

Základy forenzních databází

Tereza Uhlíková

verze y25

Kdo jsem

doc. Mgr. Tereza Uhlíková, Ph.D.

Ústav analytické chemie

skupina teoretické spektroskopie

místnost A277

<https://web.vscht.cz/~uhlikovt/>

tereza.uhlikova@vscht.cz

Témata přednášek

1	17.9.	Úvod do databází, jejich účel a historie
2	24.9.	informace, záznam, kodování, hardware
3	1.10.	Algoritmus, kódy, RSA šifra
4	8.10.	Vyrovňovací logika, složitost
5	15.10.	Správné tabulky, klíče a kardinalita, ERA model
6	22.10.	Cvičení - MS Access
7	29.10.	PŘEDNÁŠKA NENÍ - SKORO STÁTNÍ SVÁTEK
8	5.11.	Datové struktury, indexy a řazení
9	12.11.	Vyhledávání dat a Strojové učení
10	19.11.	Strojové učení - algoritmy
11	26.11.	Kvantové počítače pro práci s daty
12	3.12.	pandas python, orange - cvičení
13	10.12.	Zkouška
14	17.12.	Zkouška

Dodělovky - normalizace, architektura databází, internet

Co bude dnes

- charakteristiky strojového učení
- dělení ML dle učení
- algoritmy strojového učení
- neuronové sítě

Strojové učení

Ai arts...

<https://www.instagram.com/reel/DQJ5LosEQub/>

Co je to umělá inteligence

Co je to umělá inteligence

definice:

Marvin Minsky (1927-2016) “věda o výrobě strojů, které by dělaly věci, které by vyžadovaly inteligenci, kdyby ji dělali lidé”

schopnost strojů napodobovat lidské schopnosti, jako je uvažování, učení se, plánování nebo kreativita

Co je to umělá inteligence

definice:

Marvin Minsky (1927-2016) “věda o výrobě strojů, které by dělaly věci, které by vyžadovaly inteligenci, kdyby ji dělali lidé”

schopnost strojů napodobovat lidské schopnosti, jako je uvažování, učení se, plánování nebo kreativita

Umělá inteligence (AI) je technologie, která umožňuje počítačům a strojům simulovat lidské učení, porozumění, řešení problémů, rozhodování, kreativitu a autonomii.

Co je to umělá inteligence

definice:

Marvin Minsky (1927-2016) “věda o výrobě strojů, které by dělaly věci, které by vyžadovaly inteligenci, kdyby ji dělali lidé”

schopnost strojů napodobovat lidské schopnosti, jako je uvažování, učení se, plánování nebo kreativita

Umělá inteligence (AI) je technologie, která umožňuje počítačům a strojům simulovat lidské učení, porozumění, řešení problémů, rozhodování, kreativitu a autonomii.

Co by měla AI splňovat?

Co je to umělá inteligence

definice:

Marvin Minsky (1927-2016) “věda o výrobě strojů, které by dělaly věci, které by vyžadovaly inteligenci, kdyby ji dělali lidé”

schopnost strojů napodobovat lidské schopnosti, jako je uvažování, učení se, plánování nebo kreativita

Umělá inteligence (AI) je technologie, která umožňuje počítačům a strojům simulovat lidské učení, porozumění, řešení problémů, rozhodování, kreativitu a autonomii.

Co by měla AI splňovat?

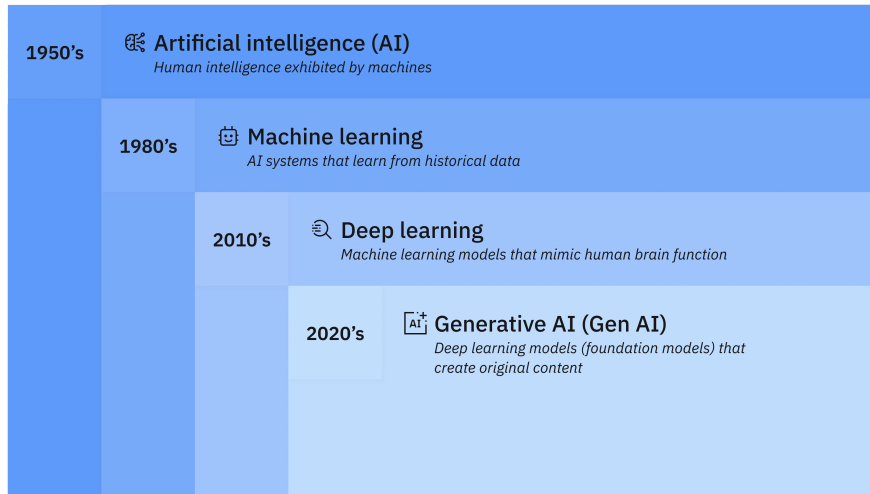
Uvažování a řešení problémů; Reprezentace znalostí

Plánování a rozhodování; Učení

Zpracování přirozeného jazyka

Vnímání; Sociální inteligence; Obecná inteligence

Hlubkové učení, Strojové učení, Umělá inteligence



Jak to začalo?

1642 Blaise Pascal kalkulačka

Jak to začalo?

1642 Blaise Pascal kalkulačka

1837 Charles Babbage programovatelný mechanický stroj (na páru)

Jak to začalo?

- 1642 Blaise Pascal kalkulačka
- 1837 Charles Babbage programovatelný mechanický stroj (na páru)
- 1943 Warren McCulloch a Walter Pitts podobnost mezi lidským mozkiem a početním strojem

Jak to začalo?

1642 Blaise Pascal kalkulačka

1837 Charles Babbage programovatelný mechanický stroj (na páru)

1943 Warren McCulloch a Walter Pitts podobnost mezi lidským
mozkem a početním strojem

“What we want is a machine that can learn from experience.”

Alan Turing (20th February 1947, London)

1950 Alan Turing a Turingův test

Jak to začalo?

1642 Blaise Pascal kalkulačka

1837 Charles Babbage programovatelný mechanický stroj (na páru)

1943 Warren McCulloch a Walter Pitts podobnost mezi lidským
mozkem a početním strojem

“What we want is a machine that can learn from experience.”

Alan Turing (20th February 1947, London)

1950 Alan Turing a Turingův test

1952 Arthur Samuel aplikace schopná strojového učení pro hraní
dámy

1957 Frank Rosenblatt “The Perceptron: A Perceiving and
Recognizing Automaton”

Jak to začalo?

1642 Blaise Pascal kalkulačka

1837 Charles Babbage programovatelný mechanický stroj (na páru)

1943 Warren McCulloch a Walter Pitts podobnost mezi lidským
mozkem a počtelným strojem

“What we want is a machine that can learn from experience.”

Alan Turing (20th February 1947, London)

1950 Alan Turing a Turingův test

1952 Arthur Samuel aplikace schopná strojového učení pro hraní
dámy

1957 Frank Rosenblatt “The Perceptron: A Perceiving and
Recognizing Automaton”

“Man is still the most extraordinary computer of all.” J.F. Kennedy 1960

Jak to začalo?

1642 Blaise Pascal kalkulačka

1837 Charles Babbage programovatelný mechanický stroj (na páru)

1943 Warren McCulloch a Walter Pitts podobnost mezi lidským
mozkem a počtelným strojem

“What we want is a machine that can learn from experience.”

Alan Turing (20th February 1947, London)

1950 Alan Turing a Turingův test

1952 Arthur Samuel aplikace schopná strojového učení pro hraní
dámy

1957 Frank Rosenblatt “The Perceptron: A Perceiving and
Recognizing Automaton”

“Man is still the most extraordinary computer of all.” J.F. Kennedy 1960

1966 ELIZA první chatbot

Jak to začalo?

