

ROB LAB 3-4

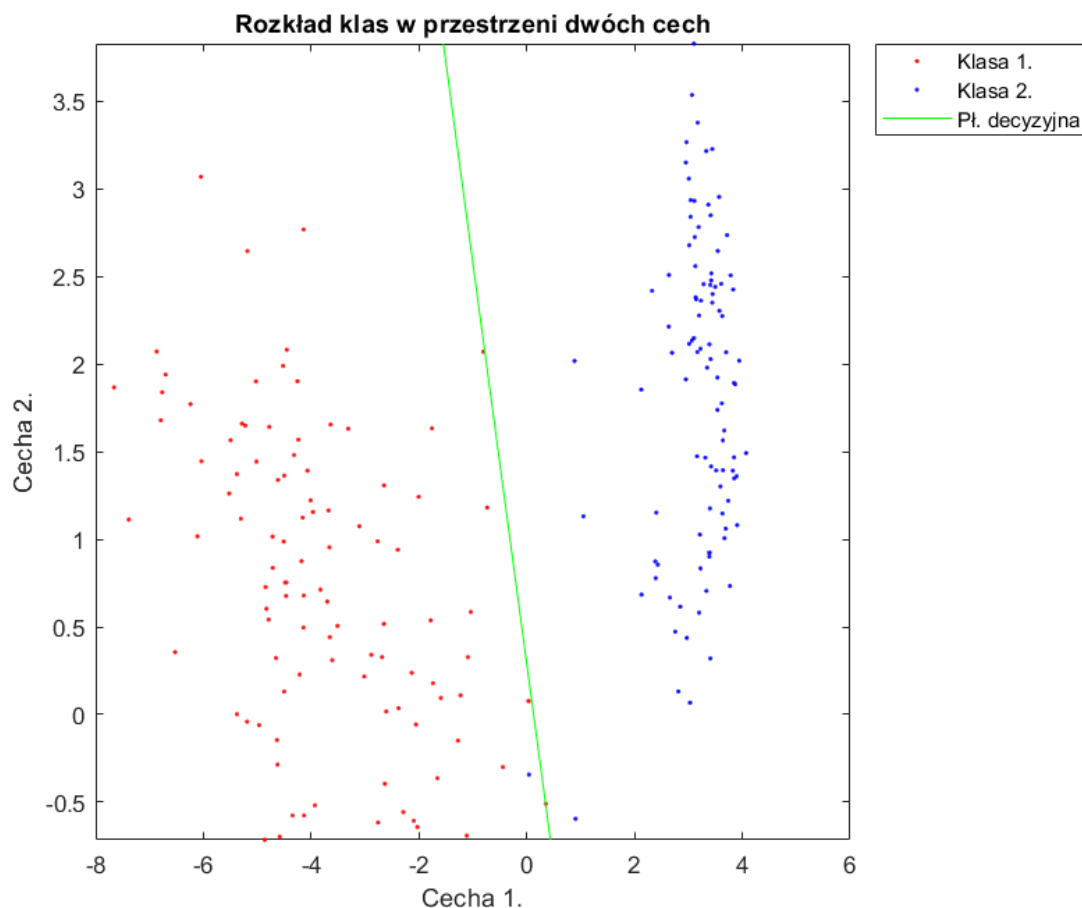
Marcin Hanas

Opis metody klasyfikacji

Do klasyfikacji próbki potrzebne są płaszczyzny, których współrzędne wyznaczane są z wykorzystaniem metody preceptorowej:

1. Losowane są pierwsze współrzędne przestrzeni
2. Sprawdza się, które próbki są błędnie sklasyfikowane (iloczyn skalarny płaszczyzny i próbki jest niedodatni)
3. Dla każdej współrzędnej wyznacza się sumę błędnie sklasyfikowanych próbek i dodaje do odpowiadającej współrzędnej płaszczyzny (po pomnożeniu przez współczynnik uczenia)
4. Sprawdza się, czy jest spełniony warunek stopu (wszystkie próbki poprawnie sklasyfikowane lub 200 iteracji) i jeśli nie, to wraca się do punktu 2.

Sposób działania metody preceptorowej dla 100 próbek klasy 1. i 100 próbek klasy 2. przedstawiono na wykresie:



Płaszczyzny te tworzy się zawsze dla dwóch grup próbek, w klasyfikatorze one vs one są to dwie klasy, a w przypadku klasyfikatora one vs rest – jedna klasa i grupa próbek, które nie należą do tej klasy.

Aby określić klasę próbki, należy zbadać każdym klasyfikatorem, do której klasy należy (poprzez sprawdzenie znaku iloczynu skalarnego płaszczyzny i próbki). Następnie odbywa się głosowanie klasyfikatorów. W przypadku one vs one, próbka jest sklasyfikowana do danej klasy, jeśli wszystkie klasyfikatory, w których występuje ta klasa, zagłosowały na przydzielenie próbki do danej klasy. W komitecie klasyfikatorów one vs rest, próbka jest sklasyfikowana do danej klasy, gdy tylko jeden klasyfikator (gdzie klasa ta jest klasą dodatnią) zagłosuje na przydział próbki do swojej klasy.

Wyniki klasyfikacji

Poniżej przedstawiono wyniki klasyfikacji komitetem klasyfikatorów one vs one oraz one vs rest ze standardową (400) oraz rozszerzoną (780) liczbą cech.

