

Dane **clu**:

X^1	X^2	P_1	P_2	P_3
5	1	1	1	1
5	2	1	3	1
5	3	2	1	1
1	4	2	2	1
4	4	3	3	2
2	5	3	3	3
5	5	1	3	2
1	6	3	1	3
6	6	3	3	2
7	7	2	1	2
8	8	2	2	2
2	10	2	2	3
3	9	1	2	3

1. $X = \begin{bmatrix} X^1 & X^2 \end{bmatrix}$ jest macierzą danych. Kolumny P_1, P_2, P_3 wskazują na trzy sposoby podziału tych danych na trzy grupy 1,2 i 3.

1.1 Narysuj trzy wykresy, ilustrujące te podziały. Oceń wizualnie, który z nich uważasz za najlepszy.

1.2 Dla każdego podziału oblicz:

1.2.1 środki ciężkości grup

1.2.2 bezwładność międzygrupową J_M i bezwładność wewnątrzgrupową J_W . Oceń na ich podstawie, który z tych podziałów jest najlepszy. Czy zgadza się to z oceną z punktu 1.1?

1.3 Każdy podział popraw metodą Woronoja (k- średnich) aby uzyskać najlepszy możliwy podział.

1.3.1 Narysuj trzy wykresy, ilustrujące te podziały.

1.3.2 Spośród tak uzyskanych podziałów wybierz najlepszy.

Dane **earthquake**

[Źródło: J. Ćwik, J. Mielniczuk: *Statystyczne systemy uczące się -ćwiczenia w oparciu o pakiet R*, Oficyna Wydawnicza PW, Warszawa, 2009]

zawierają trzy zmienne:

popn - wskazuje w jakich warunkach zbierano dane: *equake* - zbierane w czasie trzęsienia ziemi, *explosn* - zbierane w czasie wybuchu podziemnego

body - prędkość przemieszczania się fali podziemnej

surface - prędkość przemieszczania się fali powierzchniowej

2.1 Oblicz macierz kowariancji danych oraz środki ciężkości w obu grupach.

2.2 Znajdź współrzędne wektora dyskryminacyjnego.

2.3 Wyskaluj ten wektor tak, aby 100 było wartością progową. Na jaką grupę wskazują punkty, których iloczyn skalarny z wektorem dyskryminacyjnym jest mniejszy niż 100?

2.4 Narysuj wykres zawierający dane (*body*, *surface*) z punktami o różnych kolorach dla grup *equake* i *explosn*. Narysuj na wykresie prostą dyskryminacyjną, oddzielającą obie grupy. Które punkty danych zostały źle zaklasyfikowane?