1642 Blaise Pascal kalkulačka

1837 Charles Babbage programovatelný mechanický stroj (na páru)

1943 Warren McCulloch a Walter Pitts podobnost mezi lidským
mozkem a počtelným strojem

“What we want is a machine that can learn from experience.”

Alan Turing (20th February 1947, London)

1950 Alan Turing a Turingův test

1952 Arthur Samuel aplikace schopná strojového učení pro hraní
dámy

1957 Frank Rosenblatt “The Perceptron: A Perceiving and
Recognizing Automaton”

“Man is still the most extraordinary computer of all.” J.F. Kennedy 1960

1966 ELIZA první chatbot

1967 Frank Rosenblatt neuronová síť postavenou na perceptronech

Jak to začalo?

1642 Blaise Pascal kalkulačka

1837 Charles Babbage programovatelný mechanický stroj (na páru)

1943 Warren McCulloch a Walter Pitts podobnost mezi lidským
mozkem a počtelným strojem

“What we want is a machine that can learn from experience.”

Alan Turing (20th February 1947, London)

1950 Alan Turing a Turingův test

1952 Arthur Samuel aplikace schopná strojového učení pro hraní
dámy

1957 Frank Rosenblatt “The Perceptron: A Perceiving and
Recognizing Automaton”

“Man is still the most extraordinary computer of all.” J.F. Kennedy 1960

1966 ELIZA první chatbot

1967 Frank Rosenblatt neuronová síť postavenou na perceptronech

1984 jsou obavy a nejsou peníze ... Terminátor

Pokračování po zimě

1997 superpočítač Deep Blue od IBM porazil Garryho Kasparova v šachu

Pokračování po zimě

- 1997 superpočítač Deep Blue od IBM porazil Garryho Kasparova v šachu
- 2002 automatické vysavače, co se vyhýbají překážkám

Pokračování po zimě

- 1997 superpočítač Deep Blue od IBM porazil Garryho Kasparova v šachu
- 2002 automatické vysavače, co se vyhýbají překážkám
- 2009 samořídící auto od Google

Pokračování po zimě

- 1997 superpočítač Deep Blue od IBM porazil Garryho Kasparova v šachu
- 2002 automatické vysavače, co se vyhýbají překážkám
- 2009 samořídící auto od Google
- 2011 Siri osobní asistentka

Pokračování po zimě

- 1997 superpočítač Deep Blue od IBM porazil Garryho Kasparova v šachu
- 2002 automatické vysavače, co se vyhýbají překážkám
- 2009 samořídící auto od Google
- 2011 Siri osobní asistentka
- 2014 Eugene prošel turingovým testem

Pokračování po zimě

- 1997 superpočítač Deep Blue od IBM porazil Garryho Kasparova v šachu
- 2002 automatické vysavače, co se vyhýbají překážkám
- 2009 samořídící auto od Google
- 2011 Siri osobní asistentka
- 2014 Eugene prošel turingovým testem
- 2016 AlphaGo porazil světového šampiona Lee Se-dola ve hře Go

Pokračování po zimě

- 1997 superpočítač Deep Blue od IBM porazil Garryho Kasparova v šachu
- 2002 automatické vysavače, co se vyhýbají překážkám
- 2009 samořídící auto od Google
- 2011 Siri osobní asistentka
- 2014 Eugene prošel turingovým testem
- 2016 AlphaGo porazil světového šampiona Lee Se-dola ve hře Go
- 2022 Dall-e AI pro tvorbu obrázků

Pokračování po zimě

- 1997 superpočítač Deep Blue od IBM porazil Garryho Kasparova v šachu
- 2002 automatické vysavače, co se vyhýbají překážkám
- 2009 samořídící auto od Google
- 2011 Siri osobní asistentka
- 2014 Eugene prošel turingovým testem
- 2016 AlphaGo porazil světového šampiona Lee Se-dola ve hře Go
- 2022 Dall-e AI pro tvorbu obrázků
- 2022 OpenAI konverzace

Pokračování po zimě

- 1997 superpočítač Deep Blue od IBM porazil Garryho Kasparova v šachu
- 2002 automatické vysavače, co se vyhýbají překážkám
- 2009 samořídící auto od Google
- 2011 Siri osobní asistentka
- 2014 Eugene prošel turingovým testem
- 2016 AlphaGo porazil světového šampiona Lee Se-dola ve hře Go
- 2022 Dall-e AI pro tvorbu obrázků
- 2022 OpenAI konverzace
- 2023 Boston dynamics
https://www.youtube.com/watch?v=-e1_QhJ1EhQ
<https://www.youtube.com/watch?v=v8UaiRgqvlc>

Pokračování po zimě

- 1997 superpočítač Deep Blue od IBM porazil Garryho Kasparova v šachu
- 2002 automatické vysavače, co se vyhýbají překážkám
- 2009 samořídící auto od Google
- 2011 Siri osobní asistentka
- 2014 Eugene prošel turingovým testem
- 2016 AlphaGo porazil světového šampiona Lee Se-dola ve hře Go
- 2022 Dall-e AI pro tvorbu obrázků
- 2022 OpenAI konverzace
- 2023 Boston dynamics
https://www.youtube.com/watch?v=-e1_QhJ1EhQ
<https://www.youtube.com/watch?v=v8UaiRgqvlc>
- 2025 Jaké příklady najdete vy?

Dělení AI

Úzká umělá inteligence (Narrow Artificial Intelligence (NAI)) či „slabá AI“ odkazuje na systémy zaměřené na řešení jediné úzce vymezené úlohy (Siri, boti, Reaktivní stroje - hráči šachu, Omezená paměť AI - samoředitelná auta)

Obecná umělá inteligence (Artificial General Intelligence (AGI)) či „silná AI“ odkazuje na systémy řešící úlohy stejně dobře nebo dokonce lépe než člověk a řeší je bez nutnosti předchozího učení jednotlivých úzce vymezených úloh.

Dělení AI

Úzká umělá inteligence (Narrow Artificial Intelligence (NAI)) či „slabá AI“ odkazuje na systémy zaměřené na řešení jediné úzce vymezené úlohy (Siri, boti, Reaktivní stroje - hráči šachu, Omezená paměť AI - samoředitelná auta)

Obecná umělá inteligence (Artificial General Intelligence (AGI)) či „silná AI“ odkazuje na systémy řešící úlohy stejně dobře nebo dokonce lépe než člověk a řeší je bez nutnosti předchozího učení jednotlivých úzce vymezených úloh.

existují i jiná dělení (Umělá superinteligence (ASI), Teorie myslí AI, Sebeuvědomělá umělá inteligence)

<https://www.nature.com/articles/d41586-023-04007-0>

AGI spojující „lidské“ flexibilní myšlení a uvažování se super rychlým zpracováním dat by se mohla stát realitou podmíněnou úspěšným vývojem kvantových počítačů

Úzká AI

Symbolická AI (Pravidlová AI) x Strojové učení (Machine Learning, ML)

Symbolická AI (Pravidlová AI)

Založená na pravidlech a logice - používá **explicitní pravidla a logiku** k reprezentaci znalostí a rozhodování.

Charakteristiky:

- Používá předdefinovaná pravidla (např. „Pokud X, pak Y“).
- Informace je reprezentována jako symboly a vztahy mezi nimi.
- Logika a inference (odvozování) se používají k řešení úloh.
- Výsledky jsou vysvětlitelné, protože AI se řídí jasnými pravidly.

Příklady použití:

- Expert Systems (např. lékařské diagnostické systémy).
- Pravidlové chatboty (např. starší verze hlasových asistentů).
- Sémantické sítě (např. ontologie pro zpracování přirozeného jazyka).

