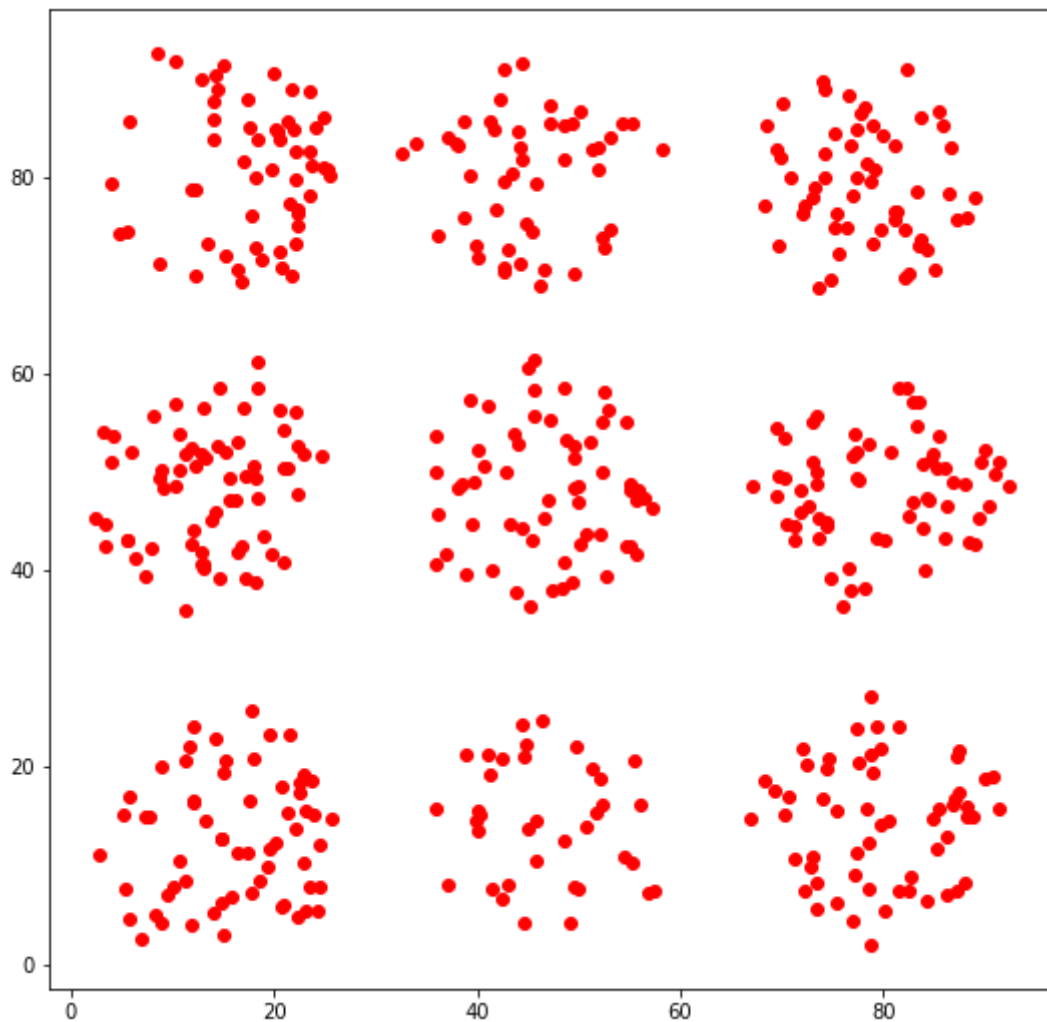


**Anna Marciniak**

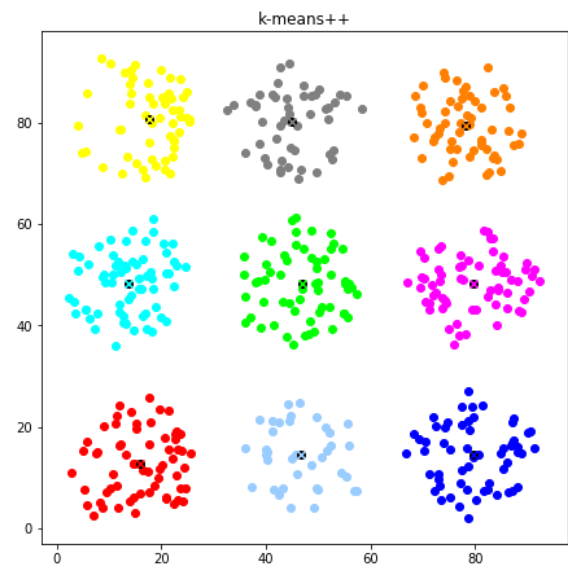
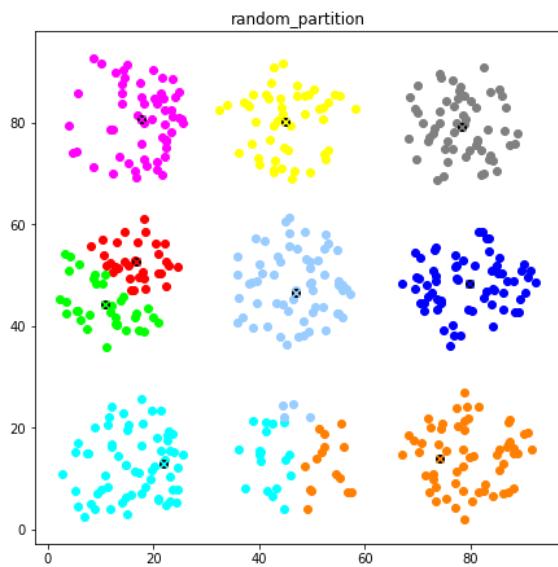
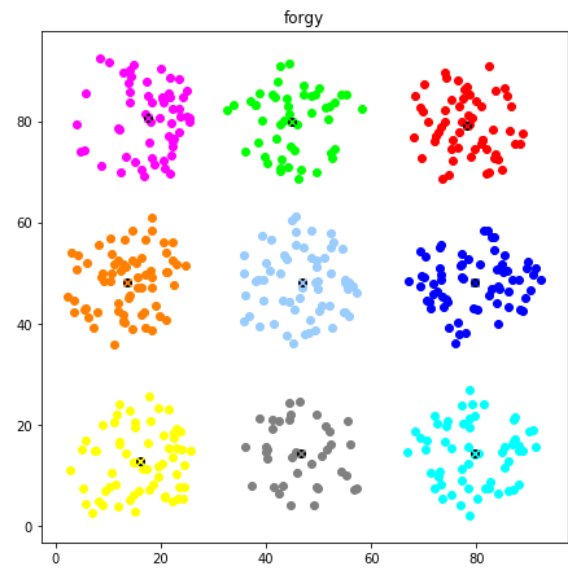
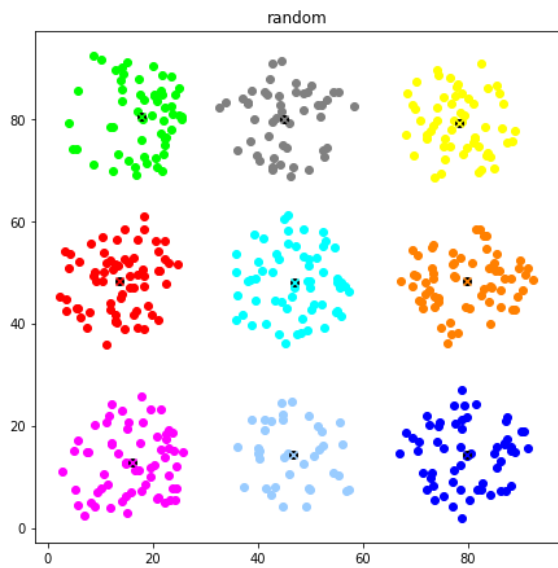
## Podstawy Uczenia Maszynowego lab 4 - **Metody inicjalizacji k-means** **Raport**

Stworzyłam zbiór na podobieństwo zbioru z rysunku.

