

Eksploracji danych

Lista 3

1. Dla danych `diamonds{ggplot2}[R]` sprawdź jakość szlifu (druga kolumna) diamentu z parametrami (`carat=24`, `depth=55.8`, `table=55`, `price=322`, `x=3.87`, `y=4.01`, `z=2.86`) (kolumny 1, 5, 6, 7, 8, 9, 10 jako zmienne objaśniające) korzystając z metody k -najbliższych sąsiadów z $k = 80$.
2. Dla danych `diamonds{ggplot2}[R]` znajdź optymalne k dzieląc zbiór na część uczącą i testową dla algorytmu k -najbliższych sąsiadów. Na podstawie części uczącej sprawdź jakość szlifu dla diamentu z Zad. 1.
3. Dla danych `Chile{carData}` na podstawie algorytmu k -najbliższych sąsiadów na podstawie cech wiek (`age`), dochód (`income`) i statusquo uzupełnij brakującą cechę `vote` dla obietków, w których brakuje tej cechy. Znajdź najpierw optymalne k . Do znalezienia obiektów z wartością NA użyj funkcji `is.na.data.frame{base}` lub `complete.cases{stats}` i `which{base}`.