

# Zaawansowane uczenie maszynowe: *ćwiczenia do wykładu 3*

Paweł Cichosz

$x$	$a_1$	$a_2$	$a_3$	$c$	$f$
1	1	1	2	0	2
2	1	2	3	0	5
3	1	3	4	0	8
4	2	1	1	0	3
5	2	2	2	0	4
6	1	1	1	1	1
7	1	2	1	1	3
8	1	3	4	1	6
9	2	2	3	1	7
10	2	3	2	1	5

1. Traktując wszystkie atrybuty jako dyskretne, wyznaczyć prawdopodobieństwa klasy 1 przewidywane za pomocą naiwnego klasyfikatora bayesowskiego wyznaczonego na podstawie podanego zbioru trenującego (z wygładzaniem Laplace'a prawdopodobieństw warunkowych wartości atrybutów w ramach klas) dla przykładów klasy 0 z tego samego zbioru.
2. Wyznaczyć błąd bezwzględny i współczynnik determinacji modelu regresji liniowej przewidującego  $f$  na podstawie  $a_1, a_2, a_3$  uzyskanego po jednokrotnym łącznym przetworzeniu wszystkich przykładów trenujących za pomocą algorytmu spadku gradientu z rozmiarem kroku 0.01 i początkowymi wartościami parametrów 0.01.
3. Wyznaczyć parametry modelu regresji liniowej przewidującego  $f$  na podstawie  $a_3$  za pomocą metody najmniejszych kwadratów.
4. Wyznaczyć parametry modelu regresji liniowej przewidującego  $f$  na podstawie  $a_1, a_2, a_3$  za pomocą metody najmniejszych kwadratów, posługując się wybranymi narzędziami obliczeń numerycznych w celu realizacji operacji mnożenia i odwracania macierzy.

5. Wyznaczyć błąd klasyfikacji modelu liniowo-progowego przewidującego  $c$  na podstawie  $a_1, a_2, a_3$ , uzyskanego po jednokrotnym przetworzeniu kolejno wszystkich przykładów trenujących za pomocą algorytmu *prosty perceptron* z początkowymi wartościami parametrów 0.01.
6. Dobrać parametr  $w_4$  (składnik stały) modelu liniowo-progowego minimalizujący błąd klasyfikacji przy predykcji  $c$  na podstawie  $a_1, a_2, a_3$ , jeśli  $w_1 = -0.3$ ,  $w_2 = 1$  i  $w_3 = -0.5$ .
7. Wyznaczyć błąd zespołowego modelu klasyfikacji uzyskanego przez połączenie dwupoziomowych drzew decyzyjnych (z dwoma poziomami podziałów i liśćmi na trzecim poziomie) do predykcji pojęcia  $c$  na podstawie par atrybutów  $(a_1, a_2)$ ,  $(a_2, a_3)$ ,  $(a_1, a_3)$ , traktując atrybuty  $a_1$  i  $a_2$  jako dyskretne (podziały równościowe) oraz atrybut  $a_3$  jako numeryczny (podziały nierównościowe).
8. Wyznaczyć błąd średniokwadratowy zespołowego modelu regresji uzyskanego przez połączenie dwupoziomowych drzew regresji (z dwoma poziomami podziałów i liśćmi na trzecim poziomie) do predykcji funkcji  $f$  na podstawie par atrybutów  $(a_1, a_2)$ ,  $(a_2, a_3)$ ,  $(a_1, a_3)$ , traktując atrybuty  $a_1$  i  $a_2$  jako dyskretne (podziały równościowe) oraz atrybut  $a_3$  jako numeryczny (podziały nierównościowe).