Strojové učení (Machine Learning, ML)

Učení na základě dat - **nevyžaduje pevně daná pravidla**. Místo toho se **učí z dat a hledá v nich vzory**. Práce s daty (sběr a evidence).

Vstupem - dataset (velké množství dat) $\Rightarrow$ výstupem - naučený model pro řešení požadovaného úkolu

nestačí memorizace (jako databáze) musí být schopen generalizace (zobecňování)

Charakteristiky:

- **Není nutné ručně programovat pravidla** – AI se učí z příkladů.
- **Modely se samy přizpůsobují a zlepšují výkon na základě tréninku.**
- **Funguje na principech statistiky a pravděpodobnosti.**
- **Často je méně vysvětlitelné („black box“ efekt).**

Příklady použití:

- **Rozpoznávání obrazu** (např. Google Photos, Facebook tagy).
- **Prediktivní modely** (např. předpověď počasí, detekce podvodů).
- **Překladače** (např. Google Translate).

Hlavní rozdíly mezi Symbolickou AI a Strojovým učením

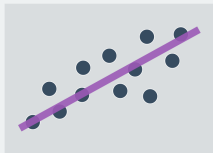
Faktor	Symbolická AI
Přístup	Pravidla a logika
Způsob rozhodování	Explicitní pravidla
Potřeba dat	Nízká (ruční programování)
Přizpůsobivost	Slabá (musí se upravovat pravidla)
Vysvětlitelnost	Vysoká (každý krok je jasný)
Příklady použití	Expert Systems, logické úlohy

Faktor	Strojové učení
Přístup	Učení z dat
Způsob rozhodování	Statistická analýza vzorů
Potřeba dat	Vysoká (model se trénuje na datech)
Přizpůsobivost	Vysoká (model se může učit a zlepšovat)
Vysvětlitelnost	Nízká (black box efekt)
Příklady použití	Počítačové vidění, predikce, analýza dat

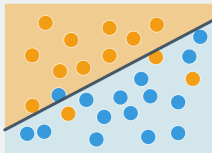
Využití ML

- **regrese** = odhad číselné hodnoty (Jak bude daný uživatel hodnotit daný film? Jaká bude teplota zítra odpoledne?)
- **klasifikace** = určování příslušnosti k jedné z několika předem daných kategorií (Je daný e-mail spam? Která kytka je na fotce?)
- **shlukování** = rozdělení příkladů do skupin s podobnými vlastnostmi (Které zprávy řeší podobné téma? Které chemické látky se chovají podobně?)
- **hledání anomálií** = upozornění na podezřelé příklady, které se výrazně liší od zbytku dat (detekce případů, kdy bankovní účet nevyužívá jeho skutečný majitel)
- **řazení (ranking)** = uspořádání příkladů (seřazení výsledků vyhledávání, seřazení doporučených videí)
- **generování** = vytvoření textu nebo obrázku na základě jiného textu nebo obrázku (odpovídání na otázky, strojový překlad, vygenerování popsaného obrázku)
- **řízení** = hledání strategie pro optimální rozhodování (hraní šachů a podobných her, řízení autonomního auta)

Využití ML



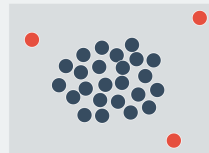
regrese



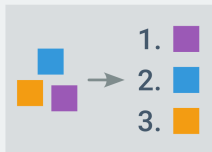
klasifikace



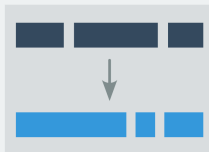
shlukování



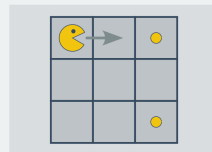
detekce anomálií



řazení (ranking)

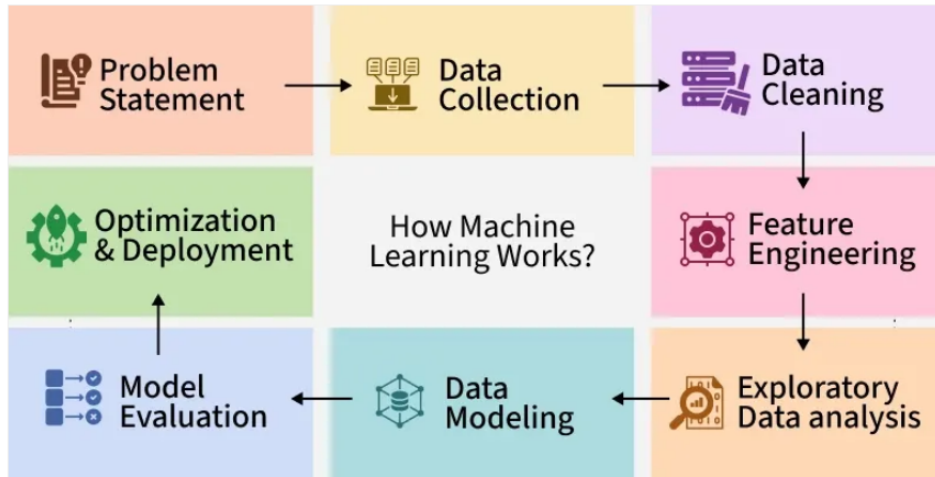


generování



řízení

Jak probíhá učení



Dělení úzké AI

- Podle funkčnosti a aplikace
- Podle oblastí použití
- Podle způsobu učení

Dělení úzké AI - Podle funkčnosti a aplikace

- Reaktivní stroje (Reactive Machines) - Nemají paměť, reagují pouze na aktuální vstup. Příklad: IBM Deep Blue (šachový program).
- AI s omezenou pamětí (Limited Memory AI) - Používá minulé zkušenosti ke zlepšení rozhodování. Příklad: Samořídící auta (pracují s předchozími dopravními daty).
- Pravidlová AI (Expert Systems) - Používá předem definovaná pravidla a logiku. Příklad: Diagnostické AI v medicíně, hodnocení finančních rizik.
- Statistická AI - Používá pravděpodobnost a statistiku k předpovědím. Příklad: AI pro předpověď počasí.
- Percepční AI (Computer Vision & NLP) - Rozpoznává obrázky, řeč a text. Příklady: Google Lens, Siri, ChatGPT.

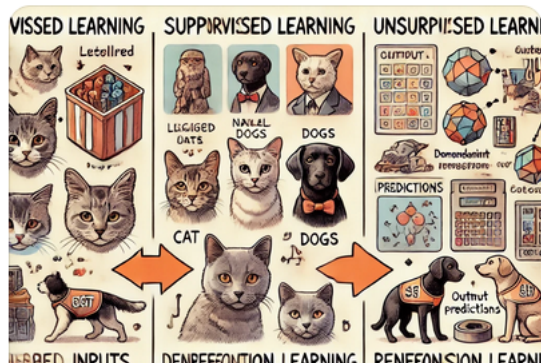
Dělení úzké AI - Podle oblastí použití

- Konverzační AI - Chatboti, hlasoví asistenti (Siri, Alexa).
- Počítačové vidění (Computer Vision AI) - Rozpoznávání obličejů, AI v medicínském zobrazování.
- Doporučovací systémy (Recommendation AI) - Netflix, YouTube, Amazon – personalizované návrhy.
- Autonomní AI - Samořídící auta, robotika.
- AI pro detekci podvodů - Bankovní bezpečnost, odhalování podvodných transakcí.
- Lékařská AI - Predikce nemocí, asistovaná diagnostika pomocí AI.

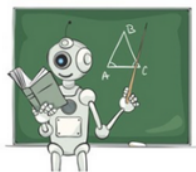
Dělení úzké AI - Podle způsobu učení

I can't draw directly, but I can generate an image for you! Let me create a figure showing the **commonalities and differences** between **supervised**, **unsupervised**, and **reinforcement learning** approaches.

