Umělá inteligence, rizika, příležitosti

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?

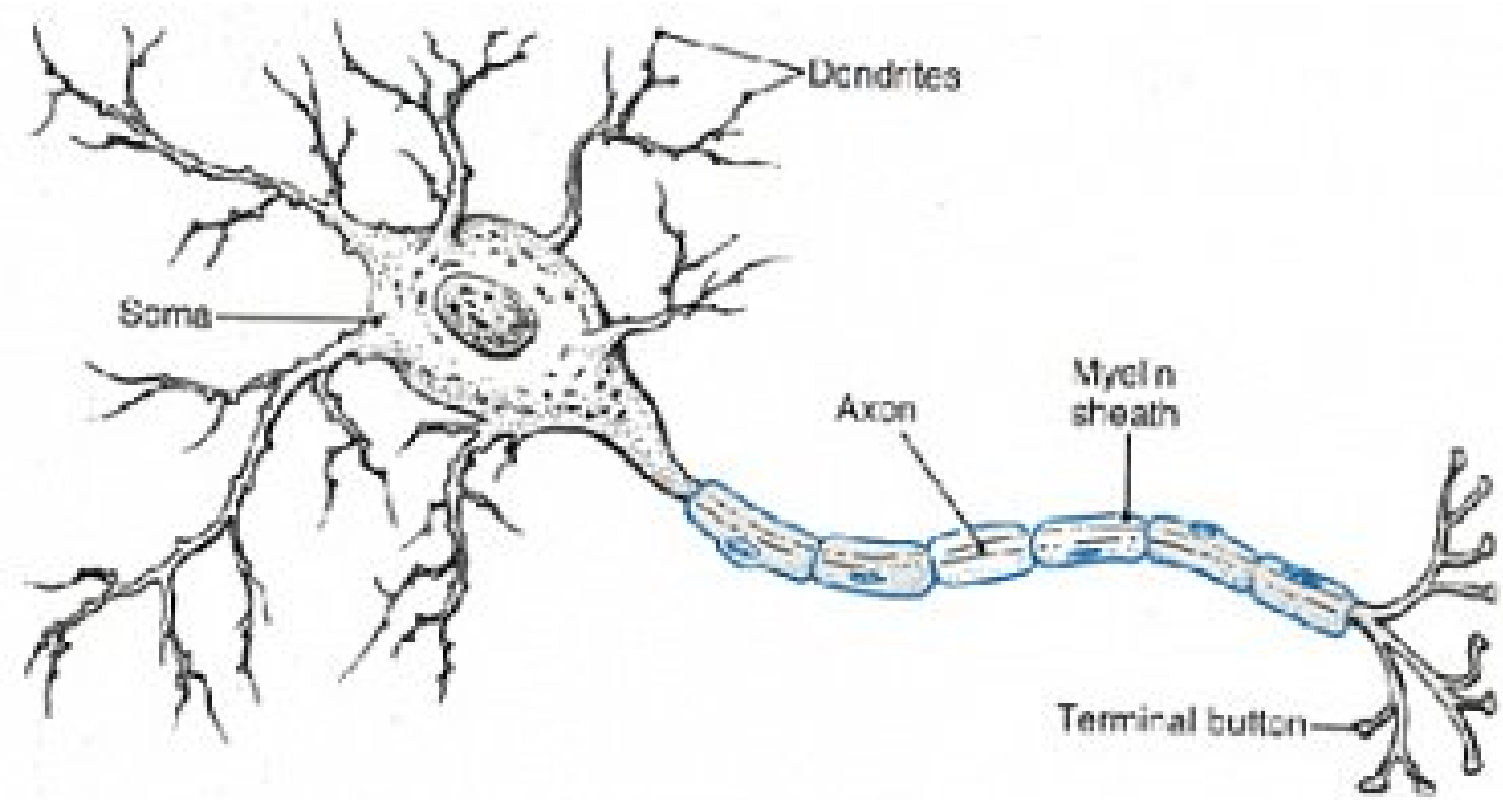
Program

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.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. Další zdroje.

Program

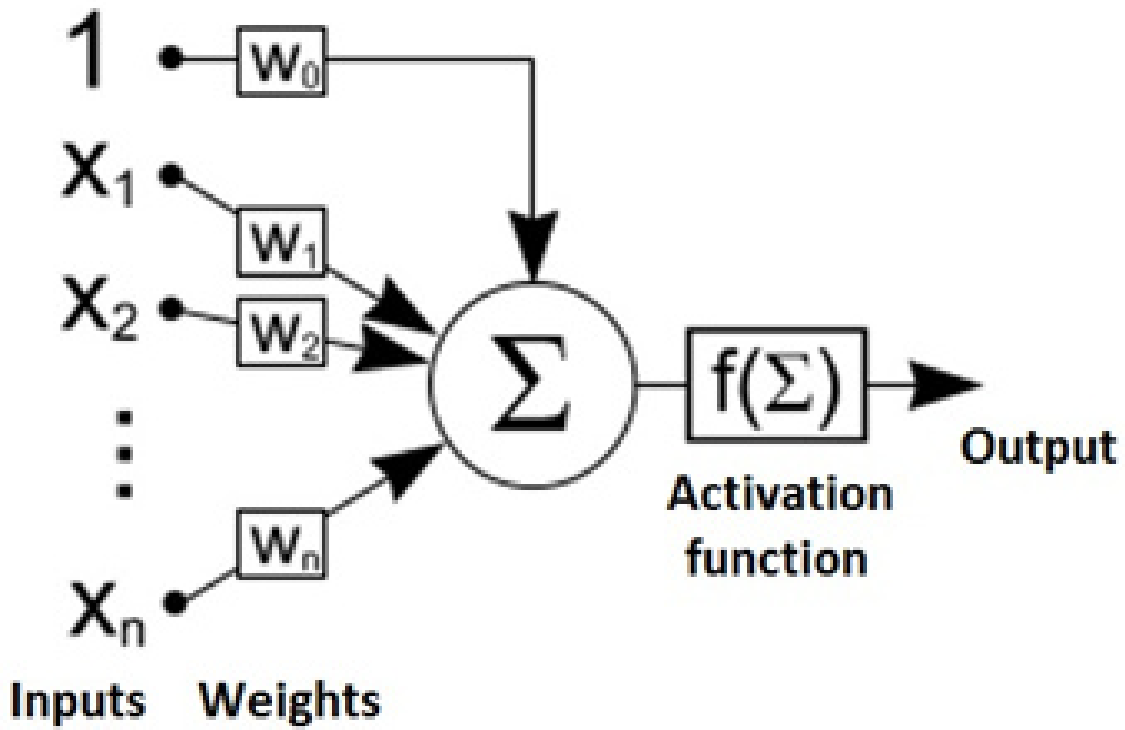
1. Co jsou hluboké neuronové sítě.
2. Geoffrey Hinton, „otec“ hlubokých sítí a techno pesimista.
3. Kai-fu Lee, taiwanský vědec, vizionář, investor, techno realista.
4. Naše digitální stopa a její rizika.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. Další zdroje.

Nervová buňka



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

Umělý neuron (perceptron, 1957)

