

# Zaawansowane uczenie maszynowe: *ćwiczenia do wykładu 11*

Paweł Cichosz

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

1. Traktując atrybuty  $a_1, a_2, a_3$  jako numeryczne, wyznaczyć niepewność predykcji modelu regresji logistycznej o parametrach  $-0.6, 2.3, -1.5, -0.2$  (odpowiednio dla atrybutów  $a_1, a_2, a_3$  i składnika stałego) na podstawie entropii.
2. Traktując atrybuty  $a_1, a_2, a_3$  jako numeryczne, wyznaczyć niepewność predykcji modelu regresji logistycznej o parametrach  $-0.6, 2.3, -1.5, -0.2$  (odpowiednio dla atrybutów  $a_1, a_2, a_3$  i składnika stałego) na podstawie marginesu ufności.
3. Wyznaczyć predykcje naiwnego klasyfikatora bayesowskiego używającego jako zbioru trenującego przykładów 1, 2, 6, 7 dla pozostałych przykładów oraz dołączyć do zbioru trenującego dwa spośród nich o największej niepewności predykcji. Wyznaczyć predykcje nowego modelu oraz korelację predykcji prawdopodobieństwa klasy 1 obu modeli dla niewykorzystanych przykładów. W obliczeniach nie stosować wygładzania prawdopodobieństw, a prawdopodobieństwa zerowe zastąpić przez 0.01.
4. Wyznaczyć predykcje naiwnego klasyfikatora bayesowskiego używającego jako zbioru trenującego przykładów 1, 2, 6, 7 dla pozostałych przykładów oraz dołączyć do zbioru trenującego dwa spośród nich o najmniejszej niepewności predykcji, używając klas przewidywanych przez model jako etykiet. Wyznaczyć różnicę

błędu klasyfikacji naiwnego klasyfikatora bayesowskiego utworzonego na pierwotnym i powiększonym zbiorze przykładów. W obliczeniach nie stosować wygładzania prawdopodobieństw, a prawdopodobieństwa zerowe zastąpić przez 0.01.