

The background of the slide features a blurred image of a person's hands holding a silver smartphone. In the lower portion, a laptop keyboard is visible. A large, semi-transparent blue rectangle with a white wavy pattern is overlaid on the center of the image, containing the title and author information.

Od perceptrónu po hlbokú neurónovú sieť

Peter Lacko, Fakulta informatiky a informačných technológií, Slovenská technická univerzita v Bratislave

Prečo sú neurónové siete dnes populárne?

- BIG DATA
 - Zhromažďujú sa veľké objemy dát, vytvorila sa infraštruktúra
- GPUs
 - NVidia Tesla P100 GPU dokáže dodať výpočtový výkon 4.7Tflops pričom spotrebuje len 250W elektrickej energie
 - Podľa zoznamu TOP500 by počítač s takýmto výkonom bol najlepším superpočítačom na svete v roku 2000
- Veľké spoločnosti
 - Firmy ako Google, Microsoft, Baidu a Facebook podporujú výskum a vývoj v oblasti umelej inteligencie

<http://www.nvidia.co.uk/object/tesla-p100-uk.html>

<https://www.top500.org/lists/2000/06>

Prístupy učenia

- Učenie s učiteľom (supervised learning)
 - používame množinu označených objektov (vektorov vlastností, čo môžu byť body obrázku, alebo výsledky testov pacienta), ku každému objektu máme pridelenú aj jeho kategóriu, pes, mačka, alebo chrípka, angína.
 - regresia, klasifikácia, predikcia
 - napr. odhaľovanie podvodného správania u zamestnancov a zákazníkov, predikcia vývoja cien

Prístupy učenia

- Učenie bez učiteľa (unsupervised learning)
 - používame množinu objektov, ktorej prvky nemáme predom ohodnotené. Snažíme sa týmto objektom priradiť nejaké ohodnotenie (napríklad priradiť ich do skupiny podobných objektov)
 - zhlukovanie, analýza hlavných komponentov (PCA)
 - napr. segmentácia zákanzík a hľadanie podobných skupín

Prístupy učenia

- Učenie posilňovaním (reinforcement learning)
 - nepoužívame ohodnotené objekty, algoritmus teda nedostáva "správne" riešenie úlohy , ale ohodnocujeme jeho správanie (odmena alebo trest) - teda dávame mu len informáciu, ako dobre robí to čo má
 - napr. získavanie stratégií pri obchodovaní

Riešenie problému strojového učenia

- DÁTA - musíme im rozumieť
- DÁTA - musí ich byť dostatočné množstvo
- DÁTA - musia obsahovať to, čo potrebujeme algoritmus naučiť
- Výber a tréning modelu

Neurónové siete – Perceptrón

- základným stavebným blokom neurónovej siete je neurón a jedným z prvých prístupov učenia bolo použiť perceptrónu
- perceptrón je v najjednoduchšej podobe binárny klasifikátor, ktorý mapuje vstupy x na výstupné hodnoty 0 alebo 1 podľa:

