

Implementace metod získávání znalostí

Zadání

Jakub Zárybnický (xzaryb00), Jan Hammer (xhamme00)

I. PŘEHLED

Kohonenova mapa (**self-organizing map**, SOM) je typ neuronové sítě trénovaný pomocí učení bez učitele (unsupervised learning). Používá se pro snížení dimenzionality dat, často pro projekci do 2D/3D prostoru za účelem vizualizace. Také se používá pro klasifikační úkoly, kdy výsledkem učení jsou shluky. Podobnými metodami jsou **k-means clustering** nebo **support vector machines**, které mohou dostávat lepší výsledky v klasifikačních úkolech, ale zato nemají tak čitelné vizualizační výstupy.

V naší práci budeme SOM implementovat v jazyce Python za pomoci knihovny Numpy. Budeme vycházet z primárních zdrojů (např. [3]) a výsledek práce demonstrujeme na veřejně dostupnými sadami dat.

II. CÍLE

- 1) Implementovat knihovnu pro vytváření, učení, ukládání, načítání a klasifikaci pomocí **self-organizing map**
- 2) Volitelné rozšíření: stejná funkčnost pro **growing self-organizing map**
- 3) Implementovat skript, který knihovnu použije pro operace s mapou a který umí vizualizovat výsledky učení.
 - `./som.py learn -i training_data.csv -m model.pickle`
 - `./som.py classify -i test_data.csv -m model.pickle`
 - `./som.py visualize -m model.pickle`
- 4) Demonstrovat použití skriptu a knihovny nad sadou dat **Iris** [0] a volitelně sadami dat **Digit recognizer** [1] a **Gene expression dataset (Golub et al.)** [2], porovnat přesnost vytvořeného modelu s referenční implementací a vytvořit vizualizace tohoto modelu.
- 5) Sepsat písemnou zprávu obsahující strukturu knihovny a použitý algoritmus, výsledky získané na testovacích sadách dat, úspěšnost našeho modelu a získané vizualizace.

REFERENCES

- [1] <https://www.kaggle.com/uciml/iris>
- [2] <https://www.kaggle.com/c/digit-recognizer>
- [3] <https://www.kaggle.com/crawford/gene-expression>
- [4] Teuvo Kohonen: Self-Organizing Maps, ISBN: 9783540679219