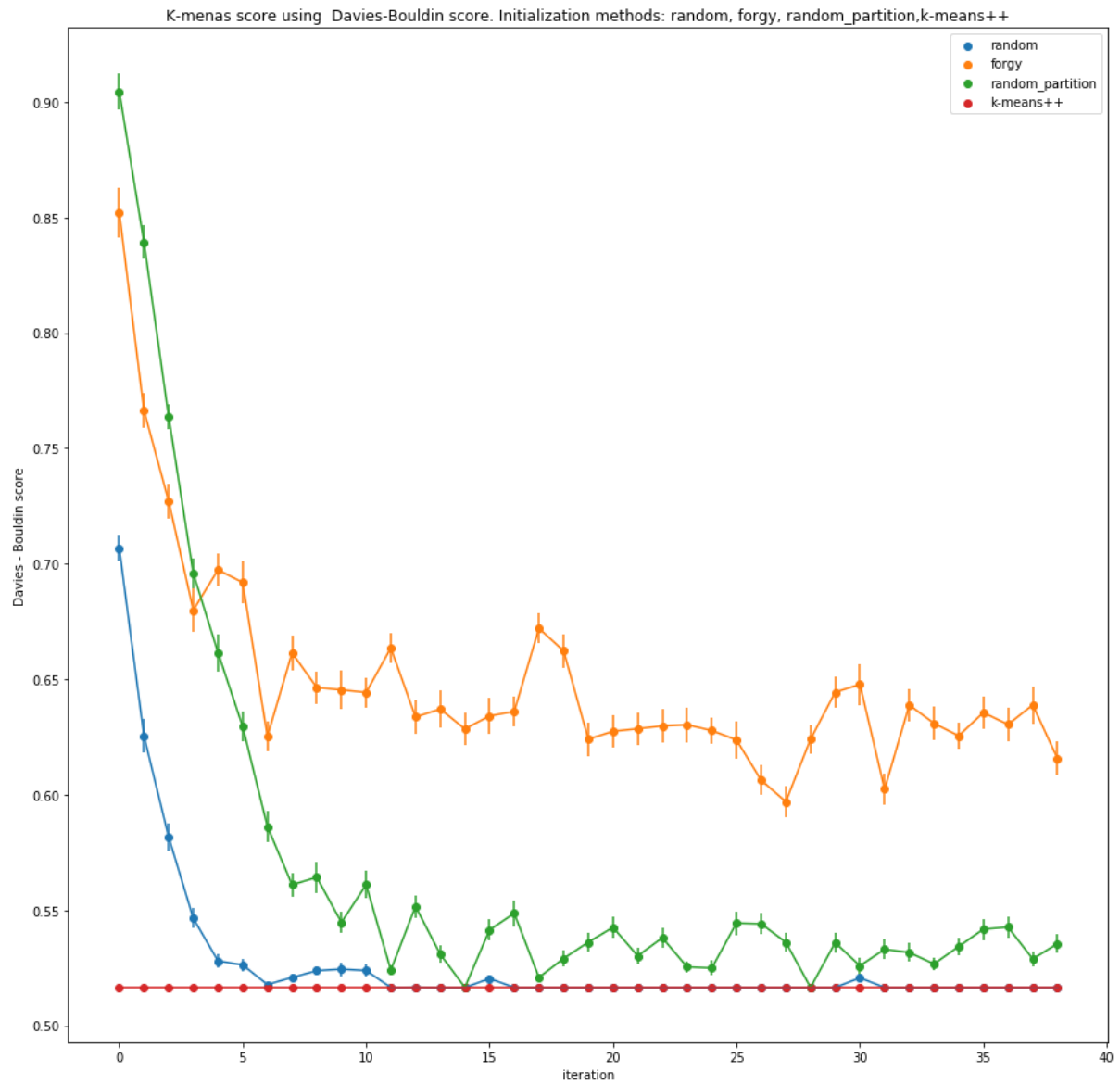


Następnie zastosowałam metodę k means do znalezienia klastrów.