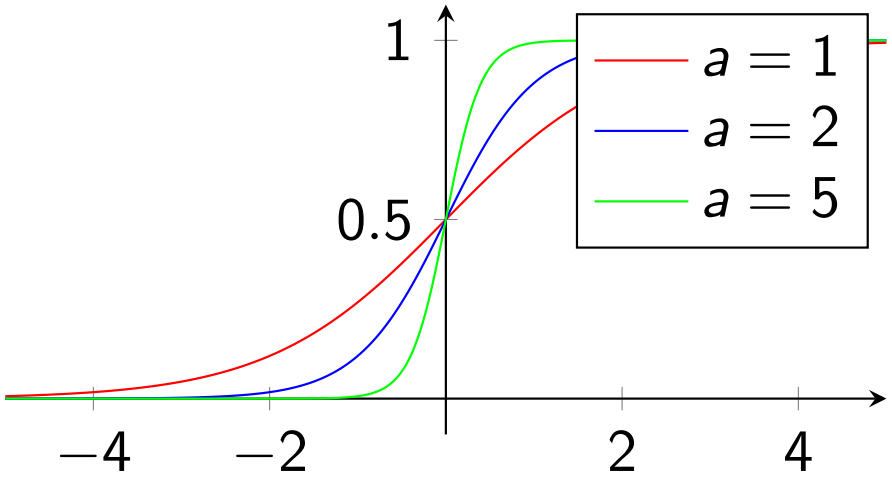


https://www.root.cz/clanky/biologicke-algoritmy-4-neuronove-site

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

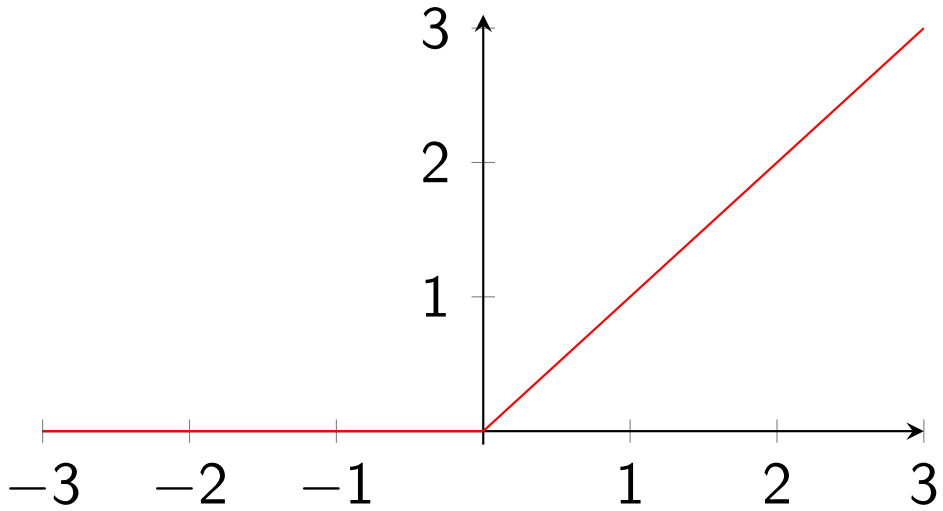
Příklady aktivačních funkcí

Sigmoid



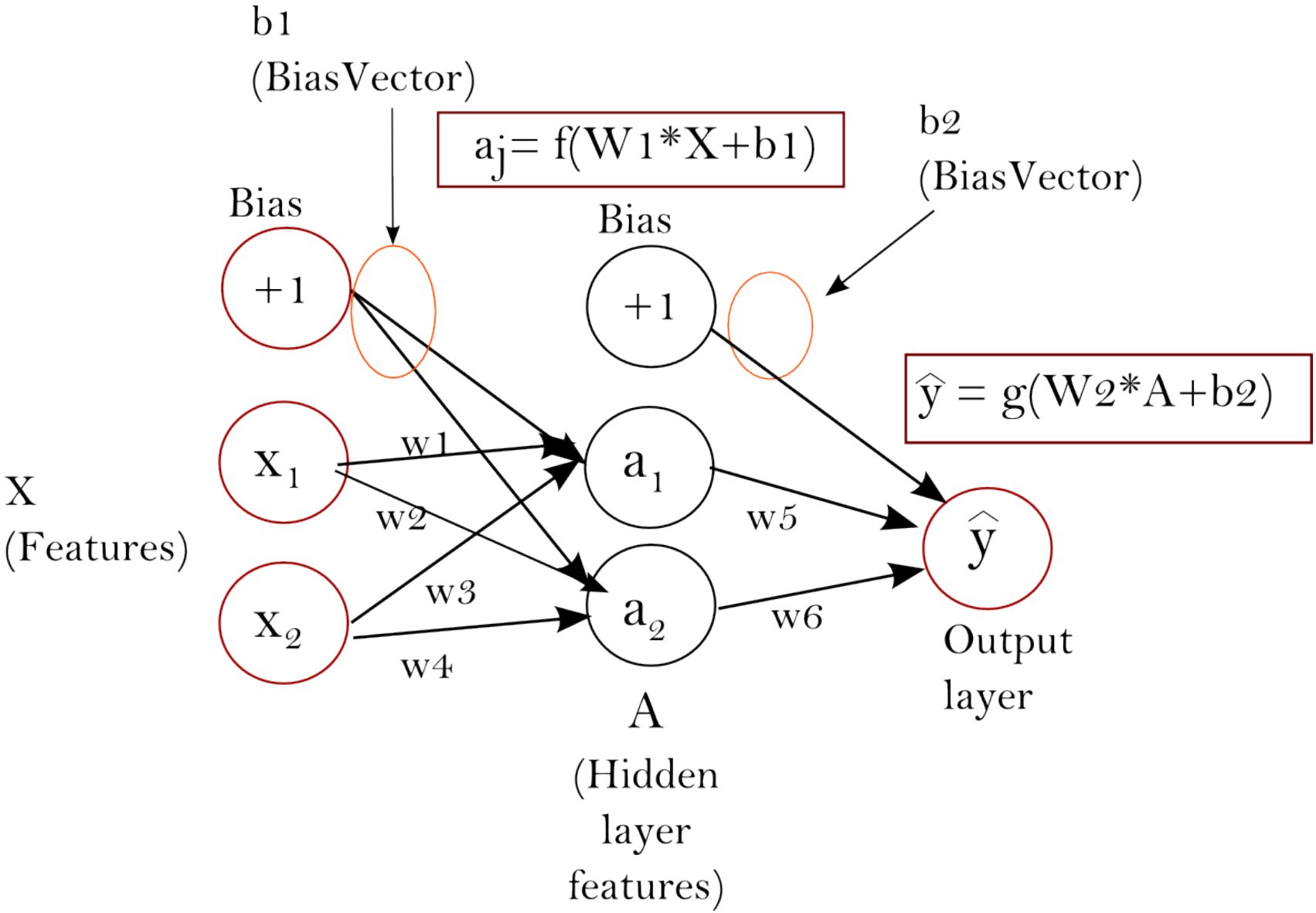
$$f(\Sigma) = \frac{1}{1 + \exp(-a\Sigma)}$$

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

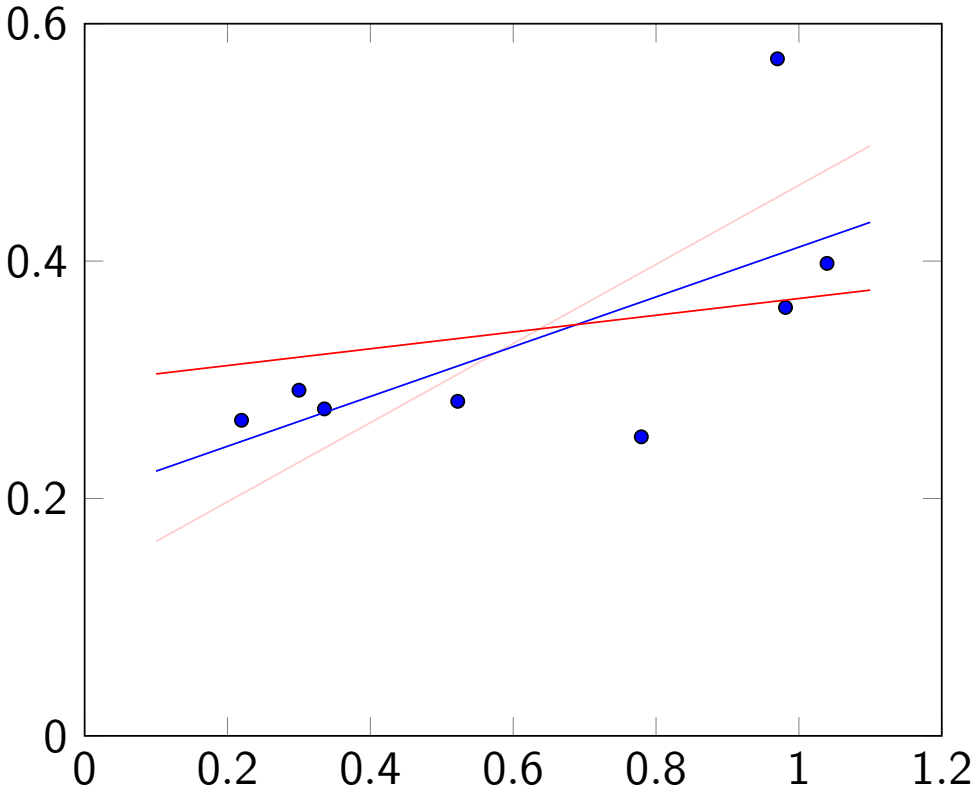
$$\begin{aligned} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i + nb &= \sum_{i=1}^{n} y_i \end{aligned}$$