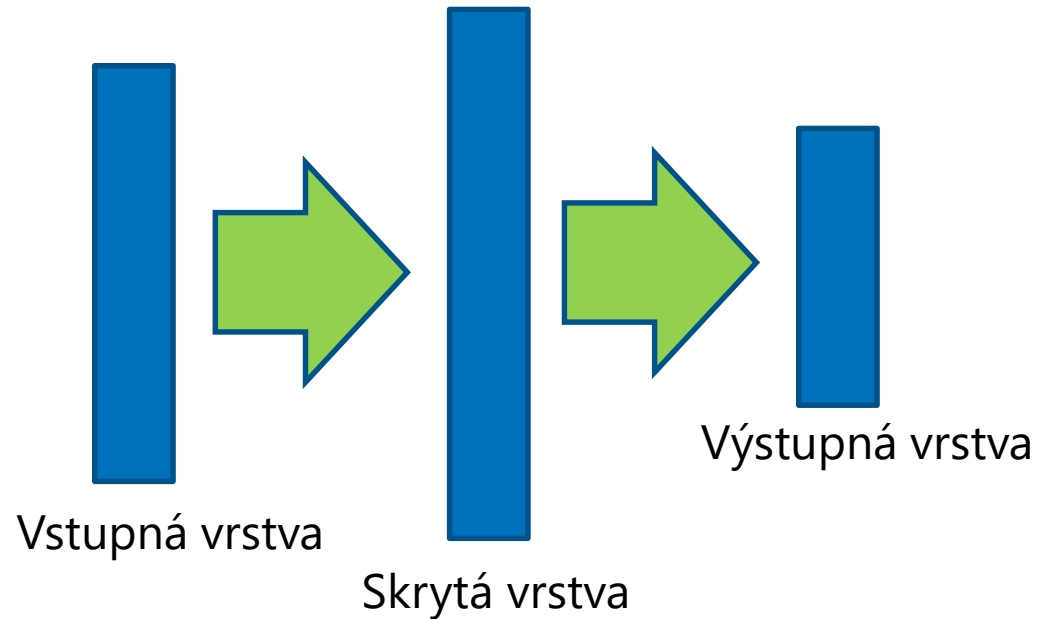
$$h(\mathbf{x}) = \begin{cases} 1 & : ak \mathbf{w} \cdot \mathbf{x} + b > 0 \\ 0 & : inak \end{cases}$$

Čo perceptrón ale nevie

- Perceptrón sa trénuje pomocou zmeny váh tak, že pre každý vstupný vektor z trénovacej množiny sa vypočíta výstup perceptrónu a váhy sa upravujú tak, aby sa minimalizovala chyba klasifikácie.
- Perceptrón rieši LEN lineárne separovateľné problémy

Viacvrstvové neurónové siete (MLP)

Dopredné šírenie

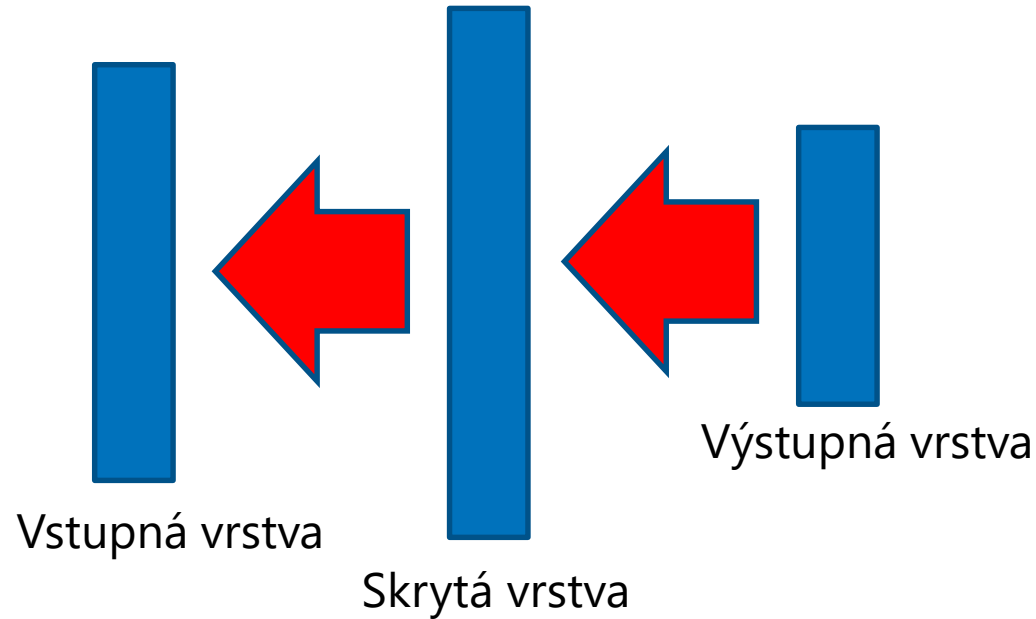


Viacvrstvové neurónové siete (MLP)

Na trénovanie viacvrstvovej neurónovej siete sa môže použiť učenie so spätným šírením chyby, ktoré je založené na vypočítaní chyby na výstupe neurónovej siete a spätnej propagácii tejto chyby cez neurónovú sieť s úpravou váh.

