

Wnioskowanie i klasyfikacja w analizie danych

.....

Projekt dla chętnych. Algorytm k -średnich.

.....

W pliku weight-height.csv zapisano wyniki pomiaru wzrostu i wagi w grupie mężczyzn i kobiet.

Wylosuj 5 obserwacji z grupy mężczyzn i 5 obserwacji z grupy kobiet. Otrzymane wartości zaokrąglaj do całości. Otrzymany zbiór punktów na potrzeby projektu uznamy za zbiór 10 obserwacji, które charakteryzują dwie cechy (wzrost i wagę).

Dokonaj podziału obserwacji na 2 grupy stosując algorytm k -średnich. Obliczenia wykonaj "na piechotę" (przynajmniej 3 pierwsze iteracje). W obliczeniach można wspomagać się komputerem (proszę przygotować plik).

Rezultaty przeprowadzonej analizy porównaj z wynikiem zastosowania metody k -średnich w R.

Raport powinien być napisany w profesjonalnym edytorze typu LaTeX.

Ocenie będzie podlegało Państwa zaangażowanie w wykonanie projektu. Zupełne minimum to zastosowanie metody k -średnich do podziału obserwacji na grupy i porównanie wyników z R). Aby uzyskać wyższą punktację, należy dokonać głębszej analizy problemu, przykładowo można zbadać jak wynik zależy od wartości początkowych czy sprawdzić, jaka jest optymalna liczba klastrów – te fragmenty teorii warto doczytać i również zaprezentować w raporcie (proszę pamiętać o komentarzach i wnioskach). Można również samodzielnie zaimplementować algorytm k -średnich w R. To są tylko propozycje – od Państwa zależy, czy i w jaki sposób pogłębią Państwo samodzielnie wiedzę na temat algorytmu k -średnich i zaprezentują ją w raporcie.

Projekt należy wykonać samodzielnie i przedstawić do oceny na ostatnich zajęciach 14.06.2022.

Z projektu można uzyskać od 0 do 10 punktów (5 za zarobienie minimum i kolejne punkty za zaangażowanie w projekt). Oznacza to, że wykonanie pracy domowej daje Państwu możliwość podniesienia oceny końcowej o 0.5 stopnia. Wykonanie projektu można więc potraktować jako nadrobienie zaległości w punktacji do lepszej oceny lub jako koło ratunkowe.