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

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

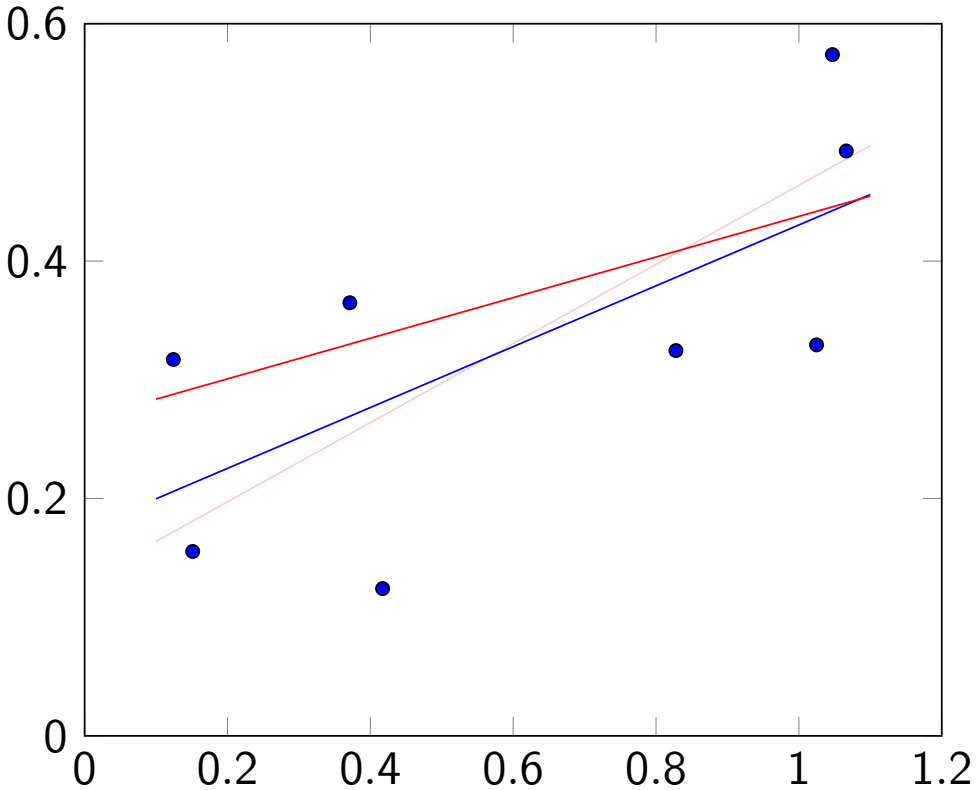
Metoda nejmenších čtverců – gradient descent method

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.



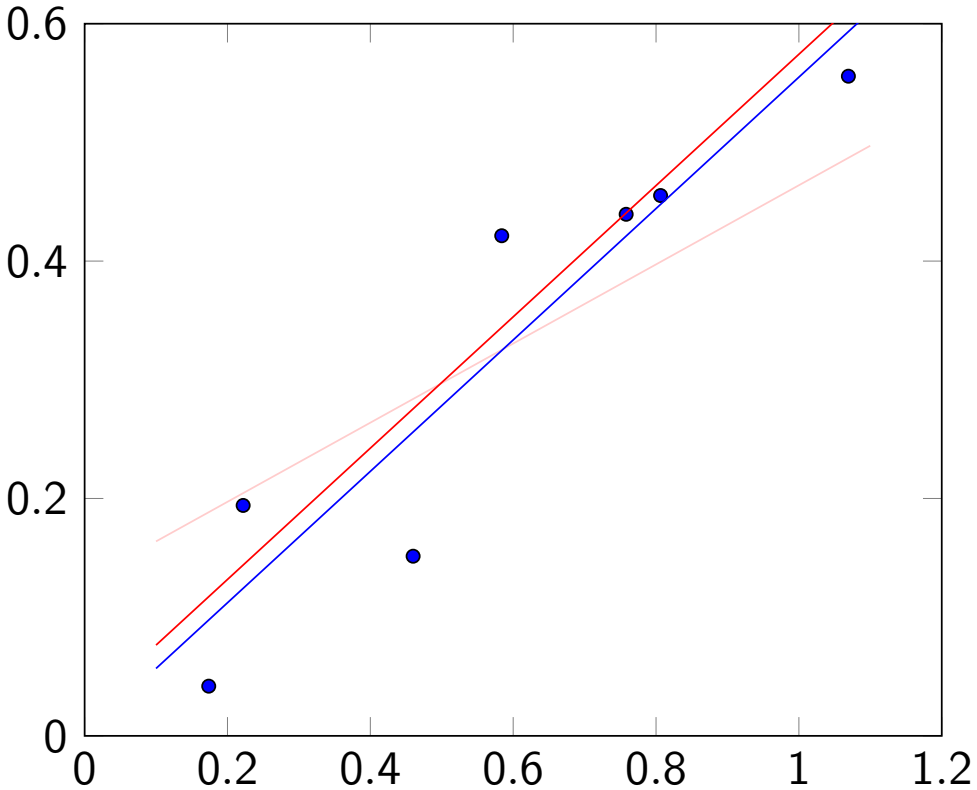
Metoda nejmenších čtverců – gradient descent method

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.



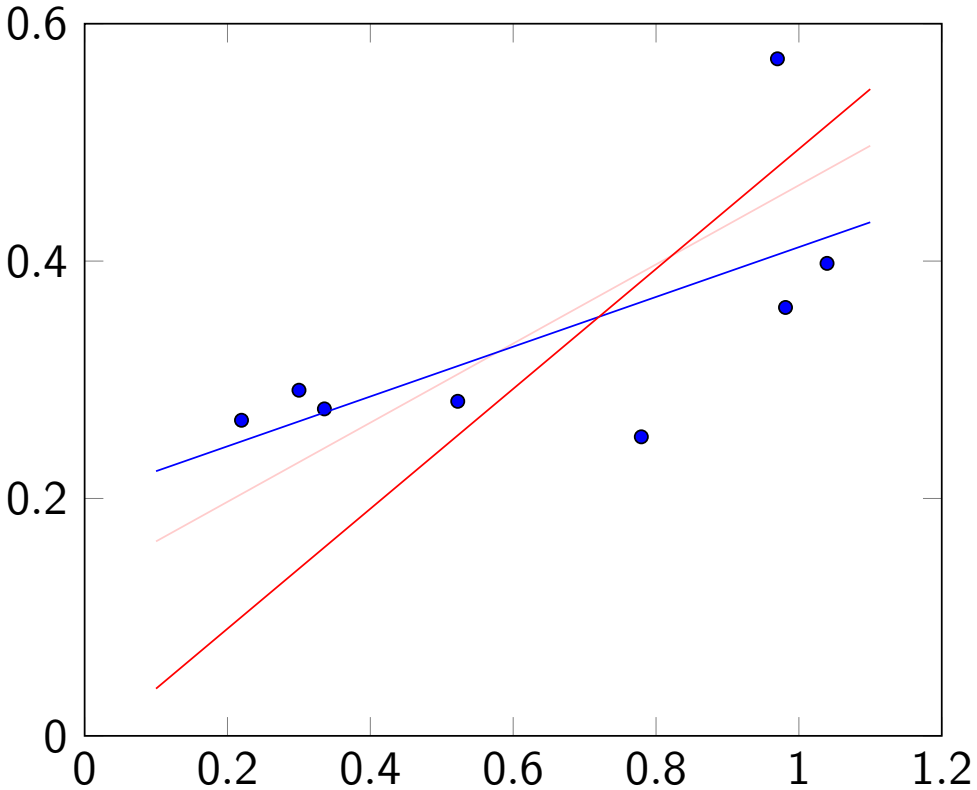
Metoda nejmenších čtverců – gradient descent method

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.



