

# Raport - Zaawansowane metody klasyfikacji oraz analiza skupień – algorytmy grupujące i hierarchiczne

Filip Michewicz 282239  
Wiktor Niedźwiedzki 258882

10 czerwca 2025 Anno Domini

## Spis treści

<b>1</b>	<b>Zaawansowane metody klasyfikacji</b>	<b>2</b>
1.1	Rodziny klasyfikatorów/uczenie zespołowe . . . . .	2
1.1.1	k-NN . . . . .	2
1.1.2	Naiwny klasyfikator bayesowski . . . . .	2
1.1.3	Drzewa klasyfikacyjne . . . . .	2
1.2	Gówno . . . . .	2
1.2.1	Jądro liniowe . . . . .	2
1.2.2	Jądro wielomianowe . . . . .	3
1.2.3	Jądro radialne . . . . .	4
1.2.4	Jądro sigmoidalne . . . . .	4

## Spis wykresów

## Spis tabel

1	Średnia poprawa dokładności klasyfikacji za pomocą drzewa klasyfikacyjnego, z podziałem na algorytmy uczenia zespołowego oraz liczbę replikacji . . . . .	2
2	Jądro liniowe . . . . .	3
3	Jądro wielomianowe - wielokrotny podział . . . . .	3
4	Jądro wielomianowe - cross-validation . . . . .	3
5	Jądro wielomianowe - bootstrap . . . . .	3
6	Badanie wpływu stopnia wielomianu na dokładność - wielokrotny podział, najbardziej dokładna kombinacja gamy i kary dla opcji default (stopień 3 . . . . .	4
7	Jądro radialne - wielokrotny podział . . . . .	4
8	Jądro radialne - cross-validation . . . . .	4
9	Jądro radialne - bootstrap . . . . .	4
10	Jądro sigmoidalne - wielokrotny podział . . . . .	5
11	Jądro sigmoidalne - cross-validation . . . . .	5
12	Jądro sigmoidalne - bootstrap . . . . .	5

# 1 Zaawansowane metody klasyfikacji

W pierwszej części zadania zastosujemy algorytmy *ensemble learning* (bagging, boosting i random forest) w celu poprawy dokładności cech klasyfikacyjnych. W drugiej natomiast poznamy i ocenimy nową metodę klasyfikacji - metodę wektorów nośnych (SVM).

Zadanie zostanie wykonane na zbiorze danych *wine*, którego szczegółowy opis znajduje się w poprzednim raporcie.

## 1.1 Rodziny klasyfikatorów/uczenie zespołowe

Wyróżniamy trzy algorytmy uczenia zespołowego (ang. ensemble learning):

- **Bagging** - generujemy B-bootstrapowych replikacji zbioru uczącego, na podstawie których tworzymy B klasyfikatorów. Następnie łączymy je w klasyfikator zagregowany, który przydziela dane cechy do klas za pomocą reguły "głosowania większości" (w przypadku remisu wybiera losowo). Każdy klasyfikator powstaje niezależnie (w sensie takim, że wyniki poprzednich nie mają wpływu na generowanie nowych).
- **boosting** - podobnie jak w bagging, tworzymy klasyfikator zagregowany złożony z wielu pojedynczych klasyfikatorów. Jednak różnica jest taka, że klasyfikatory powstają sekwencyjnie. Na początku każda cecha w zbiorze ma przypisaną taką samą wagę. Z każdą kolejną iteracją natomiast waga zwiększa się dla uprzednio źle sklasyfikowanych przypadków.
- **random forest** (dla drzew klasyfikacyjnych) - metoda podobna do bagging z tą różnicą, że klasyfikatory powstają na podstawie różnych m-elementowych podzbiorach cech (m mniejsze bądź równe wszystkim cechom).