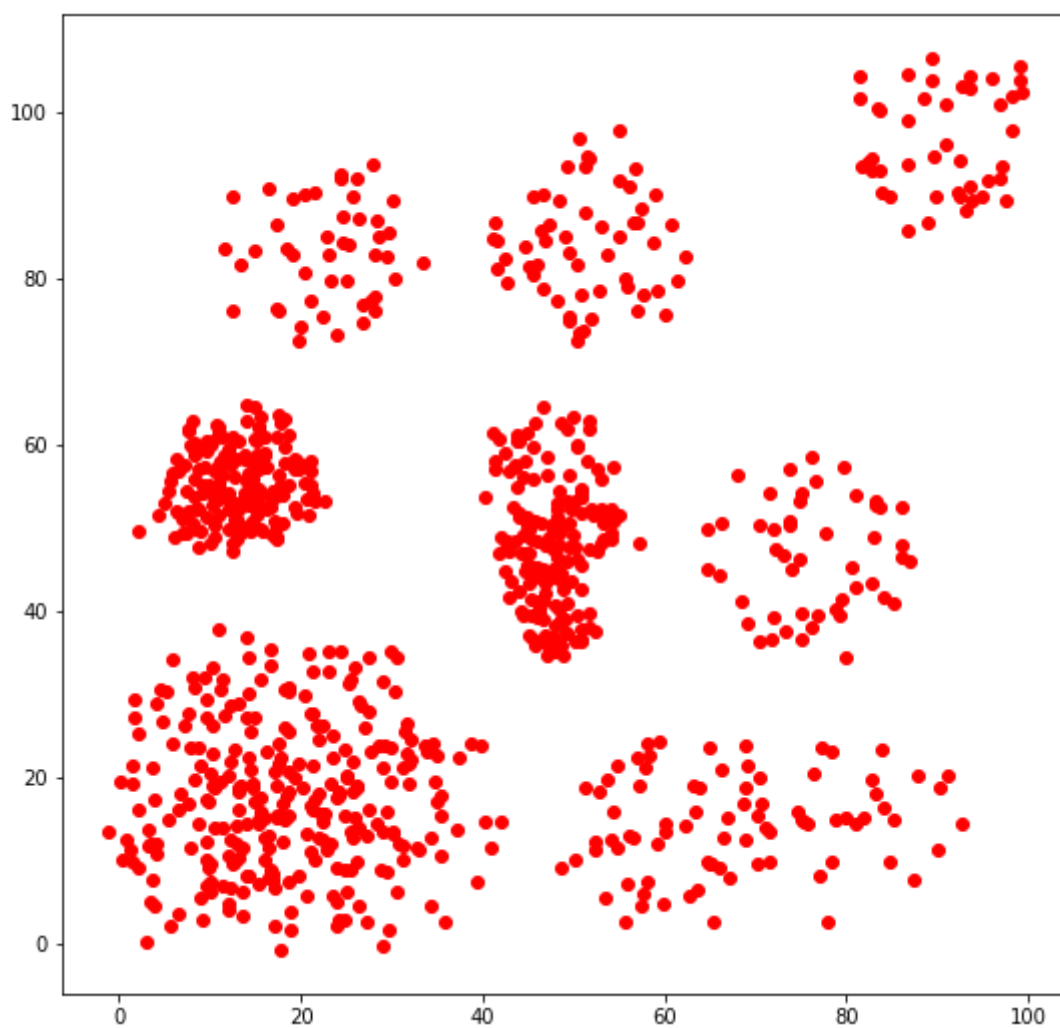
Clusters found using k means with particular init methods



Dla iteracji od 1 do 40 policzyłam index Davies-Bouldin'a, dla każdej metody inicjalizacji powtórzyłam to 20 razy, policzyłam średnią i odchylenie standardowe. Wyniki naniosłam na wykres.