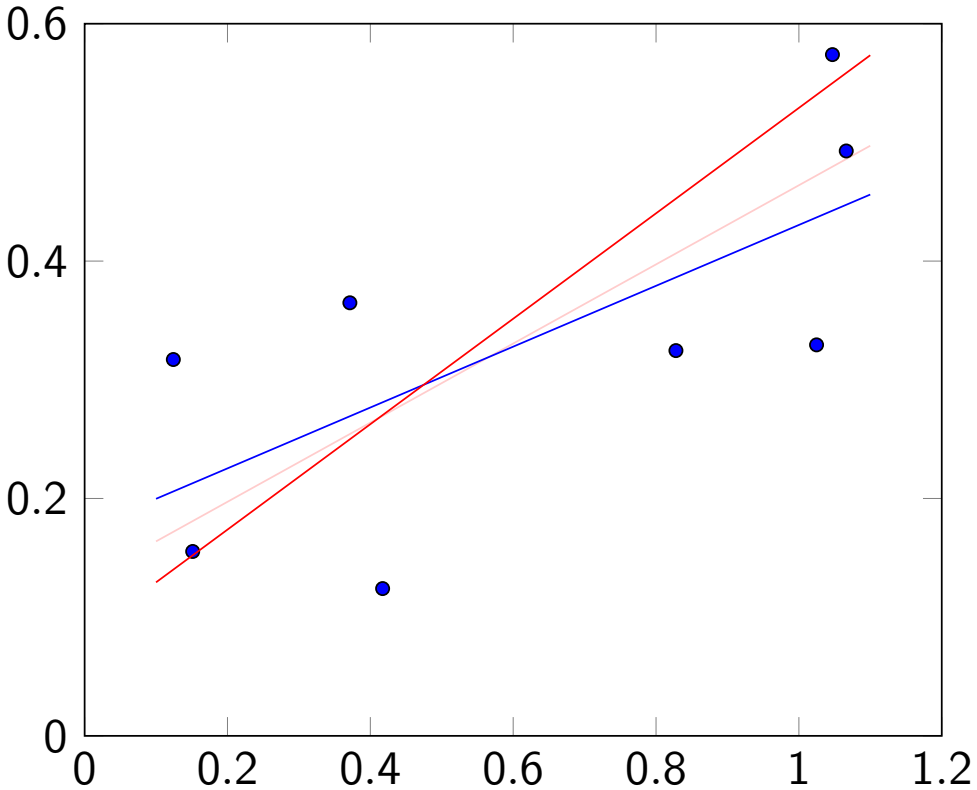
Metoda nejmenších čtverců – gradient descent method

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.



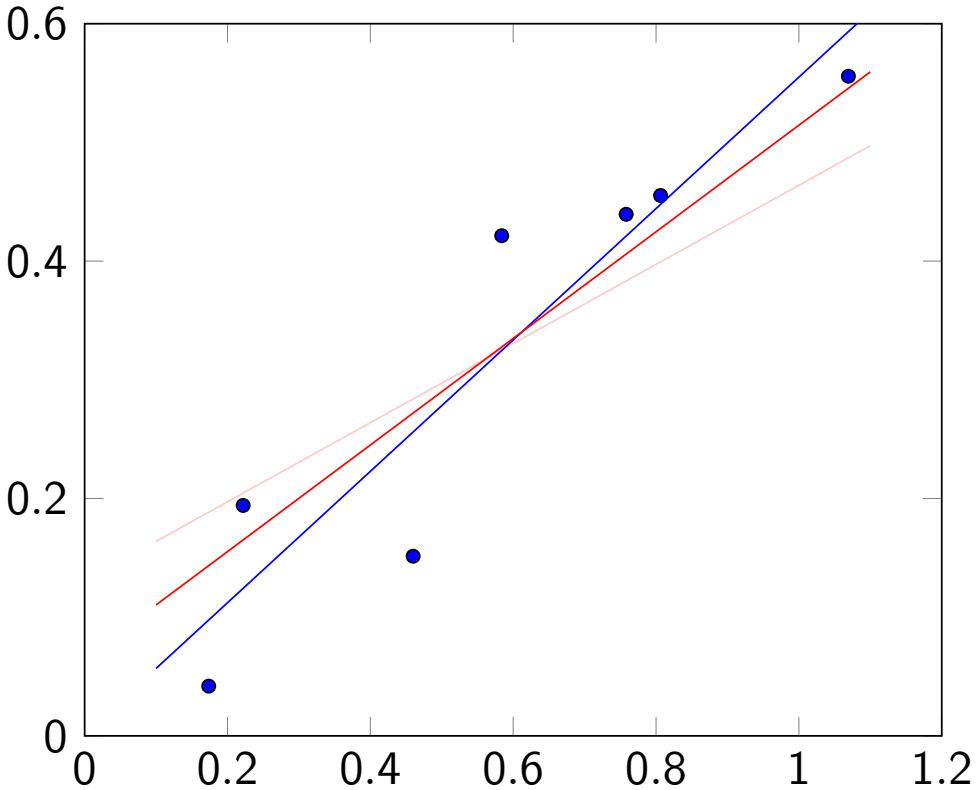
Metoda nejmenších čtverců – gradient descent method

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.



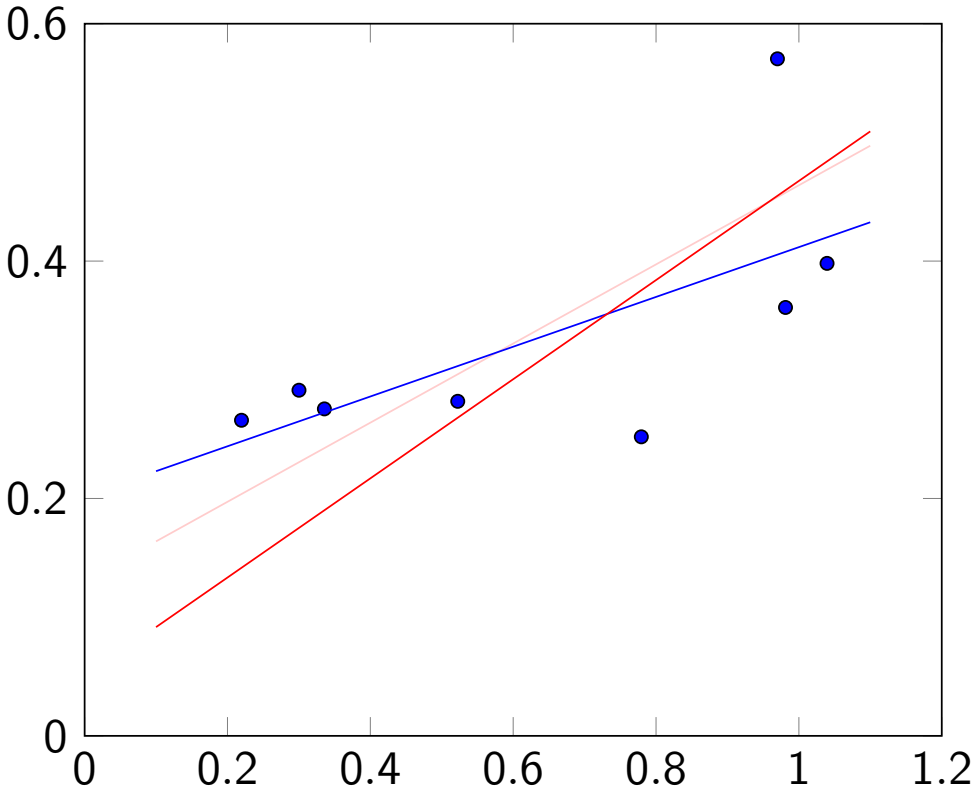
Metoda nejmenších čtverců – gradient descent method

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.



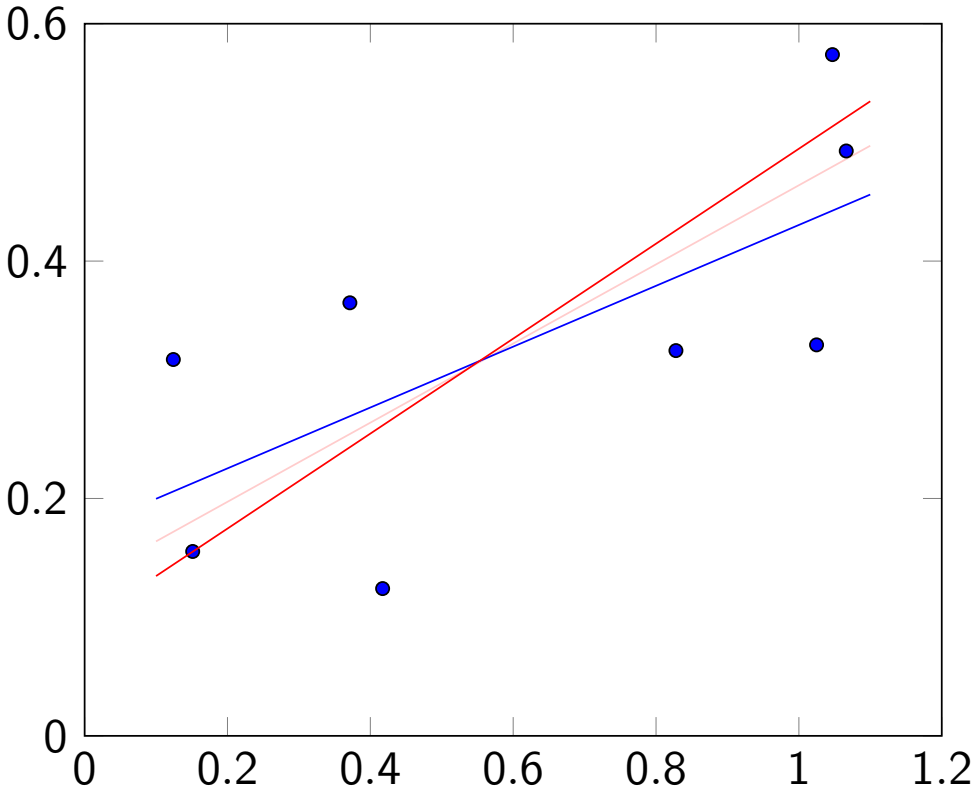
Metoda nejmenších čtverců – gradient descent method

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.



