

Analiza skupień

Damian Wysokiński

9/11/2020

Opis zadania

Celem projektu było zaimplementowanie i przetestowanie algorytmu spektralnego analizy danych.

Algorytm składa się z 4 głównych elementów:

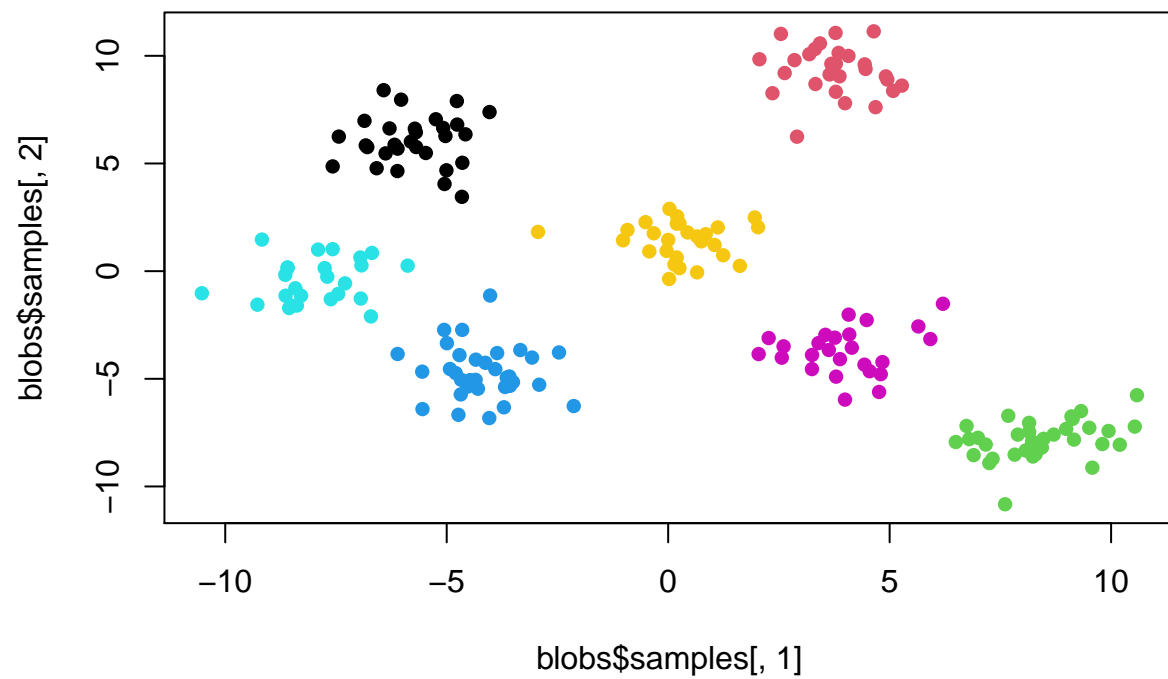
1. znalezienie M najbliższych sąsiadów wszystkich punktów
2. stworzenie grafu sąsiedztwa
3. wyznaczenie odpowiednich k wektorów własnych jego laplasjanu
4. zastosowanie algorytmu k średnich w nowej przestrzeni danych

