

## ROB LAB 5-6 – Adrian Wiśniewski

Klasyfikator tworzę wykorzystując głosowanie z wagami zgodnie z regułą Bayesa opisaną na wykładzie. Zgodnie z zaleceniami, zbiór danych jest dzielony na dwie połowy, z których jedna jest wykorzystana do policzenia wag, a druga do przeprowadzenia testów jakości. Problem brakujących próbek i zerowych wartości prawdopodobieństwa w iloczynie rozwiązałem poprzez znalezienie najmniejszej wartości prawdopodobieństwa występującej we wszystkich macierzach prawdopodobieństw wszystkich klasyfikatorów i zamianę wartości zerowych w tych macierzach przez znalezione minimum pomniejszone dziesięciokrotnie. Klasyfikator generuje serię wyników dla różnych wartości progu alfa.

### Wyniki

Zbiór danych można podzielić na następujące rozłączne grupy przykładów:

Grupa przykładów	Liczność	Opis
Złośliwe	0.23%	Klasyfikatory są jednomyślne, ale głosują źle
Złe	0.13%	Żaden klasyfikator nie podał poprawnej klasy
Dobre	4.09%	Co najmniej jeden klasyfikator podał poprawną klasę
Proste	95.55%	Klasyfikatory są jednomyślne i głosują dobrze

Z całego zbioru danych jedynie około 4% (400 sztuk na 10 000) stanowi dane ciekawe z punktu widzenia metaklasyfikacji. Dane należące do grupy prostej będą klasyfikowane przez dowolny rozsądny algorytm poprawnie, natomiast dane złośliwe zawsze będą generowały błąd. Z tego względu skuteczność budowanych metaklasyfikatorów w dużym stopniu zależy od tego jak przykłady zostaną podzielone. W tabeli poniżej przedstawiono wyniki dziesięciu kolejnych klasyfikatorów zbudowanych na losowym podziale zbioru danych względem wartości progu alfa. Tabela zawiera w każdej komórce kolejno wartość minimalną, średnią oraz maksymalną.

Próg (alfa)	Skuteczność	Błąd	Odrzut
	0,5 98,36 - 98,59 - 98,80	1,18 - 1,40 - 1,64	0,00 - 0,01 - 0,02
	0,7 98,32 - 98,53 - 98,74	1,12 - 1,34 - 1,60	0,06 - 0,13 - 0,26
	0,8 98,26 - 98,50 - 98,70	1,10 - 1,32 - 1,60	0,14 - 0,18 - 0,28
	0,9 98,26 - 98,47 - 98,64	1,08 - 1,28 - 1,58	0,16 - 0,25 - 0,34
	0,95 98,24 - 98,41 - 98,60	1,02 - 1,24 - 1,56	0,20 - 0,34 - 0,44
	0,97 98,18 - 98,38 - 98,52	1,00 - 1,21 - 1,54	0,28 - 0,42 - 0,54
	0,99 98,00 - 98,26 - 98,42	0,94 - 1,11 - 1,34	0,48 - 0,63 - 0,78
	0,995 97,88 - 98,19 - 98,36	0,84 - 1,04 - 1,26	0,60 - 0,77 - 1,02
	0,999 97,62 - 98,00 - 98,16	0,66 - 0,88 - 1,06	0,96 - 1,12 - 1,34
	0,9995 97,62 - 97,94 - 98,12	0,64 - 0,85 - 1,04	0,98 - 1,21 - 1,40

*Ilustracja 1: Wyniki testów dla dziesięciu kolejnych prób*

Uśredniając wyniki metoda potrafi zbudować klasyfikatory zmniejszające liczbę błędnych decyzji przy zwiększeniu decyzji wymijających i lekkim pogorszeniu skuteczności klasyfikacji. Wśród zbudowanych klasyfikatorów można znaleźć także klasyfikatory lepsze niż referencyjne np.:

Próg	Skuteczność	Błędy	Decyzje wymijające
0,97	0,9848	0,0106	0,0046

## ***Implementacja***

- `[result] = labmeta(rec_in, tstl_in)` – funkcja fasadowa wykonująca całe laboratorium.
  - `rec_in` – zbiór danych
  - `tstl_in` – etykiety zbioru danych
  - `result` – macierz zawierająca w kolumnach wartości kolejno: próg, skuteczność, błąd, odrzut
- `[labels] = classifyMetaBayes(tvec, tlab, tstv, threshold)` – funkcja budująca model metaklasyfikatora z użyciem reguły Bayesa i klasyfikująca przykłady zgodnie ze zbudowanym modelem
  - `threshold` – wartość progu alfa
  - `labels` – etykiety nadane przez metaklasyfikator