Viacvrstvové neurónové siete (MLP)

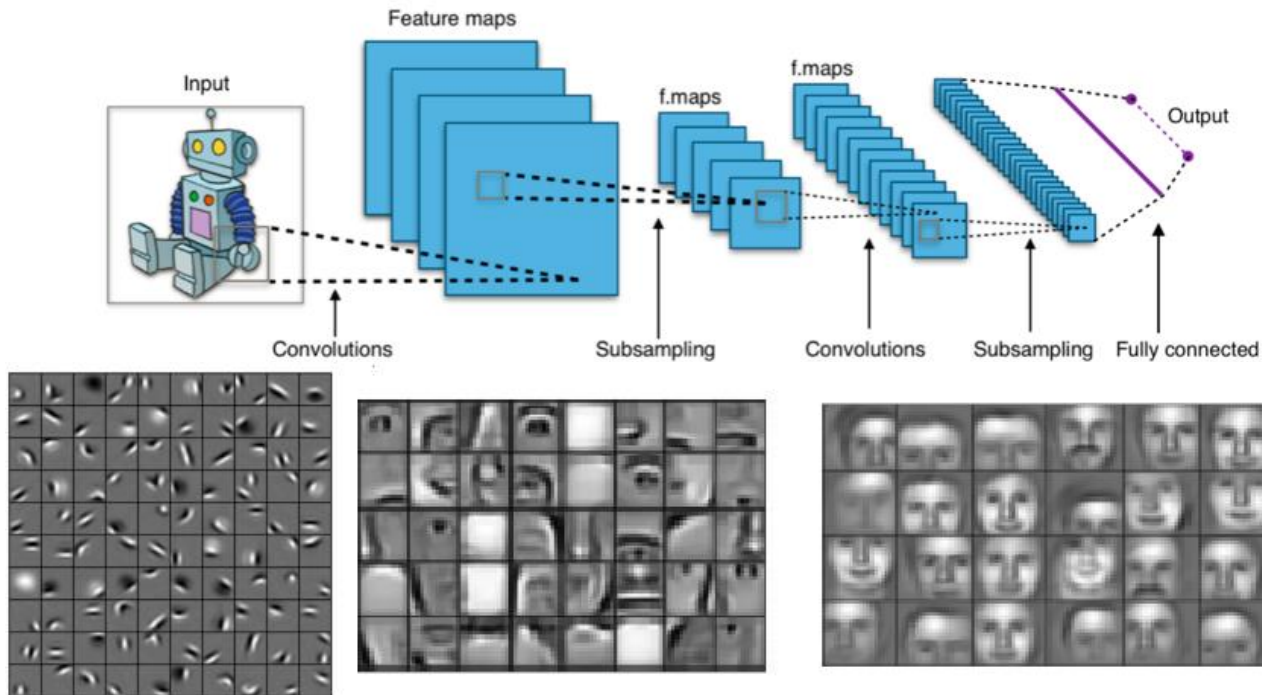
Spätné šírenie chyby



Od MLP k hlbokým sieťam

- Sieť s jednou skrytou vrstvou je univerzálny aproximátor
- ALE
 - Požiadavka na veľkosť skrytej vrstvy môže extrémne narásť
 - Problém pretrénovania – sieť nengeneralizuje, ale učí sa naspamäť
- Riešenie – hlboké siete
- ALE
 - Problém miznúceho a explodujúceho gradientu
 - Problém pretrénovania

Konvolučné siete



H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng. "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations." In ICML 2009.