### 1.1.1 k-NN

### 1.1.2 Naiwny klasyfikator bayesowski

### 1.1.3 Drzewa klasyfikacyjne

## MOŻE BYĆ NIEPOPRAWNIE W CHUJ

Tabela 1: Średnia poprawa dokładności klasyfikacji za pomocą drzewa klasyfikacyjnego, z podziałem na algorytmy uczenia zespołowego oraz liczbę replikacji

	1	5	10	20	30	40	50	100
Bagging	19.08	52.50	47.25	64.91	61.25	51.00	66.50	57.71
Random Forest	87.44	88.16	87.55	84.90	84.69	86.26	87.59	84.90
Boosting	73.49	70.44	73.56	74.10	79.94	75.98	75.62	67.32
Średnia	60.01	70.37	69.45	74.64	75.29	71.08	76.57	69.98

## 1.2 Gównio

W tej części przeprowadzona będzie klasyfikacja na podstawie metody wektorów nośnych, z podziałem na różne funkcje jądrowe.

## COŚ O TYM CO TO WOGÓLE JEST

### 1.2.1 Jądro liniowe

Tabela 2: Jądro liniowe

	0.001	0.01	0.1	1	10	100	1000
Wielokrotny podział	37.00000	97.17000	97.67	96.00	97.33000	96.50000	97.50
Cross-validation	33.33000	50.12000	90.07	89.51	89.51000	89.51000	89.51
Bootstrap	38.32000	96.95000	97.17	95.68	96.50000	97.24000	96.49
Średnio	36.21667	81.41333	94.97	93.73	94.44667	94.41667	94.50

### 1.2.2 Jądro wielomianowe

Tabela 3: Jądro wielomianowe - wielokrotny podział

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	38.33	94.67	93.50	94.50	95.17	94.67	93.67	94.33	95.00	96.00
0.01	36.00	96.33	96.17	94.50	95.50	92.00	96.00	92.50	96.17	95.50
0.1	36.67	96.67	95.17	94.83	95.83	94.00	94.00	96.00	95.50	94.50
1	37.67	94.17	94.33	95.17	95.83	95.50	96.83	94.67	95.33	95.33
10	42.17	93.50	94.33	95.50	96.33	95.17	94.83	92.83	95.00	96.00
100	83.67	95.67	95.33	94.17	95.67	95.83	95.33	96.50	95.17	94.67
1000	94.17	95.17	96.33	94.50	95.17	95.33	96.67	94.00	94.83	94.67

Tabela 4: Jądro wielomianowe - cross-validation

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	39.97	96.08	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63
0.01	39.97	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63
0.1	39.97	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63
1	39.97	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63
10	44.41	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63
100	87.71	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63
1000	97.22	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63	96.63

Tabela 5: Jądro wielomianowe - bootstrap

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	37.32	94.98	94.37	94.29	94.99	93.51	93.90	94.12	95.36	93.43
0.01	38.50	95.28	94.76	94.64	94.63	93.34	94.84	93.23	93.96	94.35
0.1	37.80	94.37	94.76	95.04	91.91	94.74	95.83	94.87	95.25	94.58
1	36.54	93.41	94.57	93.31	92.90	94.94	95.91	95.26	93.86	94.78
10	37.50	93.64	94.03	94.91	93.41	94.28	94.52	93.66	94.56	95.10
100	86.86	93.10	94.88	93.02	94.07	94.44	96.11	94.66	94.76	93.18
1000	94.01	93.32	94.20	95.02	95.13	94.83	93.71	94.45	93.73	93.25

Najlepsza gamma: **6.67**, najlepsza kara: **1**.

Badamy tylko na podstawie wielokrotnego podziału, bo tak i chuj.

Tabela 6: Badanie wpływu stopnia wielomianu na dokładność - wielokrotny podział, najbardziej dokładna kombinacja gamy i kary dla opcji default (stopień 3

	2	3	4	5	6	7
Dokładność	89.17	94.5	88	88.67	78.67	80.83

### 1.2.3 Jądro radialne

Tabela 7: Jądro radialne - wielokrotny podział

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	39.80	39.80	39.80	39.8	39.8	39.8	39.8	39.8	39.8	39.8
0.01	39.80	39.80	39.80	39.8	39.8	39.8	39.8	39.8	39.8	39.8
0.1	80.26	39.80	39.80	39.8	39.8	39.8	39.8	39.8	39.8	39.8
1	98.86	57.71	39.80	39.8	39.8	39.8	39.8	39.8	39.8	39.8
10	98.33	62.22	40.92	39.8	39.8	39.8	39.8	39.8	39.8	39.8
100	97.78	62.22	40.92	39.8	39.8	39.8	39.8	39.8	39.8	39.8
1000	97.78	62.22	40.92	39.8	39.8	39.8	39.8	39.8	39.8	39.8

