

Klasyfikacja niekompletnych danych za pomocą Naiwnego Klasyfikatora Bayesa

Pliki uczące kompletne: car_learn.arff, mushroom1_learn.arff

Pliki uczące z brakującymi atrybutami (różne dla każdej grupy): car_learn_nr.arff, mushroom1_learn_nr.arff

W plikach tych usunięto ok 5% atrybutów warunkowych. Atrybuty decyzyjne są kompletne.

Zadania do wykonania:

1. Przygotować po 3 zestawy danych uczących kompletnych, w których
 - próbki niekompletne usunięto,
 - brakujące atrybuty zastąpiono wartością najczęstszą w klasie, do której należy próbka,
 - dokonano podziału niekompletnych próbek.

Do przygotowania danych, należy wykorzystać oprogramowanie opracowane w ramach poprzedniego laboratorium.

2. Napisać program realizujący klasyfikację za pomocą naiwnego klasyfikatora Bayesa.

Można założyć, że dane uczące będą kompletne (lub uzupełnione). Natomiast klasyfikator powinien mieć możliwość klasyfikacji próbek z brakującymi atrybutami.

3. Dla danych uczących kompletnych oraz dla każdego z trzech podejść (wymienionych w p. 1):
 - dokonać klasyfikacji danych w pliku testowym kompletnym (car_test.arff, mushroom1_test.arff),
 - dokonać klasyfikacji danych w pliku testowym z brakującymi atrybutami warunkowymi (ok. 5% braków) (car_test_nr.arff, mushroom1_test_nr.arff)

Dla każdego klasyfikowanego zestawu danych testujących, należy określić błąd klasyfikacji (% błędnie klasyfikowanych próbek).

Wyniki przedstawić w tabeli.

Dane uczące	Dane testujące	Błąd [%]
car_learn (kompletne)	car_test	
car_learn_nr (braki usunięte)	car_test	
car_learn_nr (braki wypełnione)	car_test	
car_learn_nr (podział)	car_test	
car_learn (kompletne)	car_test_nr	
car_learn_nr (braki usunięte)	car_test_nr	
car_learn_nr (braki wypełnione)	car_test_nr	
car_learn_nr (podział)	car_test_nr	

Dane uczące	Dane testujące	Błąd [%]
mushroom1_learn (kompletne)	mushroom1_test	
mushroom1_learn_nr (braki usunięte)	mushroom1_test	
mushroom1_learn_nr (braki wypełnione)	mushroom1_test	
mushroom1_learn_nr (podział)	mushroom1_test	
mushroom1_learn (kompletne)	mushroom1_test_nr	
mushroom1_learn_nr (braki usunięte)	mushroom1_test_nr	
mushroom1_learn_nr (braki wypełnione)	mushroom1_test_nr	
mushroom1_learn_nr (podział)	mushroom1_test_nr	

Wnioski