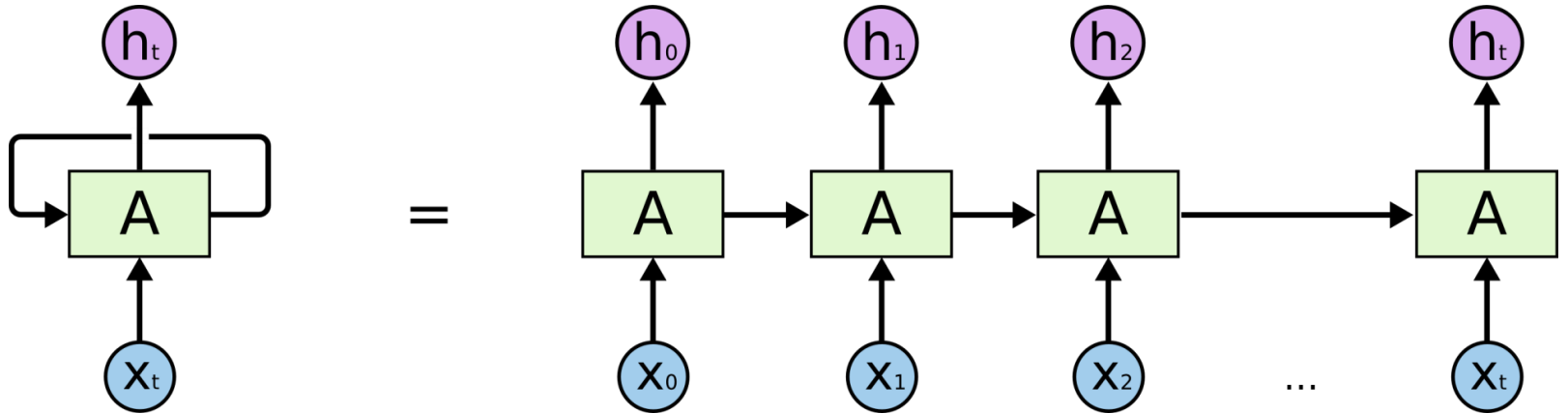
Konvolučné siete

- Konvolučné neurónové siete sú dopredné neurónové siete inšpirované vizuálnym kortexom živých organizmov, úspešne sa používajú na analýzu obrazu a textu
- Počas trénovania si sieť na vrstvách vytvára rôzne stupne abstrakcie vstupu. Nižšie vrstvy sa špecializujú na odhaľovanie nízkoúrovňových črt, pričom ďalšie vrstvy sa viacej špecializujú a vytvárajú zložite koncepty

Konvolučné siete

- Pozostávajú zo striedajúcich sa konvolučných a vzorkovacích vrstiev
 - Konvolučná vrstva je reprezentovaná množinou menších filtrov, ktoré sa postupne aplikujú na celú plochu obrázku.
 - Vzorkovacia vrstva sa používa pre nelineárne zmenšenie výstupov konvolyčnej vrstvy (tradične max pooling)
- Výstup celej siete je tvorený vrstvami plne prepojených neurónov, ako pri normálnych dopredných sieťach

Rekurentné siete



Christopher Olah, Understanding LSTM Networks, 2015, <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Rekurentné siete

- Na rozdiel od dopredných sietí obsahujú aj prepojenia, ktoré vytvárajú cykly
- Majú schopnosť „pamätať si“ predchádzajúce vstupy
- Používajú sa na sekvenčné množiny dát
 - rozoznávanie hlasu, generovanie hlasu
 - predikcie počasia
 - analýza textu

Projekty, ktoré riešime FIIT STU

- Použitie architektúry WaveNet a LSTM siete na predikciu:
 - spotreby el. energie
 - počasia
- Učenie s posilňovaním na vytváranie povelov nákupu a predaja akcií na burze
- Eyetracking cez webkameru
- Prepis audio záznamu do notového zápisu
- Segmentácia zákazníkov webového portálu
- Predikcia konverzie čitateľa na novinovom portáli so spoplatneným obsahom