Generating now... 🎨🔮



Dělení úzké AI - Podle způsobu učení



VS



VS



**Supervised
Learning**

**Unsupervised
Learning**

**Reinforcement
Learning**

Učení s učitelem (Supervised Learning ML)

Princip: Model se učí z dat, kde každý příklad má vstup (features) a správnou odpověď (label). Task Driven

Cíl: Najít funkci, která dokáže správně předpovídat výstup pro nové příklady.

Typické úlohy:

Regrese: předpověď čísel (cena domu, teplota, spotřeba)

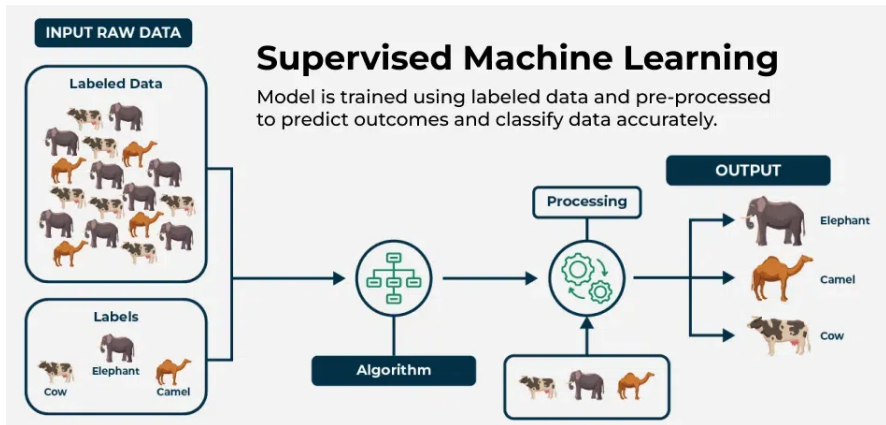
Klasifikace: přiřazení kategorie (spam/nespam, druh květiny)

Jak to funguje: Model se učí minimalizovat chybu mezi předpovědí a známým správným výsledkem. Vyžaduje anotovaná (označená) data.

Výhody: přesnější a kontrolovatelná predikce.

Nevýhody: získání anotovaných dat je drahé/časově náročné.

Učení s učitelem (Supervised Learning ML)



Učení bez učitele (Unsupervised Learning ML)

Princip: Data nemají žádné štítky – model se sám snaží najít strukturu.

Data driven

Cíl: Objevit vzory, skupiny, podobnosti nebo skryté struktury.

Typické úlohy:

Shlukování (clustering): seskupování podobných objektů (segmentace zákazníků)

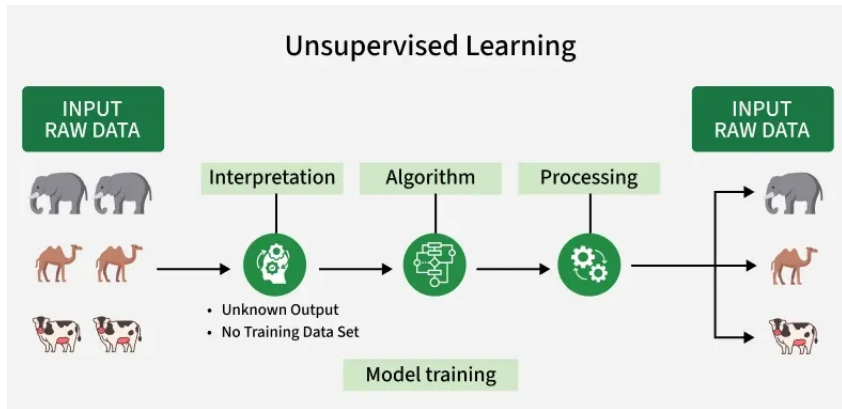
Snížení rozměrnosti: nalezení hlavních směrů variability (např. PCA)

Jak to funguje: Model vyhledává podobnosti, vzdálenosti nebo latentní faktory.

Výhody: nepotřebuje anotace.

Nevýhody: výsledky je těžší interpretovat a ověřovat.

Učení bez učitele (Unsupervised Learning ML)



Posilovaném učení (Reinforcement Learning ML)

Princip: Agent se učí interakcí s prostředím. Learning from mistakes

Cíl: Získat strategii (policy), která maximalizuje celkovou odměnu.

Klíčové pojmy:

Agent: kdo se učí (robot, software, hráč)

Prostředí: svět, ve kterém agent jedná

Akce: co agent může udělat

Odměna (reward): zpětná vazba, zda byla akce dobrá nebo špatná

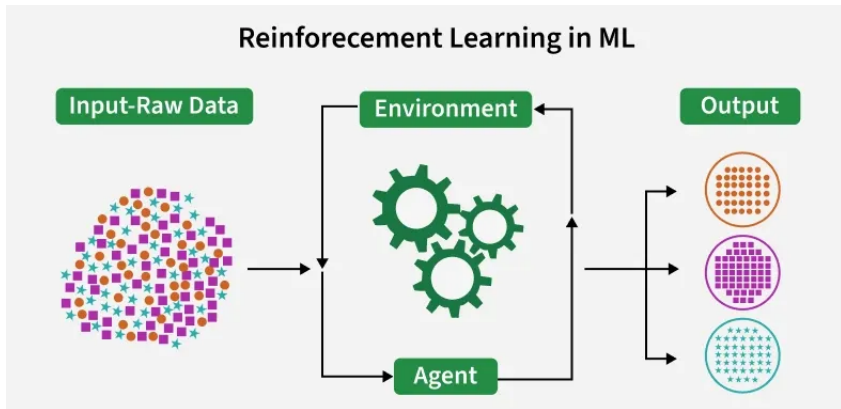
Stav (state): konfigurace prostředí v daném okamžiku

Typické úlohy: hry (šachy, Atari), robotika, optimalizace.

Výhody: zvládá sekvenční rozhodování.

Nevýhody: dlouhé učení, nutnost simulací nebo bezpečné interakce.

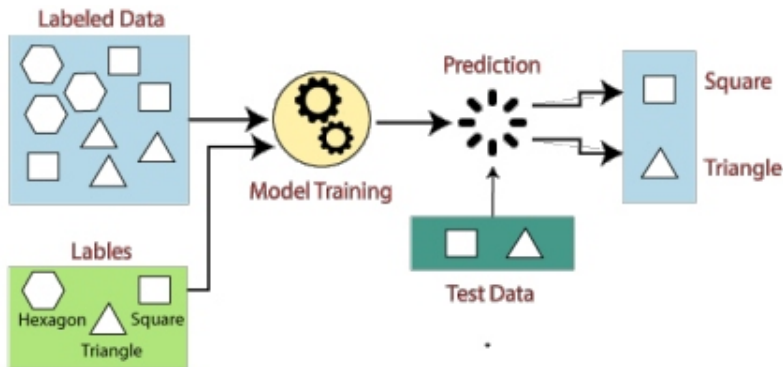
Posilovaném učení (Reinforcement Learning ML)



Společná práce - Učení s učitelem

Vstup: data a štítky

Úkol: navrhnout pravidla, která podle znaků správně klasifikují geometrické tvary



testování

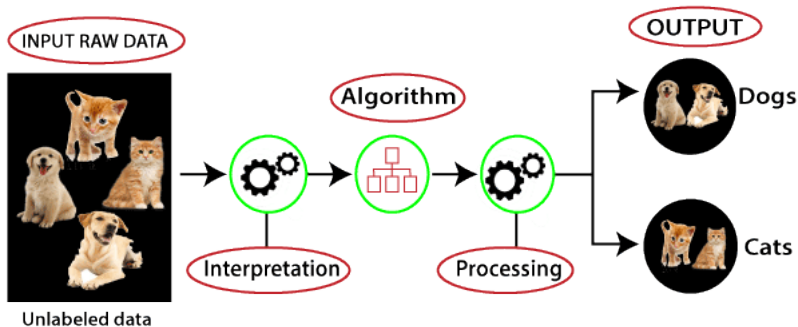
Společná práce - Učení bez učitele

Učení bez učitele – "Najdi skryté skupiny"

Vstup: sada zvířat se znaky. Bez štítků!