Testy

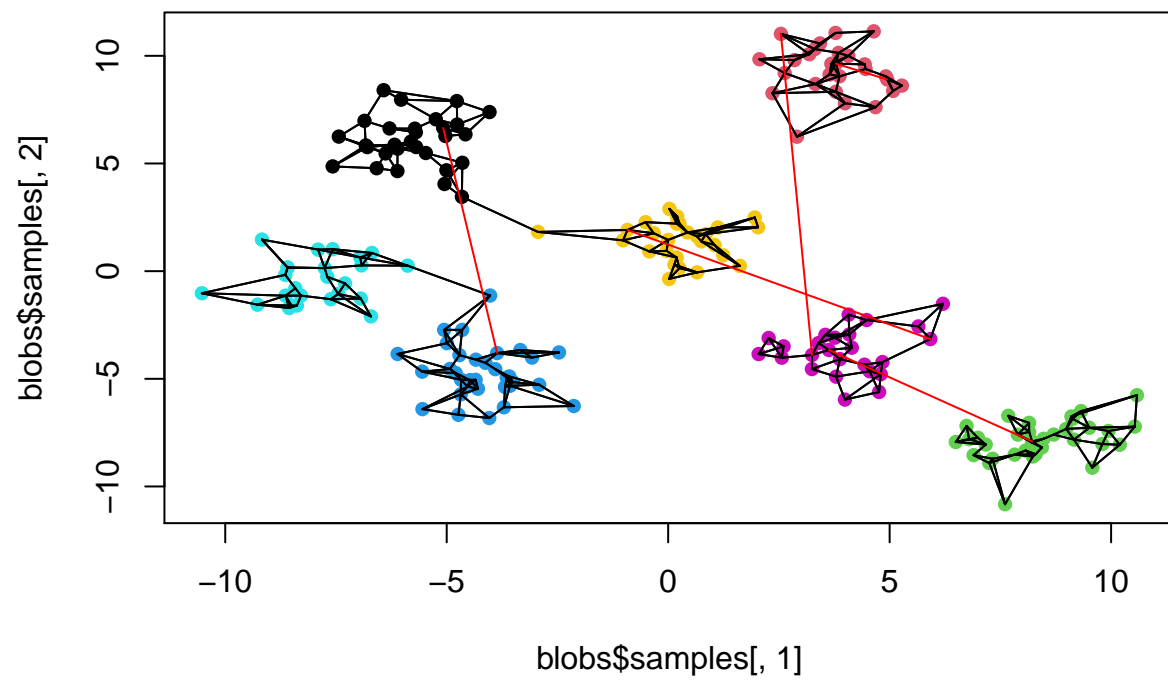
Testy danych 2-wymiarowych

Test dla następujących danych:

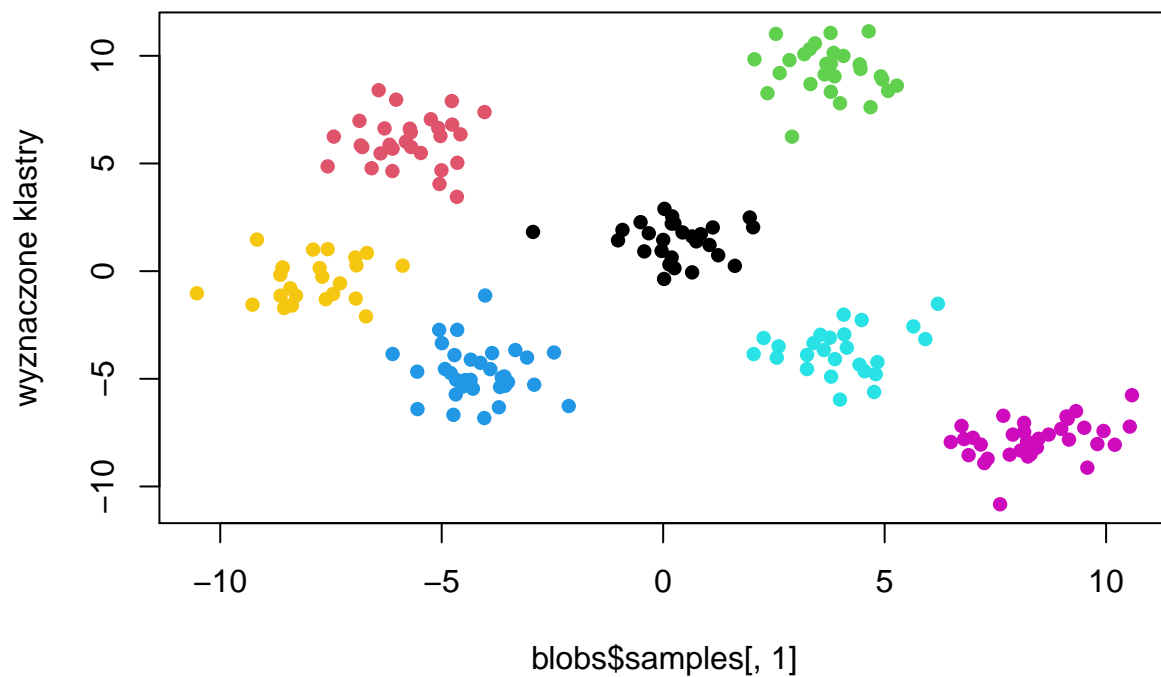
- liczba próbek 200
- liczba klastrów 7
- liczba najbliższych sąsiadów 3



Stworzenie grafu i uspójnienie go (dodane krawędzie zaznaczone czerwonymi liniami).