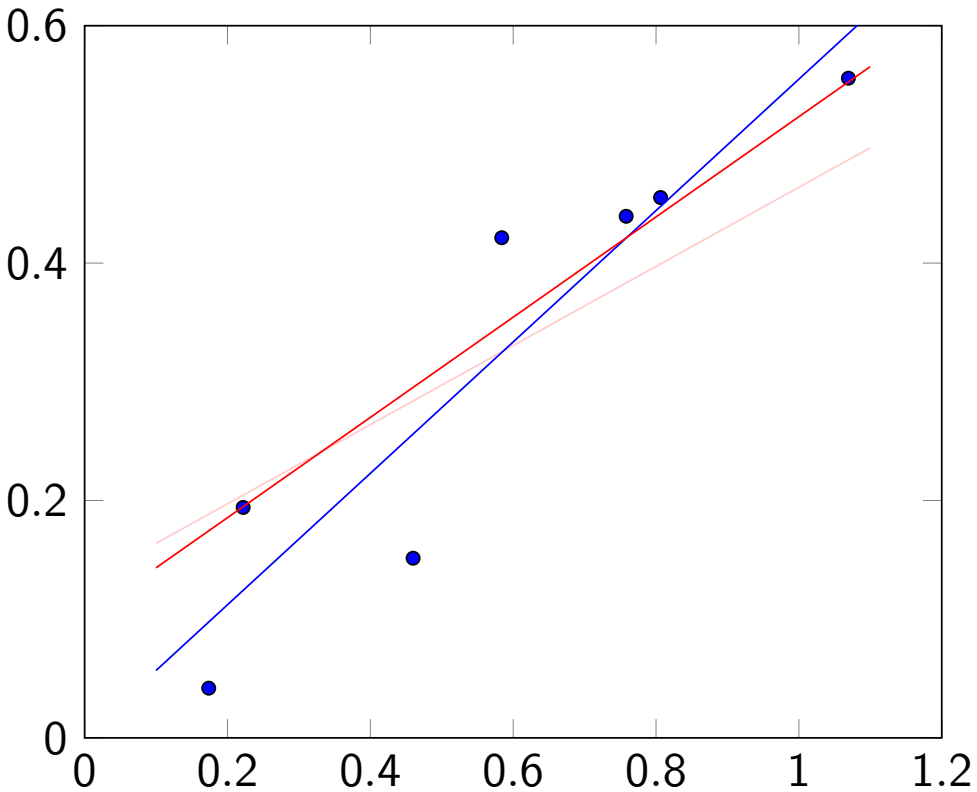
Metoda nejmenších čtverců – gradient descent method

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.



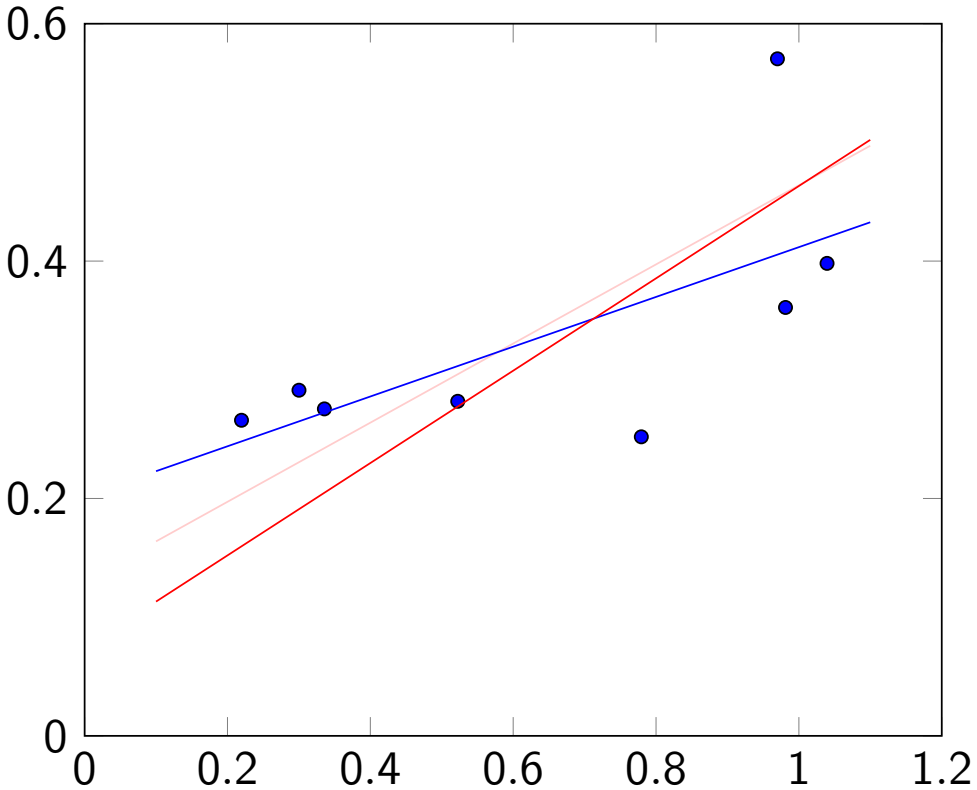
Metoda nejmenších čtverců – gradient descent method

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.



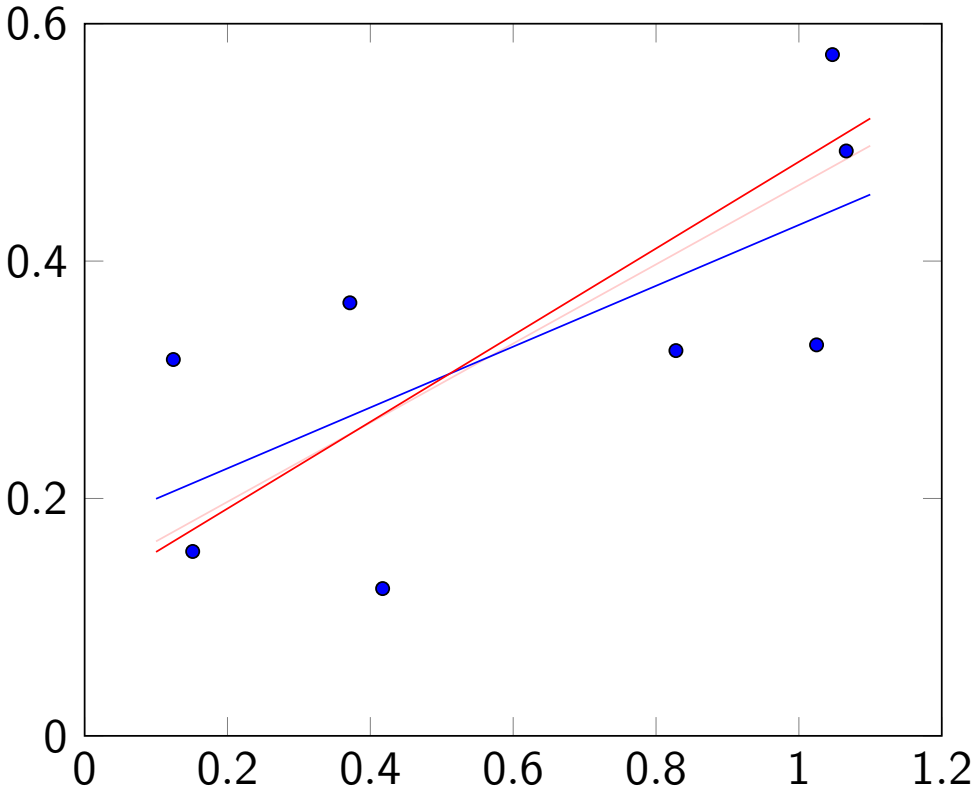
Metoda nejmenších čtverců – gradient descent method

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.



