FAKULTA STAVEBNÍ, KATEDRA GEOMATIKY
155YGEI Geoinformatika

4. Clusterizační algoritmy

Matěj Klimeš, Tomáš Zbíral
ZS 2024/25, číslo skupiny: 2, datum zpracování: 30.11.2024

---

1 Zadání

Vytvořte svůj vlastní skript pro clusterizaci bodů pomocí k-means metody. Jako bonusovou úlohu upravte skript tak, aby přijímal jako input trojrozměrná, resp. nrozměrná data.

2 Teoretický úvod

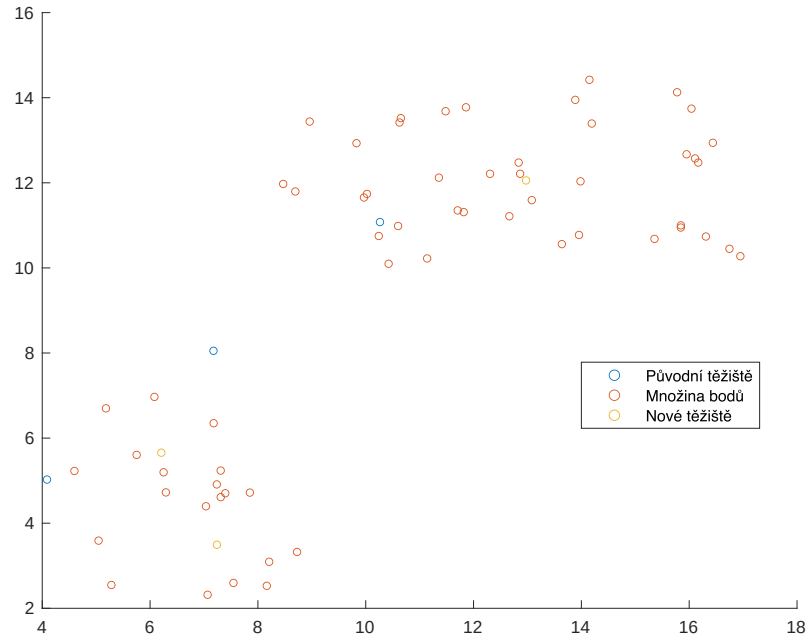
K-means je algoritmus poprvé definovaný Stuartem Loydem [1] v roce 1957. Algoritmus využívá myšlenky, že objekty, ze kterých chceme vytvořit shluky, jsou chápány jako body v euklidovském prostoru. Algoritmus funguje tak, že iterativně přiřazuje body k centroidům (v první iteraci voleným náhodně), které jsou poté přepočítány do těžiště shluků. Poté jsou body přiřazeny k novým centroidům a výpočet se opakuje. Tento výpočet je konvergentní.

3 Pracovní postup a získané výsledky

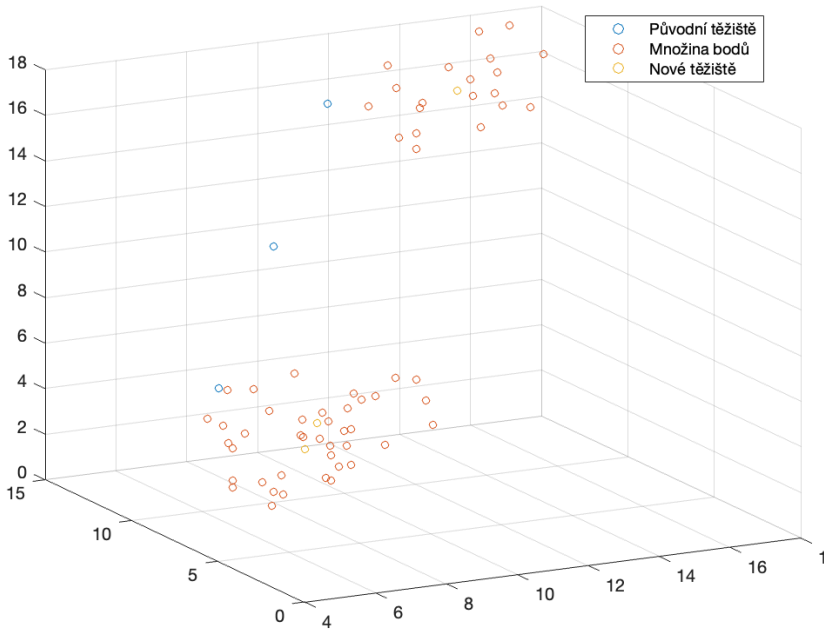
V rámci úlohy jsme nejprve vygenerovali shluky bodů, spojili je do jedné matice a inicializovali jsme těžiště pro první výpočet. Inicializace těžišť jednotlivých shluků proběhla tak, že jsme nejprve vypočetli rozdíl mezi maximálními a minimálními hodnotami $dP = \max(P) - \min(P)$, kde P reprezentuje množinu bodů. Tento výpočet proběhl v rámci všech bodů a to pro všechny dimenze. Souřadnice těžišť jsme pak vypočetli podle vzorce $T_i = \min(P) + \frac{dP \cdot i}{n_c}; i \in [1, 2, \ldots, n_c]$, kde n_c reprezentuje počet shluků, do kterých chceme bod rozřadit. V následujícím kroku jsme bodům přiřadili příslušnost k nějakému těžišti, vypočetli jeho vzdálenost od všech těžišť, a jeho přiřadili ho k těžišti, ke kterému mělo nejkratší vzdálenost. Na konci cyklu jsme vždy vypočetli posuny těžišť oproti jejich předchozí poloze a výpočet provádeli tak dlouho, dokud se tato vzdálenost nebyla menší, než 0.1. Jako poslední stojí za zmínku, že náš skript generuje grafy, pokud jsou body ve 2D (Obrázek 1), resp. 3D prostoru (Obrázek 2).

4 Závěr

Přijd emi zvláštní, že metodu k-means jsme využívali v rámci 2 různých cvičení. Osobně se mi více líbí tato úloha, kde si zabýváme "střevy"k-means algoritmu. Díky tomu by šla část 2.úlohy třeba nahradit jinou úlohou a využití tohoto algoritmu by šlo ukázat pouze v rámci přednášky (nebo použít jako velmi zajímavou bonusovou úlohu). Ocenil bych, kdyby na stránkách předmětu bylo nějaké zadání v pdf, tak jako je tomu u jiných úloh.



Obrázek 1: Ukázka skriptu na 2D datech.



Obrázek 2: Ukázka skriptu na 3D datech.

5 Přílohy

Příloha 1 - MATLAB funkce - kmeans.m
Příloha 2 - MATLAB skript, který využívá tuto funkci - U4.m

Reference

[1] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.