

Klasyfikacja optymalna Bayesa

Sprawdzenie danych

Obliczenie skrajnych wartości zbioru testowego, analiza danych statystycznych a także obejrzenie histogramów spowodowało odrzucenie ze zbioru testowego próbek numer **186** oraz **641**.

Wybór cech

Przy pomocy programu *plot2features* wyświetliłem rozkłady różnych par cech i wybrałem ostatecznie cechy **drugą** i **trzecią**.

Budowa klasyfikatorów

Klasyfikatory ponumerowałem zgodnie z numeracją metod w instrukcji. Na początku przyjąłem $h_1 = 0,01$ (oczywiście jest ono uwzględniane tylko w klasyfikatorze korzystającym z okna Parzena).

Nr. klasyfikatora	Jakość klasyfikatora
1	0.84978
2	0.97697
3	0.61568

Dobór szerokości okna

Od razu widać, że szerokość okna Parzena jest bardzo źle dobrana, bo jakość klasyfikatora wyszła bardzo słaba, a wręcz niebezpiecznie zbliżyła się do 50%. Tabela przedstawia zmianę jakości klasyfikatora przy zmianie szerokości okna.

h_1	Jakość klasyfikatora
0.01	0.61568
0.001	0.74945
0.0001	0.98958
0.00001	0.99561
0.000001	0.98958
0.0000001	0.71107

Wyraźnie widać, że przy zmniejszaniu szerokości okna jakość klasyfikatora się poprawia do momentu osiągnięcia wartości optymalnej. Dalsze zmniejszanie okna powoduje znów pogarszanie jakości klasyfikatora.

Różne zbiory testowe

Ocena wpływu doboru różnej wielkości zbiorów uczących na klasyfikację zbioru testowego.

Część zbioru train	Klasyfikator 1	Klasyfikator 2	Klasyfikator 3
$\frac{1}{10}$	0.84978		
$\frac{1}{4}$	0.97697		
$\frac{1}{2}$	0.61568		
1	0.61568		

Zmiana prawdopodobieństwa *a priori***Porównanie z klasyfikatorem 1-NN**

Uruchomienie klasyfikatora 1-NN na danych kart dało rezultat:

quality = 0.99561