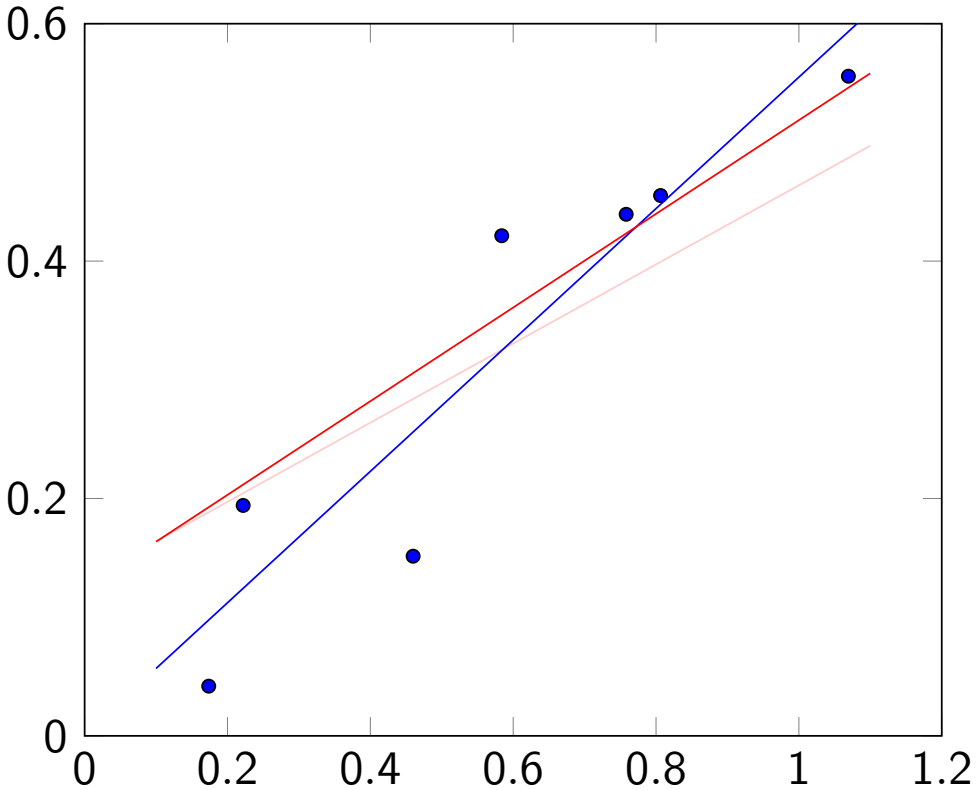
Metoda nejmenších čtverců – gradient descent method

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.

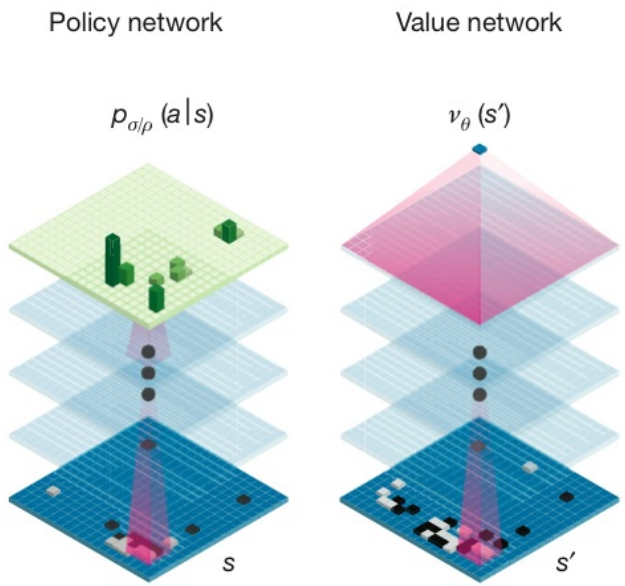


Deskové hry

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.

Video (necelých pět minut) o jednom neočekávaném tahu robota.



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.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. 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.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. 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.“ (necelá minuta)
`https://www.youtube.com/watch?v=aZ5EsdnpLMI&t=69s`

`https://www.youtube.com/watch?v=aZ5EsdnpLMI&t=381s`
(asi pět minut)

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.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. 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.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. 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.

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.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. Další zdroje.

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

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.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. Další zdroje.

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.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. 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?

Jak vznikají chatboti

Jak vznikají jazykové modely (necelé tři minuty)
https://www.youtube.com/watch?v=aZ5EsdnpLMI&t=1086s

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.

Tomáš Mikolov o obecné umělé inteligenci (v podkastu brain we are)
`https://www.youtube.com/watch?v=g64dXIU-REQ&t=3123s`
(Nebojí se personalizace, možná si nepřipouští možnost zneužití. Lidskou obecnou inteligenci chápe ne z pohledu člověka, ale života a evoluce.)
`https://www.youtube.com/watch?v=g64dXIU-REQ&t=3658s`
(O modelování evoluce)

Tomáš Mikolov v Dejvickém divadle o agi a o tom, že biosféra je inteligentnější než lidé.
`https://www.youtube.com/watch?v=EtZPzKeC7XQ&t=6847s`

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`

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.

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.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. 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.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. Další zdroje.

Harari, rozhovor v deníku N

https://denikn.cz/1615227/
nemeli-bychom-se-bat-terminatoru-ale-ai-byrokratu-rika-histor
?ref=inc&cst=
2b2bfd880bb2c8ed2999c0c4b757b3e73139c2d065d9b7a0c3b64dbd8810fe

Harari, rozhovor v deníku N

Zatím jsme ještě nic moc neviděli. Uvědomme si, že dnešní umělé inteligence jako Chat GPT jsou mimořádně primitivní. Tyto umělé inteligence se budou pravděpodobně vyvíjet ještě desítky let, staletí, tisíciletí a miliony dalších let.
Na začátku jsem mluvil o organické evoluci od jednobuněčných organismů jako améby k mnohobuněčným organismům, jako jsou dinosauři, savci a lidé. Tohle evoluci trvalo miliardy let. Umělá inteligence je nyní v podstatě ve stadiu améby. Ale nebude jí trvat miliardy let, než se dostane do stadia dinosaurů. Může to trvat jen dvacet let, protože digitální evoluce je mnohem, mnohem rychlejší než organická evoluce.
Pokud je Chat GPT améba, jak by podle vás vypadala umělá inteligence ve verzi tyranosaurus rex? Přemýšlejte o tom hodně, hodně vážně, protože s tyranosauří AI se nejspíš setkáme v roce 2040 nebo 2050, tedy ještě za života většiny lidí, kteří čtou tento rozhovor.
V případě umělé inteligence nemůžeme z definice předem naplánovat a předvídat, co všechno udělá. Pokud lze předvídat vše, co udělá, pak se nejedná o umělou inteligenci.

Harari, rozhovor v deníku N

Nevidíme už nyní rozsah a dosah dopadu umělé inteligence, i když je ještě v plenkách?
Rozhodně. Nemluvíme jen o budoucnosti, ale i o tom, co už umělá inteligence způsobila. Již jsme na vlastní oči viděli přinejmenším jednu velkou katastrofu, kdy sociální média řízená algoritmy destabilizovala demokracie a společnosti po celém světě. Tohle byla taková první ochutnávka toho, co se stane, když do světa vypustíte agenta, který se rozhoduje sám.
Algoritmy sociálních médií, používané na Twitteru/X, YouTube nebo Facebooku, jsou neskutečně primitivní. Ačkoli se jedná pouze o první generaci, tyto umělé inteligence měly obrovský dopad na dějiny. Tyto algoritmy sociálních médií dostaly za úkol zvýšit zapojení uživatelů, aby více lidí trávilo více času na Facebooku, více času na YouTube, více času na Twitteru. Co by se tak mohlo pokazit? Upoutání pozornosti a aktivita jsou přece fajn, ne? ...

