

Eksploracja Danych

r.akad. 2015/16

Zadanie projektowe

Celem zadania jest przeprowadzenie analizy, grupowania i klasyfikacji na zbiorze danych z wykorzystaniem poznanych metod i narzędzi. W szczególności wykonaj następujące zadania:

1. Określ liczbę obiektów, liczbę klas, zakresy zmienności poszczególnych atrybutów, ich wartości średnie i odchylenia standardowe dla całego zbioru i w poszczególnych klasach. Wskaż atrybuty o największej i najmniejszej zmienności zgodnie z miarami rozstępu i odchylenia standardowego. Jakie wnioski możesz wyciągnąć z tej analizy ?
2. Oceń wizualnie (analizując wykresy punktowe, macierz wykresów punktowych) czy podział na grupy reprezentowany przez atrybut decyzyjny odpowiada naturalnym skupieniom danych w przestrzeni atrybutów.
3. Przetestuj dostępne klasyfikatory, oceń czy do poprawnej klasyfikacji należy wykorzystać wszystkie atrybuty, czy wystarczy ich podzbiór ? Oceń czy wybrane atrybuty wymagają normalizacji lub standaryzacji. Jeśli tak, to wykonaj ją.
4. Przyjmij sensowną miarę jakości klasyfikacji i znajdź zgodny z nią najlepszy klasyfikator.
5. Dokonaj grupowania danych pomijając atrybut decyzyjny. Wykonaj grupowanie dla różnych liczb grup, znajdź – twoim zdaniem - optymalną liczbę grup. Czy w procesie grupowania konieczne jest wykorzystanie wszystkich atrybutów, czy wystarczy wybrać ich podzbiór ? Czy otrzymany podział jest zgodny z podziałem na klasy zawartym w atrybucie decyzyjnym ? Czy jest zgodny z wnioskami otrzymanymi w punkcie 2 ?

Powyższe zadania wykonaj w środowisku R wykorzystując poznane w trakcie wykonywania polecenia wg. instrukcji, wbudowane oraz otrzymane od prowadzącego funkcje.

Opisując w raporcie efekty analizy danych, podaj wyniki, tok rozumowania i wnioski, zamieść wykresy pokazujące najważniejsze aspekty analizy. Do raportu dołącz skrypt(y) zawierający wszystkie komendy (wraz z komentarzami), które wykonałeś. Skrypt będzie wykorzystany do weryfikacji danych z raportu. Raport wraz ze skryptem należy wgrać do iSOD do końca semestru zimowego.

Marcin Iwanowski
iwanowski@ee.pw.edu.pl