

Marcin Wardyński

Analiza skupień dla auta_all.sta

Opis metod wyznaczenia skupień

Wszystkie uwagi poczynione na temat metod klasteryzacji danych przy analizie zbioru cars.sta, aplikuje się również i w tym przypadku, więc nie będzie powtórzony.

Wynik ćwiczenia

Zbiór danych z Auta_all.sta jest tak duży, że ich ręczne porównywanie nie jest możliwe, natomiast przyglądając się drzewu utworzonemu metodą Warda można zauważyć, że odcinając dendrogram na odległości ok 50.000 uzyskamy cztery skupienia, czyli dokładnie tyle samo, ile zadaliśmy metodzie k-średnich. Ponieważ i w tym przypadku metryką bazową dla obydwu metod klasteryzacji jest odległość euklidesowa, to skupienia te powinny sobie odpowiadać.