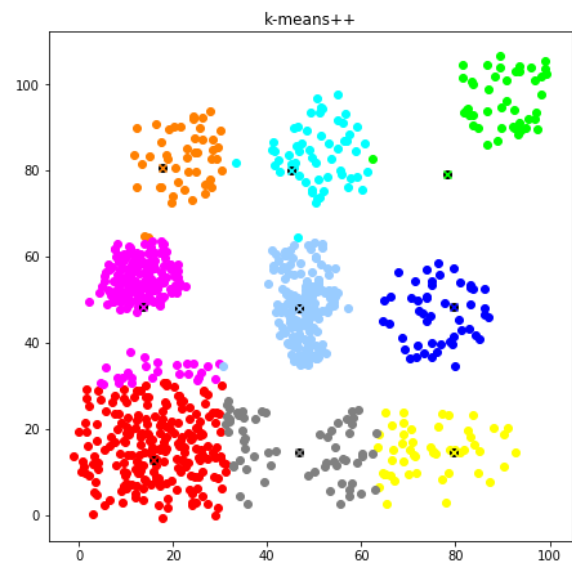
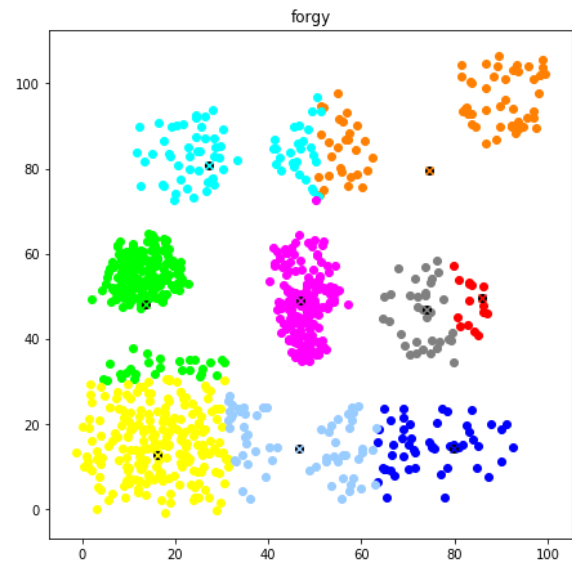
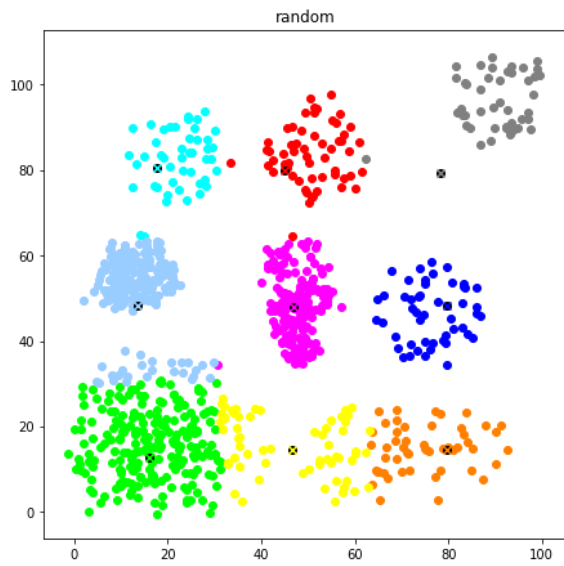


Następnie “popsułam” zbiór w następujący sposób:

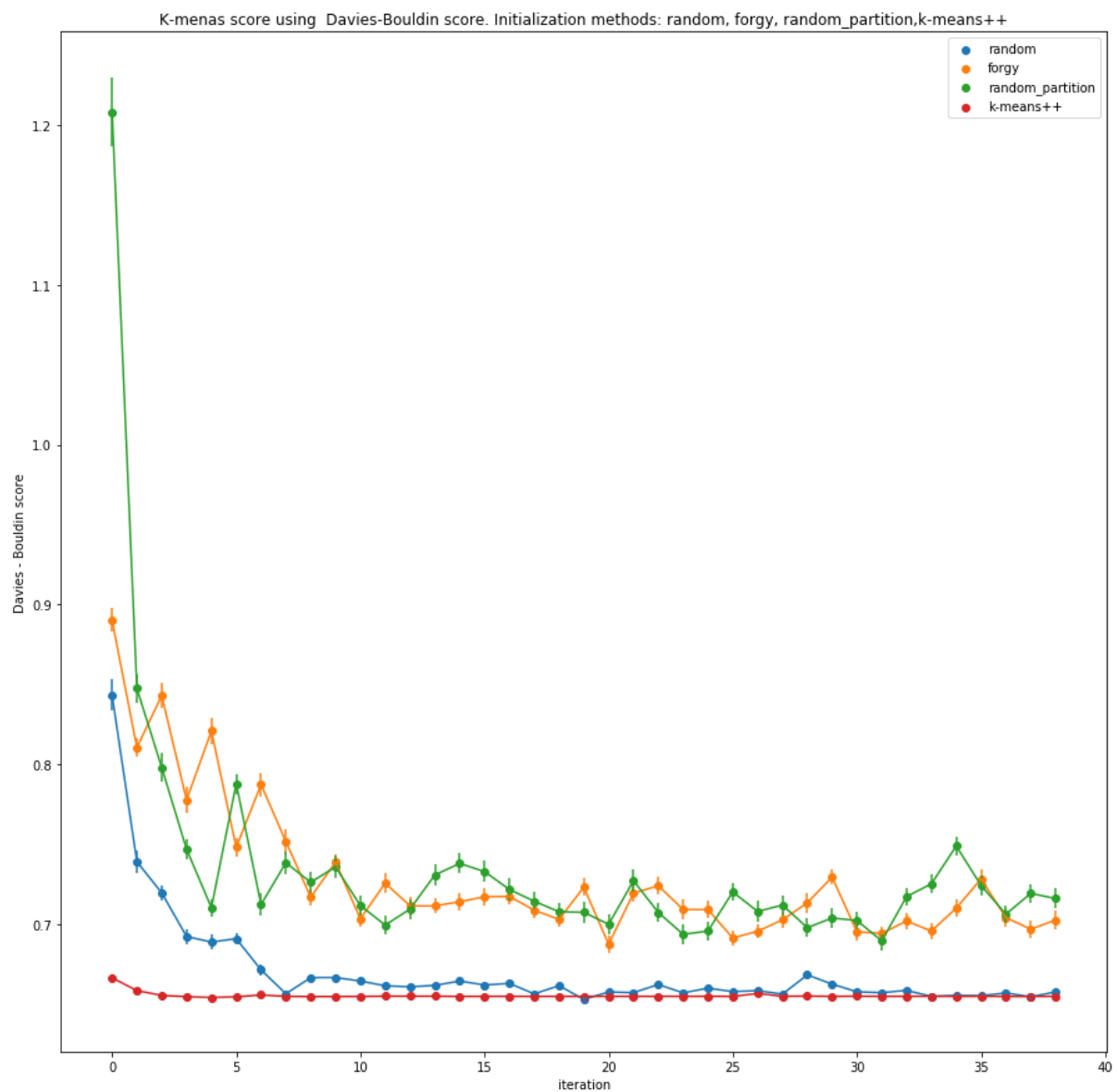


## Znalezienie klastry:

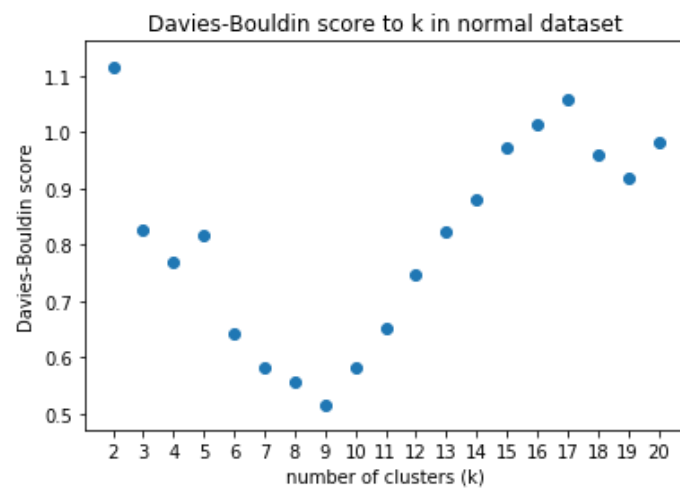
Clusters found using k means with particular init methods



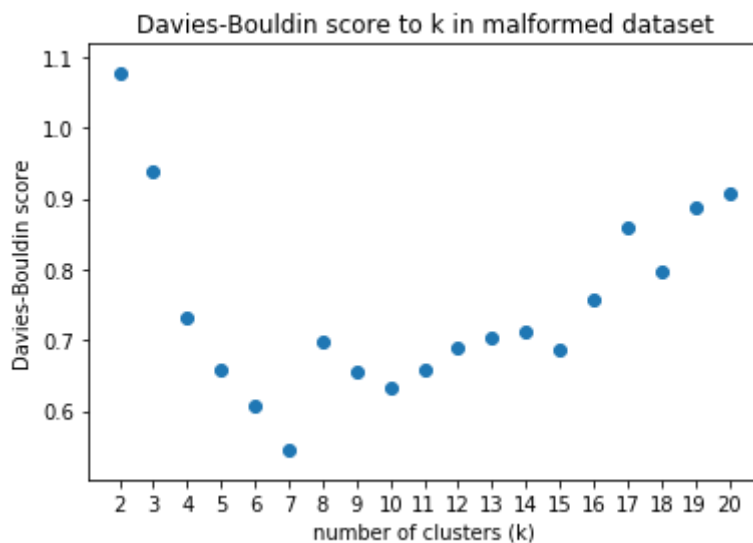
Index Davies-Bouldin'a liczony jak poprzednio



Index Davies-Bouldin'a w zależności od liczby klastrow(k) dla zbioru normalnego:



Dla zbioru "popsutego":



Metoda inicjalizacji to k-means++

Wnioski:

- Dla 'normalnego' zbioru, najmniejszą wartość score'u Davies-Bouldin'a ma k-means zainicjalizowane metodą k-means ++, random na początku słabiej sobie radzi w porównaniu do k-means ++, ale już od około 5-10 iteracji wypada porównywalnie do k-means++. Random partition zaczyna gorzej niż forgy, ale tak po (5-10) iteracjach score jest bliżej random i k-means++. Forgys zdecydowanie najgorzej wychodzi. Gdyż im mniejszy index Davies-Bouldin'a tym klasteryzacja jest lepsza, to k-means++ wybrało na początku dobre centra, random - po kilku iteracjach się centra ustawiły w odpowiednich miejscach, random partition wybrał na początku słabe centra, ale się dało poprzesuwać tak, żeby nie było bardzo źle.
- Dla 'zepsutego' zbioru k-means++ nie od razu trafił w dobre centra, ale szybko zostały one poprzysuwane, random na początku centra nie są najoptymalniejsze, ale później radzi sobie troszeczkę gorzej niż k-means ++, Random partition zaczyna bardzo słabo, ale po kilku iteracjach centra są skutecznie tak przesuwane jak w Forgys.
- Dla 'normalnego' zbioru index Davies-Bouldin'a jest najniższy dla  $k=9$ , więc można powiedzieć, że  $k=9$  jest optymalne
- Dla 'zepsutego' zbioru index Davies-Bouldin'a jest najniższy dla  $k=7$  więc  $k=9$  nie jest optymalne