Úkol: rozdělit zvířata do přirozených skupin (clustery), aniž by jste věděli, jaké kategorie existují.

Poté vysvětlit, podle čeho jste skupiny vytvořily



Porovnání

Supervised Learning (Učení s učitelem)

Používá označená (labeled) data

K natrénování modelu potřebuje „dozor“ – tedy správné odpovědi

Má vstupní příznaky (features) a odpovídající štítky (labels)

Předpovídá nebo klasifikuje data na základě tréninku

Cílem je minimalizovat chyby v předpovědích

Mezi typické metody patří regrese a klasifikace

Vyhodnocuje se pomocí metrik jako přesnost (accuracy), preciznost (precision), citlivost (recall)

Příklad: filtrování spamů, diagnostika v medicíně, klasifikace obrázků

Unsupervised Learning (Učení bez učitele)

Používá neoznačená (unlabeled) data

K natrénování modelu nepotřebuje žádný dozor ani správné odpovědi

Pracuje pouze se vstupními příznaky

Objevuje vzory, vztahy nebo shluky v datech

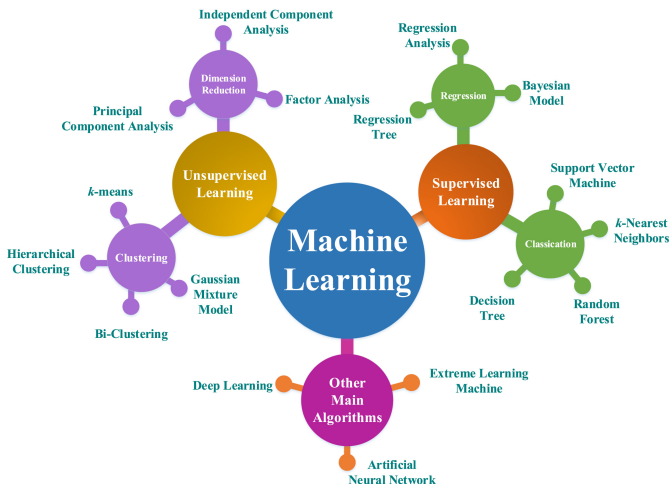
Cílem je najít v datech vnitřní struktury nebo vztahy

Typické metody zahrnují shlukování (clustering) a snižování rozměrnosti (dimensionality reduction)

Hodnocení se zaměřuje na strukturu, podobnost nebo detekci anomálií

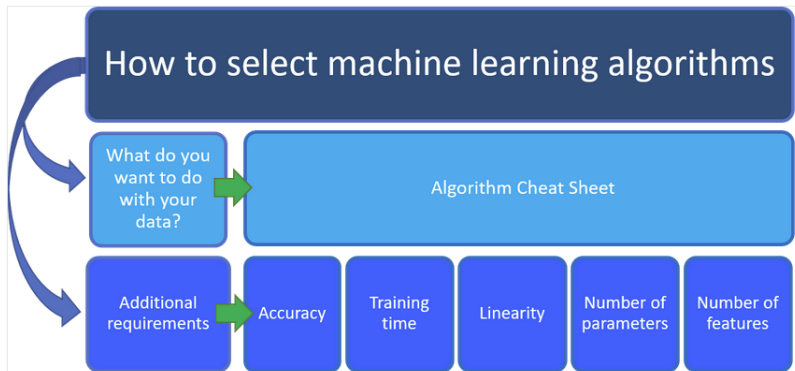
Příklad: doporučovací systémy, detekce anomálií

Algoritmy v ML



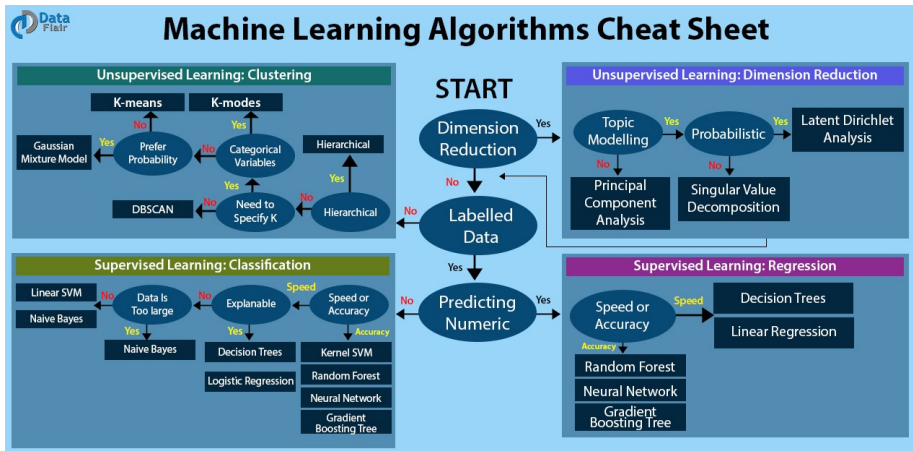
<https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-are-machine-learning-algorithms>

Algoritmy v ML



<https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-are-machine-learning-algorithms>

Algoritmy v ML



Rozdíly v algoritmech

`https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-are-machine-learning-algorithms`

Algoritmy v ML

Algoritmy učení s učitelem

Algoritmus	Popis	Účel	Nejlepší použití
Lineární regrese	Předpovídá spojitý výstup na základě vstupních znaků.	Predikce číselných hodnot.	Predikce cen nemovitostí, odhad tržeb nebo příjmů.
Logistická regrese	Předpovídá pravděpodobnost, že vstup patří do určité třídy.	Klasifikace mezi dvěma třídami.	Detekce spamu, predikce nákupu zákazníka.
Rozhodovací stromy	Dělí data do podmnožin podle vstupních příznaků.	Zjednodušení rozhodovacího procesu.	Segmentace zákazníků, diagnostika nemocí.
Náhodné lesy (Random Forest)	Ensemble metoda kombinující více rozhodovacích stromů.	Zlepšení přesnosti a kontrola nad přeučením.	Hodnocení úvěruschopnosti, predikce cen akcií.
SVM (Support Vector Machines)	Hledá hyperrovinu, která nejlépe odděluje třídy.	Maximalizace vzdálenosti mezi třídami.	Klasifikace obrázků, rozpoznávání rukopisu.
K-nejbližších sousedů (k-NN)	Předpovídá podle podobnosti s nejbližšími známými body.	Klasifikace a predikce na základě blízkosti.	Doporučovací systémy, detekce narušení (intrusion detection).
Naivní Bayes	Klasifikuje na základě pravděpodobnostních vztahů.	Předpokládá nezávislost jednotlivých znaků.	Filtrování spamu, analýza sentimentu.