Harari, rozhovor v deníku N

. . . Omyl, protože umělá inteligence zjistila, že nejsnáze zvýší zapojení uživatelů šířením nenávisti, strachu a chamtivosti, protože právě tohle lidi zajímá, máme to v povaze. Když lidem v hlavě zmáčknete tlačítko nenávisti, přiková je to k obrazovce. Na platformách sociálních médií zůstávají déle a algoritmy mohou rychle umisťovat reklamy, dokud mají vaši pozornost.
Nikdo nedal umělým inteligencím pokyn, aby šířily nenávist a pobouření. Mark Zuckerberg ani další lidé, kteří řídí Facebook nebo YouTube, neměli v úmyslu záměrně šířit nenávist. Dali těmto algoritmům moc a algoritmy udělaly něco neočekávaného a nepředvídaného, protože jsou AI. To už umělé inteligence dělávají.
Škoda není v budoucnosti. Je již v minulosti. Americká demokracie je nyní v ohrožení kvůli tomu, jak tyto extrémně primitivní AI rozštěpily naše společnosti.
Jen si představte, co napáchají sofistikovanější modely AI za deset nebo dvacet let – nebude to menší škoda, než když se z améb stali dinosauři.

Harari, rozhovor v deníku N

Národ je něco dobrého, když je řádně vybudován a udržován. Nacionalismus by neměla napájet nenávist. Pokud jste vlastenec, neznamená to, že někoho nenávidíte. Nejde o nenávist k cizincům. Nacionalismus, to je láska. Je to láska ke svým krajanům. Jde o to, že jste součástí sítě milionů lidí, z nichž většinu jste v životě nepotkali, ale přesto vám na nich záleží natolik, že jste například ochotni dát dvacet, čtyřicet, šedesát procent svého příjmu, aby se tito cizinci na druhém konci země mohli těšit dobré zdravotní péči, dosáhli na vzdělání a měli fungující kanalizaci a pitnou vodu. To je vlastenectví.

Harari, rozhovor v deníku N

V demokracii nevnímáte ostatní lidi jako své nepřátele, ale jako své politické soupeře. Říkáte: „Nepřišli zničit mě a můj způsob života. Podle mě se mýlí v navrhovaných politických opatřeních, ale nemyslím si, že by mě nenáviděli. Nemyslím si, že se mi snaží ublížit. No dobře, byl jsem u moci čtyři roky nebo osm let a vyzkoušel jsem řadu politických opatření a kroků a oni teď chtějí vyzkoušet něco jiného. Myslím, že se mýlí. Ale zkusme to a uvidíme, a pokud se po několika letech ukáže, že jejich politika je skutečně dobrá, přiznám, že jsem se mýlil. Oni měli pravdu.“

Harari, rozhovor v deníku N

Uvedu paralelní příklad z právnického světa. Když si představíte nejlepšího právníka ve Spojených státech, bude to svým způsobem autistický génius. Tento člověk může být nesmírně erudovaný a inteligentní ve velmi, velmi úzkém oboru, jako je legislativa týkající se daní z příjmů právnických osob, ale nedokáže už upéct sušenku nebo vyrobit boty. Pokud takového právníka vytrhnete z jeho prostředí a vysadíte ho v savaně, bude bezmocný – slabší než kterýkoli slon nebo lev.
Ale oni nejsou v savaně. Fungují v rámci amerického právního a byrokratického systému. A uvnitř tohoto systému jsou mocnější než všichni lvi na světě dohromady, protože tenhle jediný právník ví, jak zatáhnout za páky informační sítě, a může tak využít obrovskou moc naší byrokracie a našich systémů.
Tuhle moc umělá inteligence získává. Není to tak, že vezmete Chat GPT, hodíte ho do savany a on si tam postaví armádu. Ale když toho chatbota hodíte do bankovního systému, do mediálního systému, má obrovskou moc.

Harari, rozhovor v deníku N

Co lze udělat preventivně, aby se zabránilo šíření algoritmických duchů ve všech systémech navržených člověkem?

Nejdříve pochopit problém a správně ho vnímat – nikoliv jako vzbouřené roboty, ale jako umělou inteligenci, která přebírá moc zevnitř, jak jsme o tom před chvilkou mluvili. My lidé spěcháme s řešením problémů a nakonec řešíme špatné problémy. Než se vrhnete nabízet řešení, chvilku se nad tím problémem zastavte a nejdřív fakt pochopte, v čem spočívá.

Harari, rozhovor v deníku N

Předně mám ve výzkumném týmu čtyři lidi. Takže pokud se chci něco dozvědět o neandrtálcích nebo o umělé inteligenci, pomáhají mi další lidé, kteří se touto problematikou zabývají do hloubky. *Osobně mám informační dietu stejně, jako mají lidé dietu potravinovou.*
Informační dietu mohu každému jen doporučit, protože jsme zaplaveni příliš velkým množstvím informací a většina z nich je balast. Stejně jako si lidé dávají velký pozor na to, co jedí, měli by si dávat velký pozor i na to, jaké informace konzumují a v jakém množství.
Mám tendenci číst dlouhé knihy, a ne krátké tweety. Pokud chci opravdu pochopit, co se děje na Ukrajině nebo co se odehrává třeba v Libanonu, přečtu si o tom v závislosti na tématu několik knih, přijde na to, jestli jde o LLM, nebo o Římskou říši, o historii, o biologii, nebo informatiku.
Také si dávám informační půst. Protože většina informací je balast a informace je potřeba zpracovávat, pouhý přísun dalších informací do hlavy nijak nezaručuje, že budete chytřejší nebo moudřejší. . . .

Harari, rozhovor v deníku N

. . . Jen si zaplníte hlavu balastem.
Takže stejně jako je důležité informace přijímat, potřebujeme také čas na jejich strávení a detoxikaci hlavy. Proto denně dvě hodiny medituju. Každý rok odjíždím na třicet až šedesát dní do ústraní, kde už nepřijímám žádné nové informace. Panuje tam ticho. V meditačním centru s ostatními lidmi ani nemluvíte, jen zpracováváte, trávíte, detoxifikujete všechno, co jste během roku nashromáždili.
Vím, že to většina lidí bude pokládat za extrém. Většina lidí na to navíc nemá čas ani prostředky. *Ale přesto si myslím, že by se každý měl více zamyslet nad svým informačním jídelníčkem a také si alespoň jednou týdně třeba na den od informací odpočinout. Nebo si vyhradit pár hodin denně, kdy už další informace nebude přijímat.*

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.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. Další zdroje.

Další četba z Mindf*ck

Sociální média a internetové platformy nejsou služby - jsou to architektury a infrastruktury. Tím, že své architektury označují za „služby“, snaží se společnosti přenést zodpovědnost na zákazníky, protože odklikli „souhlas“. Ale v žádném jiném sektoru podnikání se takovéto břemeno na zákazníky neklade. Pasažéři leteckých společností nejsou žádáni, aby „odsouhlasili“ konstrukci letadel, po hotelových hostech nikdo nežádá, aby „odsouhlasili“ počet únikových východů z budovy, nikdo po nás nechce, abychom „odsouhlasili“ čistotu vody, kterou nám servírují k pití. A jako dřívější návštěvník klubů vás můžu ujistit, že když je kapacita baru nebo koncertu překročena a podmínky začnou být evidentně nebezpečné, požární inspektoři nařídí návštěvníkům, aby budovu opustili.

Další četba z Mindf*ck

Historie stavebních norem jde až do roku 64 n. l., kdy Nero po devastujícím požáru Říma, který řádil devět dní, omezil výšku domů, šířku ulic a zásobování obyvatel vodou. I když požár Bostonu v roce 1631 přiměl město zakázat dřevěné komíny a doškové střechy, první moderní stavební normy se objevily až po zničujícím požáru Londýna v roce 1666. Tak jako v Bostonu byly londýnské domy postaveny většinou ze dřeva a měly doškové střechy, takže se požár rychle rozšířil a trval čtyři dny. Schořelo 13 200 domů, 84 kostelů a téměř všechny vládní budovy. V reakci na požár král Karel II. nařídil, že nikdo „nevztyčí dům či budovu, velkou nebo malou, aniž by byla z cihel nebo kamene“. Jeho výnos rovněž nařídil rozšířit průběžné ulice, aby oheň nemohl přeskočit z jedné strany na druhou. Po dalších velkých požárech v devatenáctém století následovala příklad Londýna mnohá další města, až byli nakonec jmenováni inspektoři s pravomocí soukromé stavby prohlížet a vydávat úřední souhlas, že jsou bezpečné pro jejich obyvatele i pro veřejnost. . . .

