Reverzibilná extrakcia príznakov

Matúš Mrázik, *Školiteľ: prof. Ing. Martin Klimo, PhD.*
Fakulta riadenia a informatiky, Žilinská univerzita v Žiline

Motivácia

Rozpoznávanie vzorov hrá dôležitú úlohu v oblastiach analýzy dát a data miningu. Už niekoľko desaťročí je predmetom intenzívneho skúmania za účelom jednoduchšieho a efektívnejšieho využitia v praxi. S rýchlym vývinom hardvéru, ktorý je dnes omnoho výkonnejší, sa jeho použitie stáva rozšírenejším. Využíva sa napríklad v oblastiach ako počítačové videnie či rozpoznávanie hlasu. Stále sú však v tejto oblasti výzvy, ktorým je potrebné sa venovať. Dáta, ktorých objem sa neustále zvyšuje, sú totiž často komplexné, nevyvážené, alebo iným spôsobom zašumené.
Systém rozpoznávania vzorov zvyčajne pozostáva z troch hlavných častí. **Najskôr sa dáta predspracujú** tak, aby boli vhodné pre nasledujúce časti. Predspracovanie závisí od druhu dát (zvuk, obraz, ...). Ďalším krokom je **extrakcia príznakov**, kde sa extrahujú príznaky reprezentujúce pôvodné dáta v zmenšenom priestore. V poslednom kroku sa objekty podľa získaných príznakov **klasifikujú do jednotlivých tried**.

Ciele práce

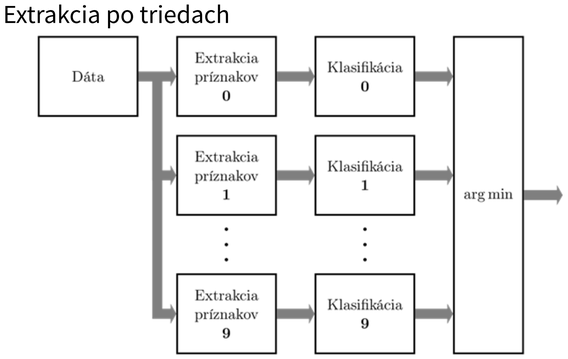
Porovnať lineárne a nelineárne metódy extrakcie príznakov
- Použitie v systéme rozpoznávania vzorov
- Porovnať spoločnú extrakciu a extrakciu po triedach
- Použitý dataset – MNIST

Rozšírený cieľ:
Vývoj metódy pre odhaľovanie cudzích vzorov
- Použitý dataset cudzích vzorov – Fashion MNIST

Porovnanie lineárnych metód extrakcie príznakov

• Spoločná extrakcia

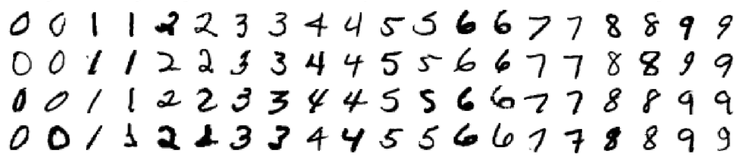
• Extrakcia po triedach

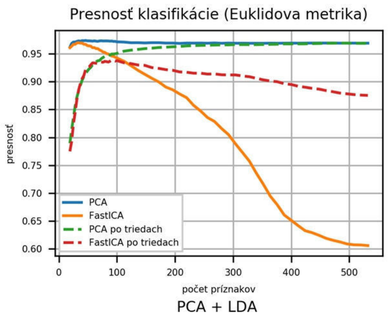


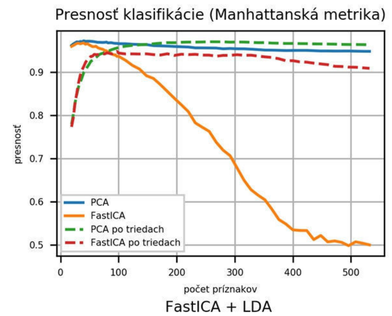
Porovnávané metódy
Lineárne
- Bez učiteľa (*Analýza hlavných komponentov (PCA), Analýza nezávislých komponentov (ICA)*)
- S učiteľom (*Lineárna diskriminačná analýza (LDA)*)

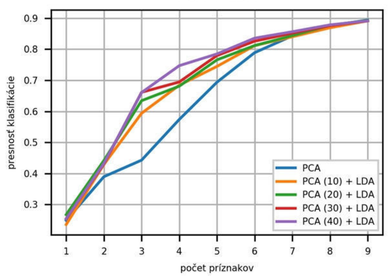
Nelineárne
- Neurónové siete, autoenkodéry