Wyznaczenie klastrow z danych:

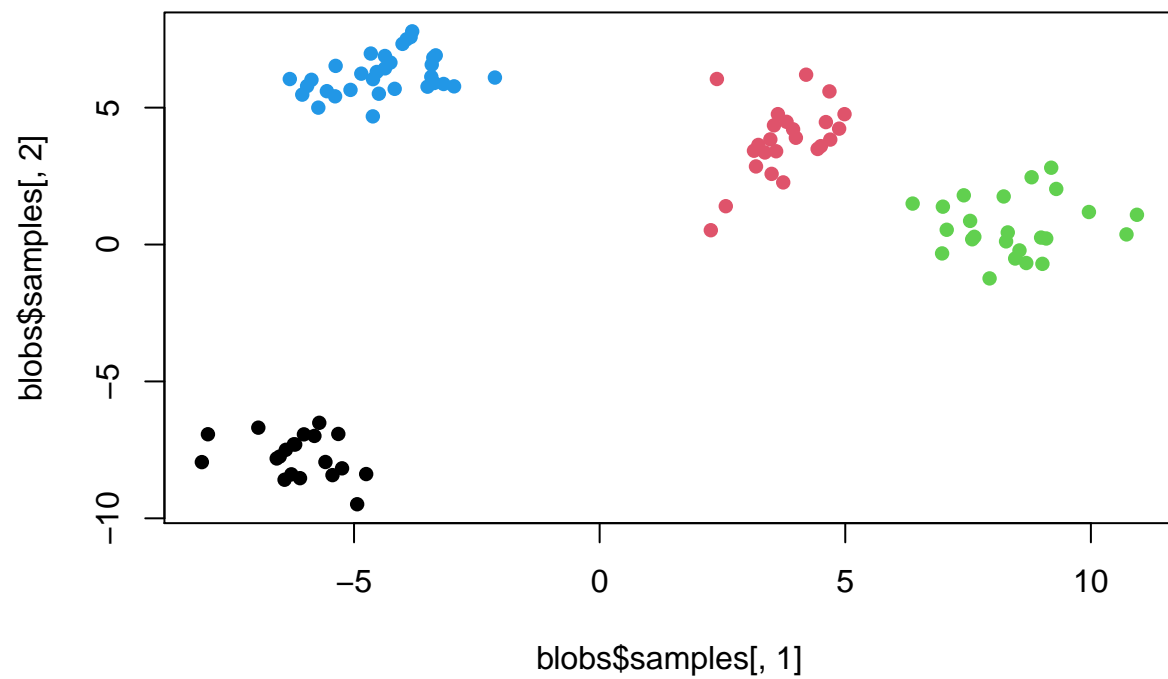


Ocena jakości klastryzacji z wykorzystaniem FM:

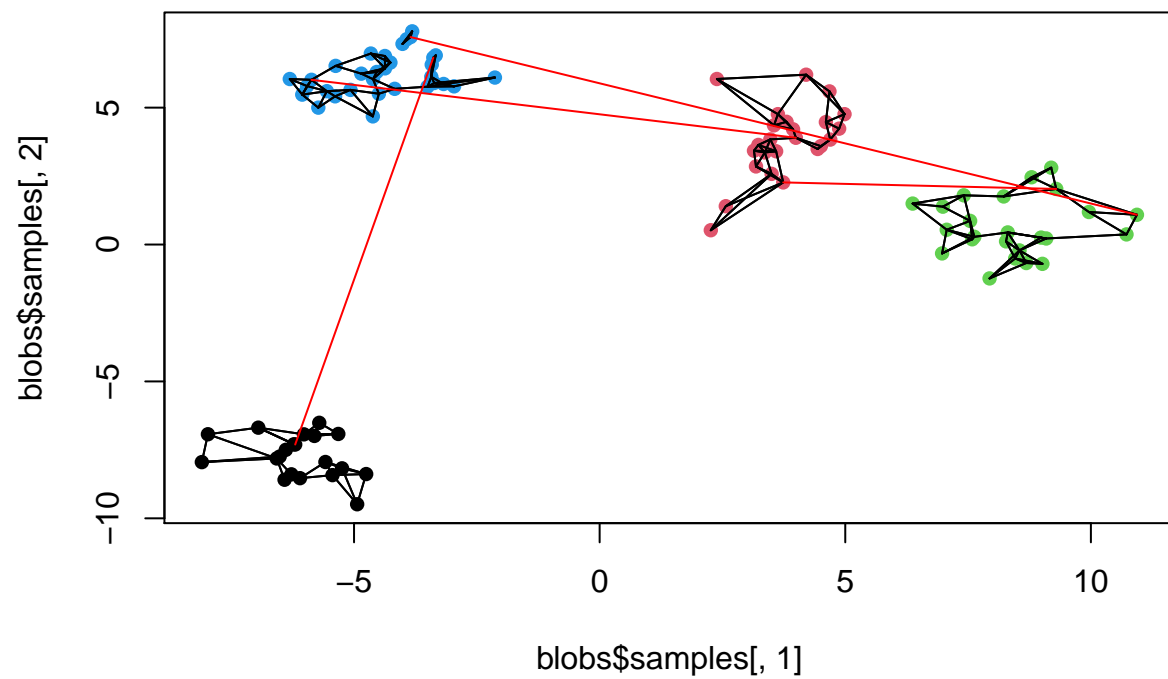
```
## [1] 1
## attr(,"E_FM")
## [1] 0.1405528
## attr(,"V_FM")
## [1] 3.721259e-05
```

Test dla następujących danych:

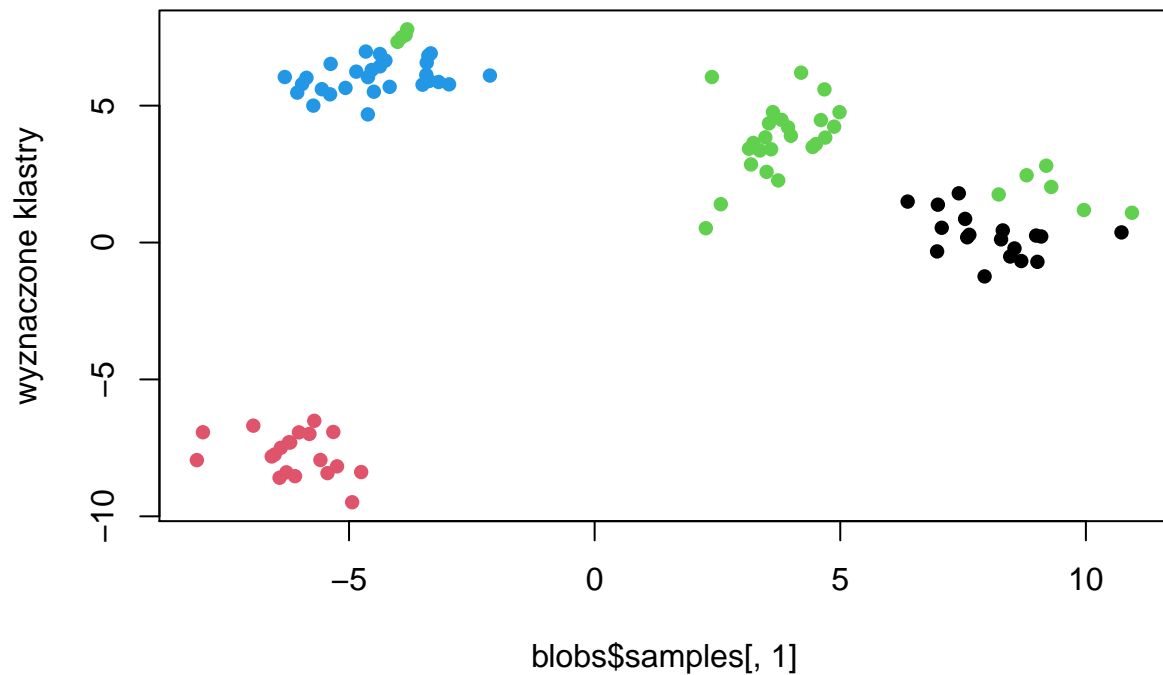
- liczba próbek 100
- liczba klastrów 4
- liczba najbliższych sąsiadów 3