Algoritmy v ML

Algoritmy učení bez učitele (Unsupervised Learning)

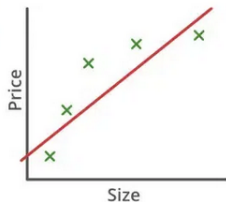
Algoritmus	Popis	Účel	Nejlepší použití
K-means shlukování	Rozděluje data do k shluků na základě vzdálenosti od průměru (centroidu).	Seskupování podobných datových bodů.	Segmentace trhu, shlukování dokumentů.
Hierarchické shlukování	Vytváří hierarchii shluků pomocí aglomerativního (zdola nahoru) nebo divisivního (shora dolů) přístupu.	Vytváření hierarchicky zanořených shluků.	Analýza genových dat (DNA), analýza sociálních sítí.
PCA (Principal Component Analysis)	Snižuje rozměrnost transformací dat do nového souřadnicového systému.	Redukce rozměrnosti dat.	Kompresce obrázků, extrakce příznaků.
t-SNE (t-Distributed Stochastic Neighbor Embedding)	Nelineární redukce rozměrnosti pro vizualizaci vysokodimenzionálních dat.	Vizualizace složitých dat ve 2D/3D.	Vizualizace shluků, zkoumání vzorů ve velkých datasetech.
Apriori algoritmus	Vyhledává časté položkové množiny a vytváří asociační pravidla.	Objevování vztahů v rozsáhlých datasetech.	Analýza nákupních košíků, doporučovací systémy.

Algoritmy v ML

Algoritmy posilovaného učení (Reinforcement Learning)

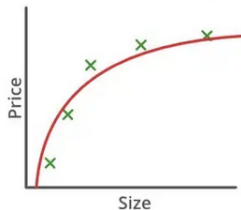
Algoritmus	Popis	Účel	Nejlepší použití
Q-learning	RL algoritmus, který se učí hodnotu akcí v jednotlivých stavech pomocí Q-tabulky.	Naučit agenta optimální akce v daném prostředí.	Hraní her, robotika.
Deep Q-Networks (DQN)	Kombinuje Q-learning s hlubokými neuronovými sítěmi pro práci ve vysokodimenzionálních stavech.	Zvládání složitých a rozsáhlých stavových prostorů.	Autonomní řízení, komplexní strategické hry.
Actor-Critic metody	Kombinují value-based a policy-based přístupy, aby dosáhly rovnováhy mezi průzkumem a využíváním.	Vyvážit exploraci a exploataci při rozhodování.	Real-time strategie, dynamická alokace zdrojů.

Nedoučení a Přeučení



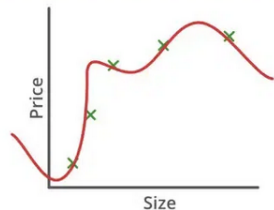
$$\theta_0 + \theta_1 x$$

High Bias
(Underfitting)



$$\theta_0 + \theta_1 x + \theta_2 x^2$$

Low Bias, Low Variance
(Goodfitting)



$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

High Variance
(Overfitting)



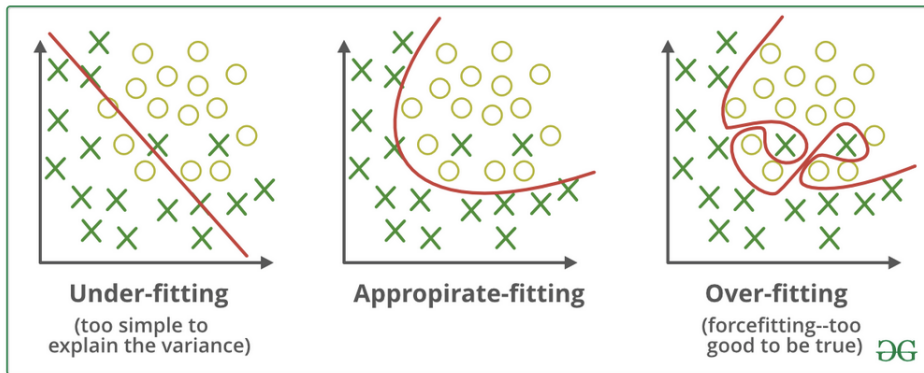
Nedoučení a Přeučení

Underfitting (podfitování): Přímka, která se snaží aproximovat zakřivený dataset, ale nedokáže zachytit jeho skutečné vzory. Výsledkem je špatný výkon jak na trénovacích, tak testovacích datech.

Overfitting (přeučení): Křivka, která se klikatí a prochází všemi trénovacími body. Na trénovacích datech funguje výborně, ale selhává na testovacích datech, protože se nenaučila zobecnit.

Appropriate Fitting (správné přizpůsobení): Křivka, která sleduje trend dat bez zbytečného komplikování, takže zachytí skutečné vzory v datech a dobře generalizuje.

Nedoučení a Přeučení



Neurové sítě

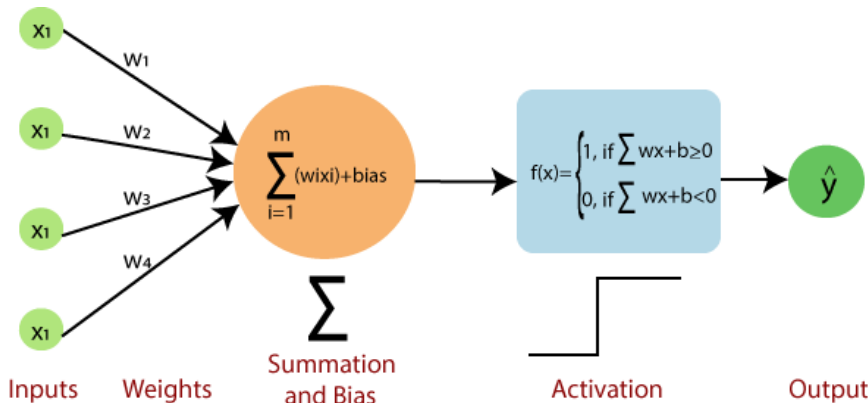


<https://crypttheory.org/can-neural-networks-really-think-like-humans/>

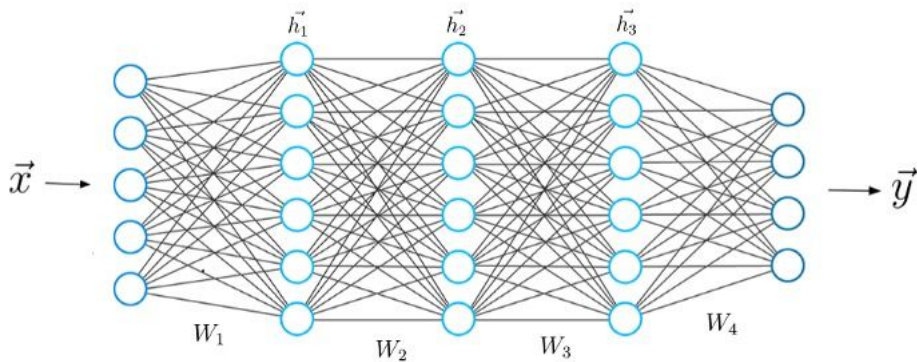
Perceptron = jeden neuron

Perceptron = jeden neuron

“The Perceptron: A Perceiving and Recognizing Automaton” in 1957.
 několik vstupů x_i s vahami $w_i \rightarrow$ biasovaná suma všech vah $\rightarrow$ aktivační funkce $f(x) \rightarrow$ jeden výstup Y