Tabela 8: Jądro radialne - cross-validation

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
0.01	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
0.1	80.16	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
1	98.33	55.03	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
10	98.30	60.62	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
100	96.01	60.62	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87
1000	96.01	60.62	39.87	39.87	39.87	39.87	39.87	39.87	39.87	39.87

Tabela 9: Jądro radialne - bootstrap

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	35.98	37.74	35.78	36.74	36.30	37.49	36.83	38.06	39.16	37.98
0.01	36.71	33.53	37.80	34.96	37.51	35.44	38.83	36.44	38.68	38.00
0.1	52.15	37.54	36.91	35.27	36.29	36.38	36.50	40.14	36.67	34.96
1	97.58	52.92	35.93	37.29	38.14	37.74	37.90	37.48	37.40	39.73
10	97.39	56.63	39.80	37.29	38.15	37.35	40.37	37.17	35.28	37.43
100	96.32	48.20	35.94	38.82	37.32	36.71	37.09	35.21	38.74	38.77
1000	96.24	55.98	40.29	36.59	34.54	35.98	37.07	36.56	34.91	38.60

### 1.2.4 Jądro sigmoidalne

Tabela 10: Jądro sigmoidalne - wielokrotny podział

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	40.00	41.50	35.33	41.67	36.67	39.00	38.00	37.00	36.50	33.67
0.01	36.00	54.17	57.33	65.00	63.67	60.33	57.50	63.67	60.17	61.33
0.1	37.83	89.67	85.17	87.83	86.67	85.83	83.00	86.17	87.33	86.00
1	98.17	84.17	82.67	81.50	81.33	82.67	82.33	80.50	82.17	80.50
10	98.17	78.00	79.83	82.50	79.67	80.50	82.67	79.67	79.17	80.00
100	96.17	79.50	80.50	79.67	77.17	80.67	80.33	78.33	80.83	78.67
1000	95.83	77.50	79.50	77.67	82.83	83.17	76.67	81.17	78.83	80.83

Tabela 11: Jądro sigmoidalne - cross-validation

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	39.90	39.90	39.90	39.90	39.90	39.90	39.90	39.90	39.90	39.90
0.01	39.90	78.66	77.55	77.55	77.55	78.10	78.10	78.10	78.10	78.10
0.1	41.63	89.31	88.17	87.61	87.06	87.06	86.50	85.95	86.54	85.95
1	97.71	79.77	76.37	78.63	78.59	75.82	76.93	75.82	74.67	75.26
10	97.16	83.66	73.07	76.41	80.36	77.52	78.07	77.52	78.69	78.14
100	96.05	81.41	70.82	75.88	79.22	78.07	79.22	78.69	78.10	76.99
1000	96.05	80.85	69.05	76.99	80.33	78.63	78.66	76.99	78.10	77.55

Tabela 12: Jądro sigmoidalne - bootstrap

	0.01	1.12	2.23	3.34	4.45	5.56	6.67	7.78	8.89	10
0.001	33.94	40.15	43.38	38.93	39.81	38.08	36.00	35.72	40.66	37.61
0.01	39.72	47.89	64.26	64.98	57.39	66.66	59.17	55.54	59.22	63.05
0.1	39.79	87.45	82.97	84.81	84.32	83.68	85.41	83.03	86.23	84.63
1	97.57	81.85	80.81	80.73	82.70	78.01	76.39	81.09	79.63	79.56
10	96.63	81.58	81.25	81.30	79.22	82.12	79.19	81.89	76.87	81.71
100	96.31	80.32	76.74	79.12	80.67	75.83	77.48	80.91	77.03	81.38
1000	95.69	82.14	78.57	79.29	76.43	79.87	77.24	78.84	79.63	82.15

elo