Další četba z Mindf*ck

. . . Objevila se nová pravidla a nakonec se termíny „bezpečnostní předpisy a normy“ staly všezahrnujícím principem, který dokázal zabránit nebezpečným nebo nepovoleným projektům, bez ohledu na přání majitelů pozemků nebo na souhlas obyvatel. Platforma jako Facebook zažívá svoje požáry už řadu let – Cambridge Analytica, ruské vměšování, myanmarská etnická čistka, hromadná střelba na Novém Zélandě – a tak jako následovaly změny po Velkém požáru Londýna, musíme dohlédnout dál než politici a zabývat se problémy digitální architektury, která ohrožuje sociální harmonii a blaho občanů.

Další četba z Mindf*ck

Internet zahrnuje nespočet různých druhů architektur, s nimiž lidé interagují každý den, někdy i několikrát denně. A jak se digitální a fyzický svět prolínají, tyto digitální architektury mají stále větší a větší vliv na naše životy. *Právo na soukromí* je fundamentální lidské právo a jako takové by mělo být hodnoceno. Jenomže soukromí je příliš často narušeno pouhým kliknutím na „souhlasím“ pod nekonečným seznamem všeobecných podmínek. Tento jakoby souhlas neustále umožňuje velkým technologickým platformám obhajovat svoje manipulativní praktiky neupřímným odvoláním se na „souhlas uživatele“. A nás to staví do pozice, kdy se nezabýváme designem – a designéry – těchto cinknutých architektur, ale neplodně se místo toho soustřeďujeme na aktivitu uživatele, který nechápe, jak byl systém navržen, a ani nemá možnost do ničeho mluvit. Nedáváme lidem možnost „vybrat“ si budovy, které mají vadnou elektroinstalaci nebo nemají nouzový východ. Bylo by to nebezpečné – a žádné podmínky připíchnuté na dveřích nepomohou architektovi, aby dostal svolení nebezpečnou stavbu realizovat. Proč by inženýři a architekti softwaru a internetových platforem měli mít výjimku?

Další četba z Mindf*ck

Tak jako v případě tradičních stavebních předpisů by ústředním rysem digitálních stavebních předpisů byl princip *bezpečnosti stavby pro uživatele*. Ten bude vyžadovat, aby platformy a jejich aplikace prošly auditem jejich zneužitelnosti a sadou bezpečnostních testů *předtím*, než bude produkt uvolněn a masově rozšířen. Břemeno důkazu bude spočívat na technologických společnostech – to ony musí dokázat, že jejich produkty jsou bezpečné pro masové použití veřejností. V souladu s tím bude zakázáno využívat veřejnost k experimentům testujícím ve velkém měřítku nové prvky, aby se občané nemohli stát pokusnými králíky. To pomůže zabránit případům jako Myanmar, kdy Facebook předem nestudoval nebezpečí, zda jeho platforma nemůže spolupracovat na zažehnutí násilí v oblasti etnického konfliktu.

Další četba z Mindf*ck

Do našich životů dnes ve velkém měřítku proniká umělá inteligence, digitální ekosystémy a software vůbec, a přitom ti, kdo tyto každodenně používané přístroje a programy vyrábějí, nejsou povinni dodržovat žádné zákony nebo vynutitelné normy, které by je přiměly pečlivě zvážit etické dopady na uživatele nebo i celou společnost. Softwarové inženýrství jako profese má vážné etické problémy, které je třeba řešit. Technologické společnosti své problematické a nebezpečné platformy nekouzlí ze vzduchu – mají zaměstnance, kteří je konstruují. A problém je v tom, že softwaroví inženýři a datoví analytici nenesou za své produkty žádnou osobní zodpovědnost. Když zaměstnavatel požádá svého inženýra, aby vytvořil systémy, které manipulují lidmi, jsou eticky pochybné nebo jsou ledabyle implementované a neberou v potaz bezpečí uživatele, neexistují žádné předpisy opravňující inženýra, aby takovou práci odmítl. Za současného stavu by takovým odmítnutím riskoval postih nebo i propuštění. . . .

Další četba z Mindf*ck

. . . I kdyby se později prokázalo, že design je neetický a odporuje zákonům, společnost stejně absorbuje případný postih, zaplatí pokuty a lidé, kteří danou technologii vyrobili, neponesou žádné následky – na rozdíl od lékaře nebo právníka, který vážným způsobem poruší etické standardy své profese. To je zvrácený stav věcí, jaký neexistuje v žádné jiné profesi. Kdyby zaměstnavatel požadoval od právníka nebo zdravotní sestry, aby provedli něco neetického, mají povinnost to odmítnout, v opačném případě jim hrozí ztráta licence. Jinými slovy, mají vážné důvody, aby se zaměstnavateli vzepřeli.

Další četba z Mindf*ck

Pokud softwaroví inženýři a datoví analytici mají být profesionály, kteří jsou hodni své pověsti a vysokých platů, musí tomu odpovídat i příslušná povinnost jednat eticky. Zákony svazující technologické společnosti nebudou zdaleka tak účinné, jak by mohly být, jestliže nebudeme vyžadovat osobní zodpovědnost lidí uvnitř těchto společností. Potřebujeme přimět inženýry, aby se začali zajímat o to, co postavili. Odpolední workshop nebo semestrální kurz o zásadách etického jednání jsou naprosto nedostatečná řešení problémů, které nám předkládají nové technologie. Nemůžeme prodlužovat současný stav technologického paternalismu, v němž šéfové ze slonovinové věže Silicon Valley vytvářejí typ nebezpečných odborníků, kteří se nehodlají zabývat škodami, jež jejich práce může potenciálně způsobit.

Další četba z Mindf*ck

Potřebujeme profesní etický kodex, na jehož plnění bude dohlížet profesní komora, jako to existuje v případě stavebních inženýrů a architektů v mnoha zemích. V těchto předpisech musejí být konkrétní tresty pro softwarové inženýry nebo datové analytiky, kteří využívají svůj talent a znalosti ke konstrukci nebezpečných a neetických technologií. Tento kodex nesmí být formulován volně, ale musí jasně, konkrétně a jednoznačně stanovit, co je přijatelné a co ne. Musí obsahovat povinnost respektovat autonomii uživatele, identifikovat a specifikovat rizika a musí projít recenzním řízením. Tento etický kodex by měl také zahrnovat požadavek vzít v úvahu důsledky provedené práce na zranitelnou část populace, včetně nepřiměřeného účinku na uživatele různých ras, pohlaví, schopností či sexuálních orientací. A když po důkladném zvážení dojde inženýr k závěru, že je žádost jeho zaměstnavatele postavit jistý prvek designu neetická, bude jeho povinností práci odmítnout a celou věc nahlásit, a pokud to neudělá, bude to mít vážné profesní důsledky. Proto musí být odmítnutí práce a hlášení chráněno zákonem před perzekucí ze strany zaměstnavatele.

Další četba z Mindf*ck

Ze všech možných druhů regulace zabrání etický kodex softwarových inženýrů největšímu množství škod, jelikož přímo přinutí stavitele, aby uvažovali o své práci *předtím, než je distribuována* na veřejnosti, a neumožňuje je zbavit se morální zodpovědnosti výmluvou, že stavitelé jen poslouchali příkazy. Technologie je často odrazem hodnot, jež společnost vyznává, takže zavedení etiky do chování technologických společností je životně důležité, pokud má naše společnost stále více záviset na jejích výtvorech. Pokud budou softwaroví inženýři skutečně osobně zodpovědni za svou práci, stanou se naší nejlepší obranou proti budoucímu zneužití technologií. *A budeme-li v roli softwarových inženýrů, měli bychom se všichni snažit, abychom si při vytváření nových architektur důvěru veřejnosti v naši práci zasloužili.*

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.
   - Cambridge Analytica, Christopher Wylie
   - The Social Dilemma, Center for Human Technology
   - Sociální sítě jako bysnysový nástroj
5. Generativní modely mění svět. Tomáš Mikolov, techno optimista.
6. Další četba
   - Juval Noach Harari
   - Mindf*ck, Christopher Wylie
7. 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

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

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

Juval Noach Harari: Nexus (stručné dějiny informačních sítí od doby kamenné k umělé inteligenci)

Juval Noach Harari: rozhovor v deníku N https://denikn.cz/1615227/nemeli-bychom-se-bat-terminatoru-ale-ai-byrokratu-rika-histor?ref=inc&cst=2b2bfd880bb2c8ed2999c0c4b757b3e73139c2d065d9b7a0c3b64dbd8810fe