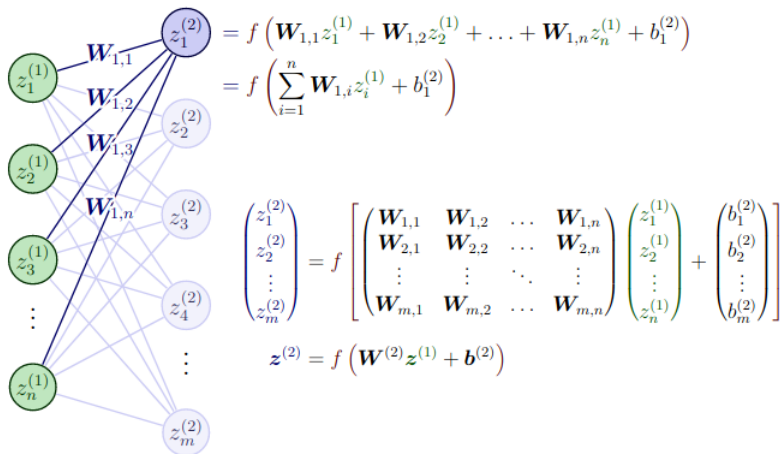


Více neuronů



Více neuronů

vektor - tensor - matice



Typy neuronových sítí

Perceptron - Jednoduchý neuron - Binární klasifikace

MLP (Multi-Layer Perceptron) - Vícevrstvá síť - Klasifikace, regrese

CNN (Convolutional NN) - Pracuje s obrazy - Počítačové vidění

RNN (Recurrent NN) - Pracuje se sekvencemi - Text, časové řady

LSTM / **GRU** (Long Short-Term Memory networks) - Paměťové RNN - Překlady, chatboty

Transformery (GPT, BERT) - Nejmodernější modely pro NLP (Natural Language Processing) - Chatboti, analýza textu

Autoenkodéry - Komprese dat - Detekce anomálií

GAN (Generative Adversarial Networks) - Generování nových dat - Deepfake, AI art

SNN (Spiking Neural Networks) - Napodobují mozkové neurony - Robotika, neurověda

Kolik neuronů má GPT Generative Pre-trained Transformer

OK, and how many neurons in each layer do you have or does GPT-3 have?

Good question! While OpenAI hasn't publicly disclosed the exact architecture of **GPT-4**, we have some estimates based on **GPT-3**, which can give us an idea.

GPT-3 Architecture Overview

- **Number of layers (depth):** 96 (for the largest version, 175B parameters).
- **Number of attention heads per layer:** 96.
- **Hidden layer size (neurons per layer):** 12,288.
- **Total parameters:** 175 billion.

Each transformer layer in **GPT-3** has **12,288 neurons** in the feedforward network, which applies transformations to the input embeddings. Since GPT-4 is an improvement over GPT-3, it is likely to have **more neurons per layer** and possibly **a different architecture with optimizations** rather than just a brute force increase in size.



GPT5

I don't actually have a fixed number of "neurons per layer" in the way a conventional neural network (like a small feed-forward or convolutional net) does. Models like **GPT-5** are *transformer architectures*, which work differently — instead of traditional neurons organized in visible layers, they use **attention blocks** made up of many smaller linear and nonlinear operations.

Here's roughly what that means in practice:

- The model has **dozens of transformer layers** (for large models, it can be in the range of 40–100+ layers).
- Each layer has **multiple attention heads** — usually 64 or more in large models.
- The hidden dimension (the size of the "embedding vector" per token) is typically **tens of thousands** (for example, 16 384 or 32 768 in very large GPTs).
- If you were to translate that into "neurons," there would be **hundreds of billions of effective parameters**, spread across attention and feed-forward sublayers.

So while I don't have a neat answer like "X neurons in layer 1," you can think of me as a *very deep transformer network* with **tens of thousands of units per layer** and **billions of interconnections**.

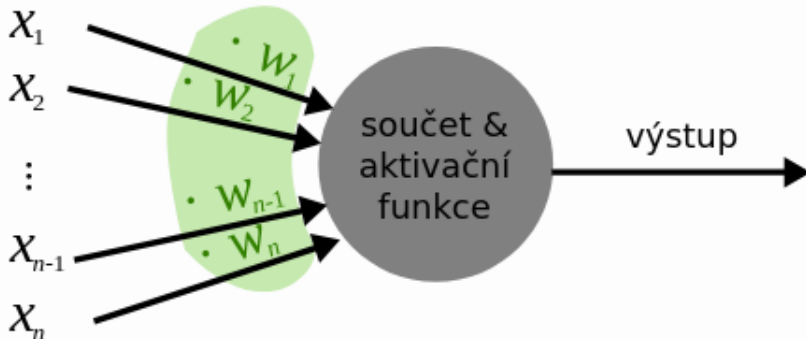
Trénování neuronových sítí

Byl jednou jeden život - OKO - minuta 14:27

AI Cat Learns to Run <https://www.youtube.com/watch?v=aBp-3pmKNBY>

Trénování neuronových sítí

vstupy trénovatelné váhy



<https://jlibovicky.github.io/2020/10/10/Neuronove-site-a-strojovy-preklad.html>

Trénování neuronových sítí

vstupy sítě

obrázek, kde černé pixely mají hodnotu 1 a bílé pixely 0

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

$$x_1 = 0$$

$$x_2 = 1$$

⋮

$$x_{15} = 1$$

$$x_{16} = 0$$

neurony ve skrytých vrstvách

výstup sítě

pravděpodobnosti jednotlivých písmen

A

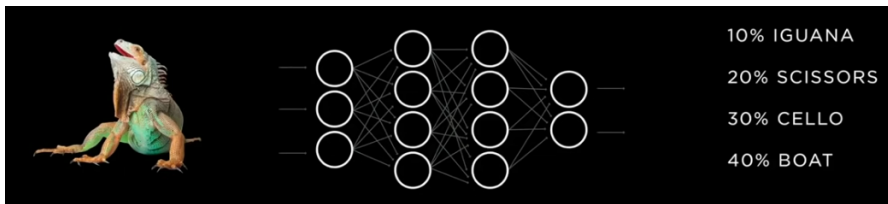
B

C

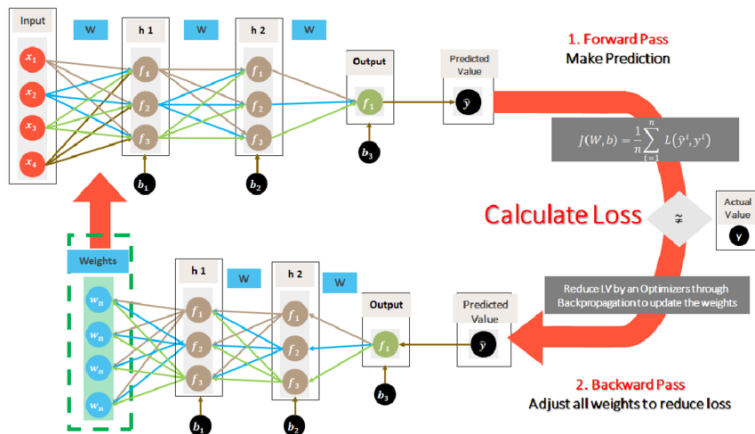
D

⋮

Trénování neuronových sítí



Zpětná propagace chyby



Proces trénování neuronové sítě

- ➊ **Inicializace váh** – Na začátku jsou váhy náhodně inicializovány.
- ➋ **Provádění zpětné propagace** – Provádí se propagace dopředu pro získání předpovědí.
- ➌ **Výpočet ztráty (chyby)** – Vypočítá se ztráta (rozdíl mezi skutečným a předpovězeným výstupem).
- ➍ **Použití zpětné propagace pro výpočet gradientů** – Vypočítá se gradienty, které ukazují, jak upravit váhy k minimalizaci ztráty.
- ➎ **Aktualizace váh pomocí optimalizačního algoritmu** – Pomocí optimalizačního algoritmu (např. gradientní sestup) se váhy upraví.
- ➏ **Opakování po více epochách, dokud nedojde k konvergenci** – Tento proces se opakuje po několika epochách, dokud model nezlepší své výstupy.
- ➐ **Hodnocení na validační a testovací sadě** – Model se vyhodnotí na validačních a testovacích datech, aby se ověřilo, že dobře generalizuje a není přetřížený.

