LABORATORIUM 6

Zadania

1. Napisać funkcje:

-d=distp(X,C,e), która wyliczy odległość euklidesową między dwoma zbiorami punktów X i C,

-d=distm(X,C,V), która wyliczy odległość Mahalanobis’a między dwoma zbiorami punktów X i C; V jest macierzą kowariancji zbioru X,

Obraz zawierający tekst, Czcionka, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.KOD NAPISANY

-[C,CX]=ksrodki(X,k), która dla zadanej macierzy wzorców X oraz liczby grup k, wyznaczy centra C i sąsiedztwa CX.

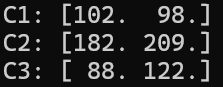
KROK 1

CZĘŚĆ KODU

Obraz zawierający tekst, Czcionka, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Funkcja k\_srodki przyjmuję X oraz k (liczbę grup, środków). Dodatkowo przyjmuję wartość bool w celu ustalenia której funkcji użyć do obliczenia odległości między punktem x a C (funkcja distp lub distm). Tworzymy dwie macierze. Macierz P przechowuje elementy oraz przynależność wektora do danej grupy. Macierz C przechowuje centroidy (centralne punkty). Następnie inicjujemy wektory macierzy C losowo.

REZULTAT

Algorytm został zastosowany dla cech ‘horsepower’, ‘engine-size’ oraz wartości k=3;

KROK 2

CZĘŚĆ KODU

Obraz zawierający tekst, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Tworzymy pomocniczą macierz P oraz listę. Następnie za pomocą funkcji distp lub distm obliczamy odległość. Przypisujemy punktowi klaster do którego punkt ma najbliżej. Następnie sprawdzamy czy Punktom zmieniły się grupy. W przypadku braku zmian przerywamy pętle.

KROK 3

Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Obliczamy dla każdego C nowy środek. Sprawdzamy czy nowe środki różnią się od starszych. W przypadku braku zmian kończymy algorytm.

KROK 4,5

Powtarzamy krok dopóki P i C nie przestaną się zmieniać.

CAŁY ALGORYTM

Obraz zawierający tekst, zrzut ekranu, oprogramowanie

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

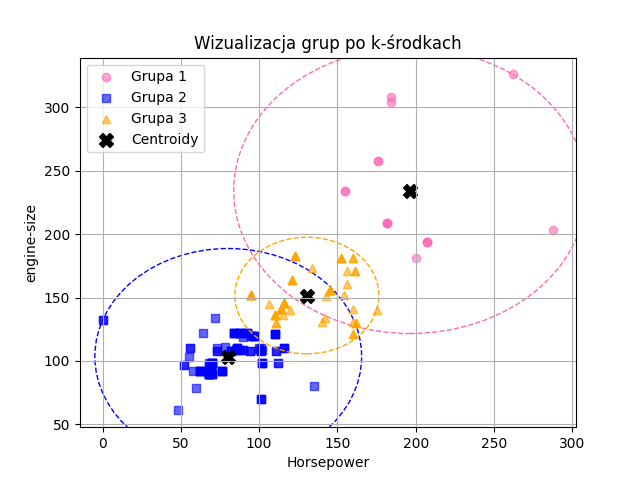
REZULTAT

Obraz zawierający tekst, zrzut ekranu, Czcionka, czarne i białe

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Zobrazowane są tutaj zmiany punktów centralnych oraz ilości punktów które należą do danej grupy.

WIZUALIZACJA



Obraz zawierający tekst, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.KOD NAPISANY DO WIZUALIZACJI

W celu lepszej wizualizacji wyników zostały dodane okręgi dla poszczególnych grup. Promień jest obliczany na podstawie odległości euklidesowej między najbardziej oddalonym punktem od centrum a samym centrum. Okręg sensownie działa tylko dla zmiennych o podobnej osi x i y. Oprócz okręgu są zaznaczone centroidy oraz pogrupowane punkty kolorem, cechą wyglądu.

OBLICZENIE JAKOŚCI GRUPOWANIA

Obraz zawierający tekst, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

REZULTAT

Jakość grupowania wyliczona na podstawie wzoru podanym w ćwiczeniu.

KOŃCOWE ZADANIE

Po zaimplementowaniu proszę w ramach testów uruchomić algorytm dla 4x wartości parametru k oraz 5x dla każdej funkcji odległości (2) i na tej podstawie spróbować określić która z nich jest lepsza (łącznie 40 przebiegów) - wyniki przedstawić w małej tabelce wyróżniając najlepszy i najgorszy wiersz

KOD NAPISANY

Obraz zawierający tekst, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Algorytm jest wywołany dla wszystkich numerycznych cech X w przypadku wartości nan wartość jest zmieniana na 0.

REZULTAT

TABELA WYNIKÓW

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| k | Metryka Euklidesowa | Metryka Mahalanobisa | Najlepszy/najgorszy wynik dla k | Najlepszy/najgorszy wynik |
| 2 | *0.0000043737323514* | *0.0000043737323514* | Wszystkie wyniki są identyczne | *Najgorsze wyniki* |
| 2 | *0.0000043737323514* | *0.0000043737323514* |
| 2 | *0.0000043737323514* | *0.0000043737323514* |
| 2 | *0.0000043737323514* | *0.0000043737323514* |
| 2 | *0.0000043737323514* | *0.0000043737323514* |
| 3 | 0.0000313783621790 | 0.0000313783621790 | Wszystkie wyniki są identyczne |  |
| 3 | 0.0000313783621790 | 0.0000313783621790 |  |
| 3 | 0.0000313783621790 | 0.0000313783621790 |  |
| 3 | 0.0000313783621790 | 0.0000313783621790 |  |
| 3 | 0.0000313783621790 | 0.0000313783621790 |  |
| 4 | 0.0000764470548922 | 0.0000764470548922 | Najgorsze  Najlepsze |  |
| 4 | 0.0000764470548922 | 0.0000764470548922 |  |
| 4 | 0.0000764470548922 | 0.0000764470548922 |  |
| 4 | 0.0000785721689182 | 0.0000764470548922 |  |
| 4 | 0.0000755442343737 | 0.0000764470548922 |  |
| 5 | 0.0001330616023926 | 0.0001365223187766 | Najgorsze  Najlepsze |  |
| 5 | 0.0001383103242069 | 0.0001383103242069 |  |
| 5 | 0.0001375572400571 | 0.0001325641301624 |  |
| 5 | 0.0001217957853958 | 0.0001658226742985 |  |
| 5 | 0.0001707103573141 | **0.0001755300274190** | **NAJWYŻSZY WYNIK** |

PODSUMOWANIE WYNIKÓW

Najwyższy wynik osiągnęła metryka Mahalanobisa. Metryka Euklidesowa raz zwyciężyła przy wartości k=4. Ale dwa razy przegrała. Sądzę, że dla tych typów danych Metryka Mahalanobisa jest bardziej stabilna i wiarygodna.