabilita brání jejich natrénování, Tomáš Mikolov si všimnul, že je možné nestabilitu detekovat, asi jedno procento trénovacích dat odstranit a poté je možné síť natrénovat.

Generativní modely, zdroje

Attention Is All You Need
`https://arxiv.org/pdf/1706.03762`

Výuková videa na kanálu 3Blue1Brown
`https://www.3blue1brown.com/topics/neural-networks`

Program

1. Co jsou hluboké neuronové sítě.
2. Geoffrey Hinton, „otec“ hlubokých sítí a techno pesimista.
3. Kai-fu Lee, taiwanský vědec, vizionář, investor, techno realista.
4. Naše digitální stopa a její rizika.
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další zdroje.

Další zdroje

Jan Hrach, Od umělého neuronu k ChatGPT
https://www.youtube.com/watch?v=o9TwtMywEuI
Prezentace: https://jenda.hrach.eu/f/gpt.pdf

Generative Large Language Multi-Modal Model
(všechno je jazyk, a to umožňuje zásadní urychlení vývoje; rychlost nazývají dvojitě exponenciální)
https://www.youtube.com/watch?v=xoVJKj8lcNQ&t=954s
V závěru videa srovnávají hrozbu z jaderné bomby s hrozbou AI. Jako lék navrhují nebrzdit vývoj, ale pouze nasazení nástrojů pro veřejnost. Před nasazením doporučují nástroje otestovat.

BBC: will Australia's social media ban for under 16-s work?
https://www.youtube.com/watch?v=-FNIAzYLCQA

Harari o roli Facebooku v masakrech v Myanmiru:
https://www.youtube.com/watch?v=_jl64f-821o&t=1554s