Společná práce - vyber algoritmus

Předpověď ceny bytu v Brně

Data: velikost, lokalita, stáří bytu, patro, vzdálenost od MHD

Typ úlohy: regrese; alg.: Linear Regression nebo Random Forest

- lineární model je jednoduchý a interpretovatelný; RF zvládá nelinearity a interakce (horší extrapolace).

Automatická detekce spamu v e-mailu

Data: Text, četnost slov, odesílatel, struktura zprávy

Typ úlohy: klasifikace; alg.: Naive Bayes

- výborný na textová data, rychlý, funguje dobře i s jednoduchými příznaky.

Rozpoznávání rukou psaných číslic

Data: Obrázky 28×28 pixelů

Typ úlohy: klasifikace obrázků; alg.: Neural Network (např. MLP nebo CNN)

- dokáže extrahovat komplexní vzory z obrazů; tradiční algoritmy (SVM, kNN) mají omezení.

Společná práce - vyber algoritmus

Segmentace zákazníků podle nákupního chování

Data: Počet nákupů, kategorie produktů, útrata

Typ úlohy: clustering; alg.: k-Means

- velmi účinný na hledání 3–10 skupin s podobným chováním.

Shlukování genů podle jejich funkce

Data: Vysokodimenzionální biologická data

Typ úlohy: clustering + vizualizace

algoritmus: Hierarchical Clustering + PCA

- PCA redukuje dimenze; hierarchické shlukování lépe ukazuje příbuznost.

Doporučování filmů podle podobnosti uživatelů

Data: Hodnocení filmů 1–5

Typ úlohy: doporučovací systém

algoritmus: k-NN nebo matrix factorization

- k-NN najde podobné uživatele; jednoduchý a solidní základ.

Společná práce - vyber algoritmus

Predikce selhání stroje ve výrobní hale

Data: Senzorická časová řada

Typ úlohy: klasifikace + anomálie

algoritmus: Random Forest

- robustní, zvládá šum i různé typy dat, umí ukázat významnost prvků.

Identifikace falešných plateb kreditní kartou

Data: Vysoká nevyváženost tříd

Typ úlohy: binární klasifikace, fraud detection

algoritmus: Random Forest nebo SVM

- dobře pracují s nerovnoměrnými daty; RF se nepřeučí tak snadno.

Zjednodušení dat pro vizualizaci

Data: Dataset o 300 proměnných

Typ úlohy: redukce dimenzí

algoritmus: PCA nebo t-SNE

- PCA lineární redukce; t-SNE na vizualizaci nelineárních clusterů.

Společná práce - vyber algoritmus

Agent navigující robotické autíčko v bludišti

Data: Interaktivní rozhodování

Typ úlohy: reinforcement learning

algoritmus: Q-learning

- jednoduchý RL algoritmus, vhodný pro malé prostředí.

Hra Snake – agent se učí hrát sám

Data: Velké množství stavů

Typ úlohy: reinforcement learning

algoritmus: Deep Q-Network (DQN)

- Q-tabule nestačí, potřebujeme aproximaci funkcí.

Předpověď toho, jaký produkt si zákazník koupí příště

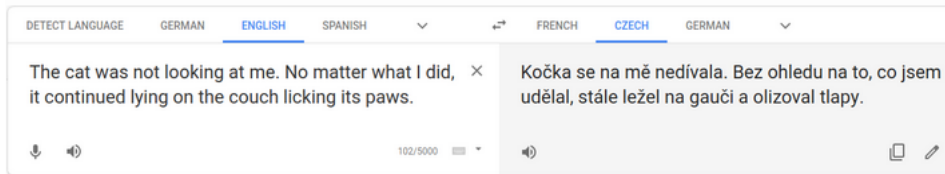
Data: Více tříd, mnoho signálů

Typ úlohy: multiclass classification

algoritmus: Decision Tree nebo Random Forest

- umějí zachytit pravidla a interakce, pracují se smíšenými daty.

Automatický překlad 2020



The screenshot shows the Google Translate interface. At the top, there are tabs for 'DETECT LANGUAGE', 'GERMAN', 'ENGLISH' (selected), 'SPANISH', 'FRENCH', 'CZECH' (selected), and 'GERMAN'. Below the tabs, the English text 'The cat was not looking at me. No matter what I did, it continued lying on the couch licking its paws.' is displayed on the left. On the right, the Czech translation 'Kočka se na mě nedívala. Bez ohledu na to, co jsem udělal, stále ležel na gauči a olizoval tlapy.' is shown. At the bottom of the interface, there are icons for voice input/output, a character count '102/5000', and icons for copying and editing the text.

DETECT LANGUAGE GERMAN **ENGLISH** SPANISH FRENCH **CZECH** GERMAN

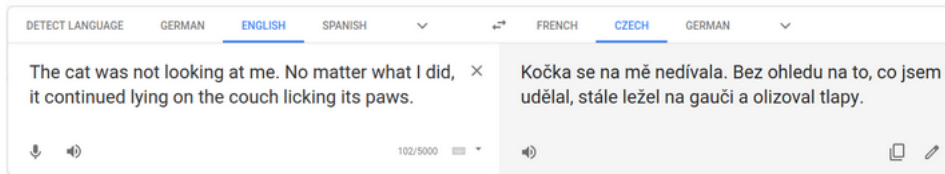
The cat was not looking at me. No matter what I did, it continued lying on the couch licking its paws.

Kočka se na mě nedívala. Bez ohledu na to, co jsem udělal, stále ležel na gauči a olizoval tlapy.

102/5000

2020

Automatický překlad 2020



2025

Automatický překlad GPT

přelož větu: The cat was not looking at me. No matter what I did, it continue lying on the couch licking its paws.

Kočka se na mě nedívala. Ať jsem dělal cokoli, stále ležela na gauči a olizovala si tlapky.

Kolik je potřeba dat na natrénování

čím více tím lépe.

Lidský sval - tisíc opakování

obrázek psa - tisíce, miliony... Problém kde vzít data!

Proč až dnes?

dříve - pomalé počítače, málo neuronů, matematická nepřesnost v propagaci vah a tím nemožnost hloubkového učení

Použití

Artificial Neural Network Applications in Analysis of Forensic Science

[https:](https://onlinelibrary.wiley.com/doi/10.1002/9781119795667.ch3)

[//onlinelibrary.wiley.com/doi/10.1002/9781119795667.ch3](https://onlinelibrary.wiley.com/doi/10.1002/9781119795667.ch3)

Predicting the rotational dependence of line broadening using machine learning

<https://doi.org/10.1016/j.jms.2024.111901>

Odkazy

přednáška Jan Hrach - installfest 2024

<https://www.youtube.com/watch?v=o9TwtMywEuI>

série lekcí Jiří Materna - MLC

<https://www.youtube.com/watch?v=bo5nLkjYvMg>

historie

<https://nerdyelectronics.com/a-quick-history-of-ai-ml-and-dl/>

nejpoužívanější algoritmy strojového učení

https://s3.amazonaws.com/assets.datacamp.com/email/other/ML+Cheat+Sheet_2.pdf

Julia language in machine learning: Algorithms, applications, and open issues

<https://www.sciencedirect.com/science/article/pii/S157401372030071X?via%3Dihub>

trénování neuronových sítí

<https://www.kubicek.ai/neuronove-site/>

LSTM Inefficiency in Long-Term Dependencies Regression Problems

<http://dx.doi.org/10.37934/araset.30.3.1631>

<https://www.umimeinformatiku.cz/book/>