Stworzenie grafu i uspořjnienie go.



Wyznaczenie nowych klastrow:



Jak pokazane na powyższym rysunku, nie został dokładnie odwzorowany początkowy podział na klastry. Duży wpływ na podział danych ma losowy wybór nowych krawędzi przy uspójnieniu grafu. Błędy pojawiają się w miejscach, gdzie były dodane nowe krawędzie.

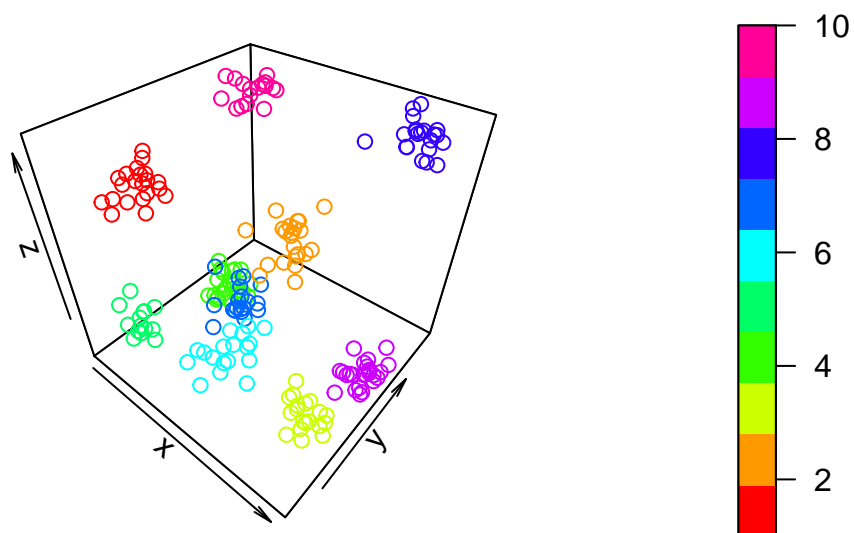
Ocena jakości klasteryzacji z wykorzystaniem FM:

```
## [1] 0.8080597
## attr(,"E_FM")
## [1] 0.2545067
## attr(,"V_FM")
## [1] 0.0001126454
```

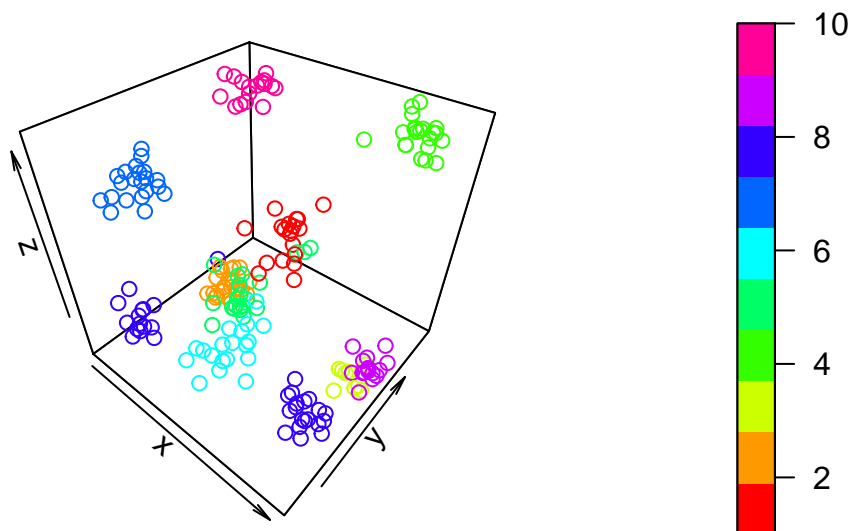
Testy danych 3-wymiarowych

Test dla następujących danych:

- 200 próbek
- 10 klastrów
- liczba najbliższych sąsiadów 3



Wyznaczenie nowych klastrów:

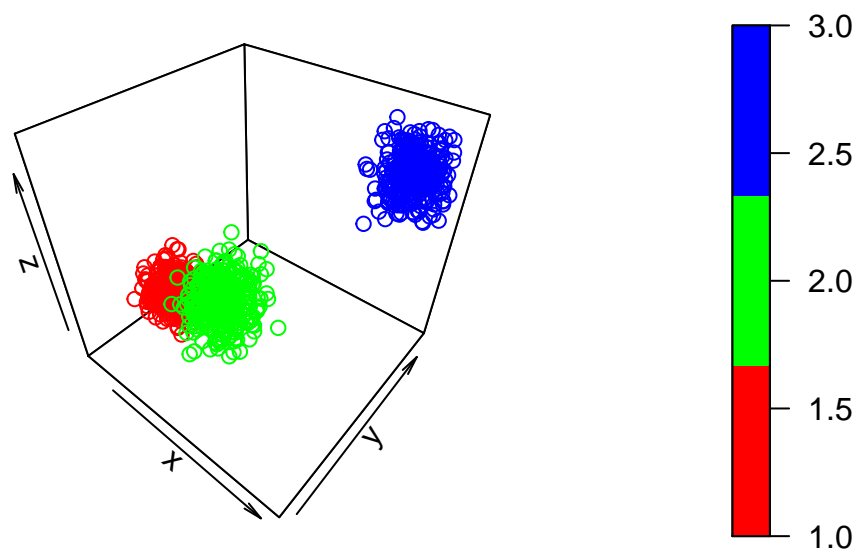


Ocena jakości klasteryzacji z wykorzystaniem FM:

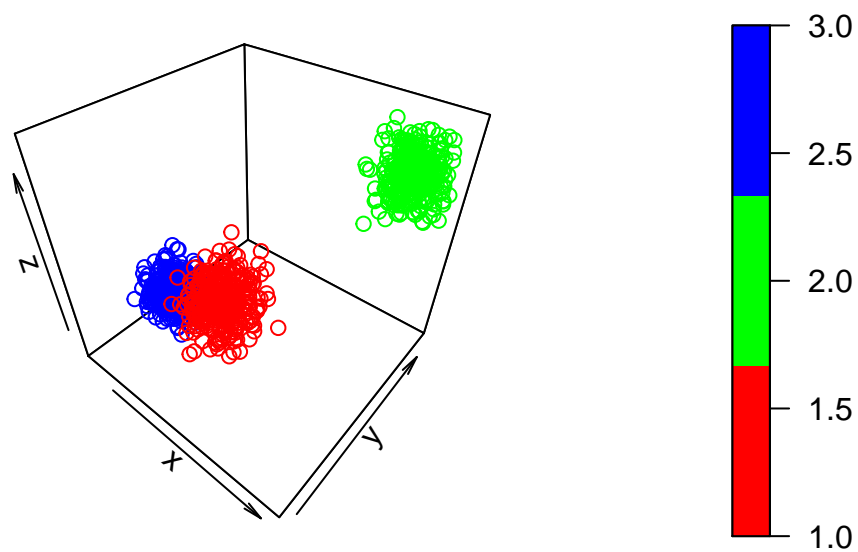
```
## [1] 0.8482128
## attr(,"E_FM")
## [1] 0.101129
## attr(,"V_FM")
## [1] 4.038286e-05
```

Testy dla następujących danych:

- 1000 próbek
- 3 klastry
- 10 najbliższych sąsiadów



Wyznaczone klastry:



Ocena jakości klasteryzacji z wykorzystaniem FM:

```
## [1] 1
## attr("E_FM")
## [1] 0.3327007
## attr("V_FM")
## [1] 8.930739e-07
```