

Wprowadzenie do uczenia maszynowego

Lista 2

Janusz Szwabiński

1. **Generowanie danych syntetycznych:** Wygeneruj syntetyczny zbiór danych z trzema wyraźnie oddzielonymi skupieniami, używając funkcji `make_blobs` z biblioteki `scikit-learn`. Zwizualizuj dane za pomocą wykresu punktowego.
2. **Implementacja algorytmu k-means:** Zaimplementuj algorytm k-means od podstaw (bez użycia `scikit-learn.cluster`). Twoja implementacja powinna obejmować:
 - Inicjalizację centroidów.
 - Pętlę, w której punkty są przypisywane do najbliższego centroidu.
 - Aktualizację położenia centroidów jako średniej z przypisanych punktów.

Przetestuj swój kod na danych z zadania 1.

3. **Zastosowanie algorytmu k-means z biblioteki scikit-learn:** Użyj klasy `KMeans` z biblioteki `scikit-learn` do pogrupowania danych z zadania 1. Porównaj wyniki z własną implementacją. Zwizualizuj otrzymane klastry, kolorując punkty według ich przynależności.
4. **Klastrowanie hierarchiczne:** Zastosuj algorytm klastrowania hierarchicznego (`AgglomerativeClustering` z `scikit-learn`) na syntetycznym zbiorze danych z zadania 1. Użyj różnych miar odległości (np. `linkage='ward'`) i zwizualizuj wyniki za pomocą dendrogramu. Omów, jak wybór miary odległości wpływa na strukturę skupień.
5. **Ocena jakości klastrowania:** Zastosuj metodę łokcia (Elbow Method) oraz wskaźnik Silhouette, aby określić optymalną liczbę skupień dla zbioru danych `Iris`. Stwórz wykres dla obu metod i zinterpretuj wyniki.

6. **Zastosowanie algorytmu DBSCAN:** Użyj klasy DBSCAN z biblioteki `scikit-learn` do pogrupowania danych z zadania 1. Przetestuj różne wartości parametrów `eps` i `min_samples` i zaobserwuj, jak wpływają one na wyniki klastrowania. Zidentyfikuj "szum" (punkty odstające).
7. **Analiza danych rzeczywistych:** Wczytaj zbiór danych dotyczący klientów (np. `Mall Customer Segmentation Data` z Kaggle). Przeprowadź analizę skupień na podstawie wybranych cech (np. wiek i roczny dochód). Zastosuj algorytm k-means i zwizualizuj otrzymane segmenty klientów. Zinterpretuj wyniki.