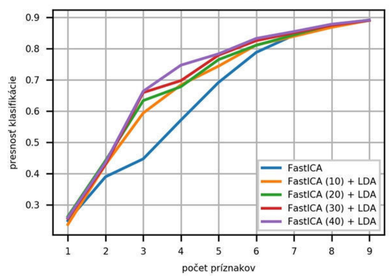
Dataset MNIST je databáza rukou písaných číslic. Obsahuje obrázky arabských číslic 0 až 9. Je to populárny dataset pre štúdium rozpoznávania obrazových vzorov, pretože rukou písané číslice sú ideálnym príkladom vyvážených dát. Trénovacia množina obsahuje celkovo 60 000 vzorov a testovacia množina obsahuje 10 000 vzorov. Všetky obrázky majú velkost 28 × 28 pixelov a sú v odtieňoch sivej.



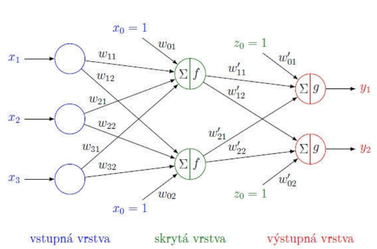


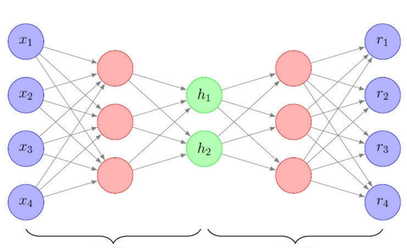






Autoenkodér je typ neurónovej siete, ktorá pozostáva z dvoch častí: ***kodéra*** reprezentovaného kódovacou funkciou ***h*** = *f*(***x***), ktorý kóduje vstupné dáta, a ***dekódera*** reprezentovaného dekódovacou funkciou ***r*** = *g*(***h***), ktorý rekonštruuje zakódované dáta. Autoenkodér sa zvyčajne trénuje tak, aby rozdiel medzi ***x*** a ***r*** bol minimálny, pričom dôležitý je výstup kodéra ***h***, ktorý **pri vhodne zvolenej architektúre dokáže z dát extrahovat relevantné informácie**.

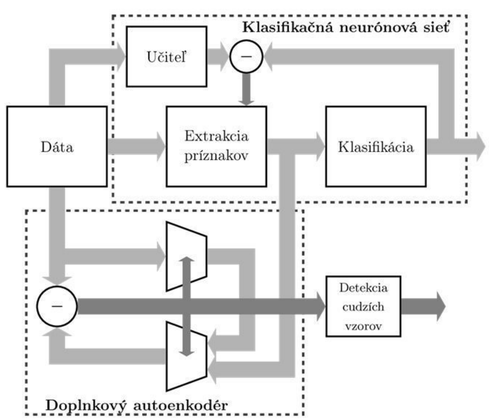


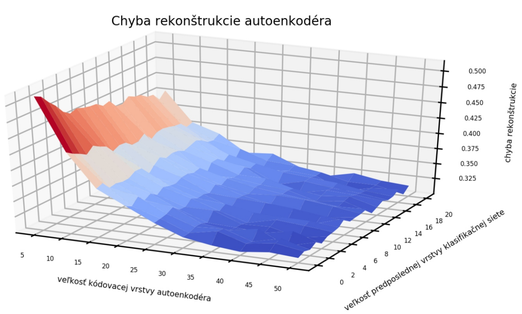


Systémy rozpoznávania vzorov predpokladajú, že vzory budú pochádzať z toho istého priestoru vzorov, do ktorého patria vzory trénovacej množiny použitej na natrénovanie systému. Používateľ pri práci so systémom rozpoznávania nevie, ako sa systém rozhoduje. Je potrebné s dostatočnou presnosťou určiť, či rozpoznávaný vstup patrí/nepatrí do priestoru vzorov trénovacej množiny.

Algoritmus detekcie cudzích vzorov

Klasifikačnú sieť budeme trénovať štandardne s celou trénovacou množinou. Každý autoenkodér budeme trénovať len so vzormi trénovacej množiny patriacimi do danej triedy. Pri testovaní systému budeme klasifikovať vzory testovacej množiny klasifikačnou sieťou. Potom sa pozrieme na doplnkové údaje získané kódovacou časťou príslušného autoenkodéra (chvost vzoru). Pre rozhodnutie o príslušnosti vzoru do priestoru vzorov použijeme dĺžku jeho chvosta (vypočítaná Euklidovou metrikou).





Publikácia: M. Klimo, O. Škvarek, M. Mrázik: „Protection against unintentional misuse of pattern recognition systems“, Pattern Recognition, Q1 2019, Special Issue on Explainable Deep Learning for Efficient and Robust Pattern Recognition