Klasyfikator one vs one

Klasa dodatnia	Klasa ujemna	Fałszywie sklasyfikowane		
		Do klasy ujemnej	Do klasy dodatniej	Łącznie
1	2	7	7	14
1	3	65	68	133
1	4	42	44	86
1	5	20	22	42
1	6	130	53	183
1	7	59	48	107
1	8	21	23	44
1	9	55	56	111
1	10	28	40	68
2	3	60	59	119
2	4	61	53	114
2	5	26	22	48
2	6	29	29	58
2	7	18	12	30
2	8	34	33	67
2	9	144	74	218
2	10	34	31	65
3	4	210	179	389
3	5	99	130	229
3	6	155	155	310
3	7	140	140	280
3	8	92	102	194
3	9	219	199	418
3	10	85	92	177
4	5	35	30	65
4	6	312	276	588
4	7	33	34	67
4	8	94	90	184
4	9	256	236	492
4	10	130	130	260
5	6	60	61	121
5	7	68	66	134
5	8	76	67	143
5	9	46	48	94
5	10	235	235	470
6	7	125	144	269
6	8	50	53	103
6	9	268	259	527

6	10	106	97	203
7	8	10	15	25
7	9	79	60	139
7	10	14	20	34
8	9	65	57	122
8	10	312	290	602
9	10	140	133	273

Numer klasy	1	2	3	4	5	6	7	8	9	10
Zbiór uczący	0.9569	0.9637	0.9006	0.8871	0.9223	0.8690	0.9420	0.9267	0.8718	0.8815
Zbiór testujący	0.9582	0.9683	0.9147	0.9059	0.9246	0.8599	0.9322	0.9086	0.8747	0.8751

	Poprawnie sklasyfikowane	Błędnie sklasyfikowane	Niesklasyfikowane
Zbiór uczący	0.9133	0.0551	0.0316
Zbiór testujący	0.9134	0.0554	0.0312

Klasyfikator one vs rest

Klasa dodatnia	Klasa ujemna	Fałszywie sklasyfikowane		
		Do klasy ujemnej	Do klasy dodatniej	Łącznie
1	Próbki spoza klasy I	380	280	660
2		330	301	631
3		1027	576	1603
4		1128	744	1872
5		861	621	1482
6		3114	961	4075
7		582	445	1027
8		670	509	1179
9		1897	1133	3030
10		876	2557	3433

Numer klasy	1	2	3	4	5	6	7	8	9	10
Zbiór uczący	0.9358	0.9511	0.8159	0.8008	0.8497	0.4038	0.8494	0.8557	0.6387	0.7897
Zbiór testujący	0.9541	0.9577	0.8023	0.8356	0.8442	0.4249	0.8382	0.8512	0.6786	0.7919

	Poprawnie sklasyfikowane	Błędnie sklasyfikowane	Niesklasyfikowane
Zbiór uczący	0.7950	0.0756	0.1294
Zbiór testujący	0.8040	0.0703	0.1257

Klasyfikator one vs one z rozszerzoną liczbą cech

Klasa dodatnia	Klasa ujemna	Fałszywie sklasyfikowane		
		Do klasy ujemnej	Do klasy dodatniej	Łącznie
1	2	2	1	3
1	3	3	10	13
1	4	3	1	4

1	5	8	2	10
1	6	4	2	6
1	7	10	10	20
1	8	0	1	1
1	9	7	3	10
1	10	5	8	13
2	3	11	10	21
2	4	4	4	8
2	5	11	6	17
2	6	4	2	6
2	7	5	5	10
2	8	16	17	33
2	9	7	20	27
2	10	11	8	19
3	4	10	17	27
3	5	4	2	6
3	6	3	3	6
3	7	2	3	5
3	8	30	20	50
3	9	8	18	26
3	10	4	7	11
4	5	3	4	7
4	6	23	35	58
4	7	1	1	2
4	8	11	6	17
4	9	20	26	46
4	10	14	19	33
5	6	1	3	4
5	7	3	7	10
5	8	22	6	28
5	9	10	3	13
5	10	24	58	82
6	7	14	9	23
6	8	5	5	10
6	9	3	15	18
6	10	6	7	13
7	8	0	1	1
7	9	9	5	14
7	10	1	1	2
8	9	10	9	19
8	10	34	39	73
9	10	17	15	32

Numer klasy	1	2	3	4	5	6	7	8	9	10
Zbiór uczący	0.9944	0.9923	0.9876	0.9870	0.9894	0.9878	0.9937	0.9860	0.9824	0.9756
Zbiór testujący	0.9714	0.9841	0.9428	0.9396	0.9450	0.9462	0.9541	0.9407	0.9333	0.9217

	Poprawnie sklasyfikowane	Błędnie sklasyfikowane	Niesklasyfikowane
Zbiór uczący	0.9877	0.0062	0.0062
Zbiór testujący	0.9483	0.0197	0.0320

Klasyfikator one vs rest z rozszerzoną liczbą cech

Klasa dodatnia	Klasa ujemna	Fałszywie sklasyfikowane		
		Do klasy ujemnej	Do klasy dodatniej	Łącznie
1	Próbki spoza klasy I	125	126	251
2		137	147	284
3		268	240	508
4		331	301	632
5		225	221	446
6		266	255	521
7		159	157	316
8		280	267	547
9		395	356	751
10		401	342	743

Numer klasy	1	2	3	4	5	6	7	8	9	10
Zbiór uczący	0.9789	0.9797	0.9480	0.9398	0.9553	0.9329	0.9588	0.9408	0.9022	0.8963
Zbiór testujący	0.9786	0.9841	0.9302	0.9376	0.9369	0.9182	0.9301	0.8969	0.8645	0.8751

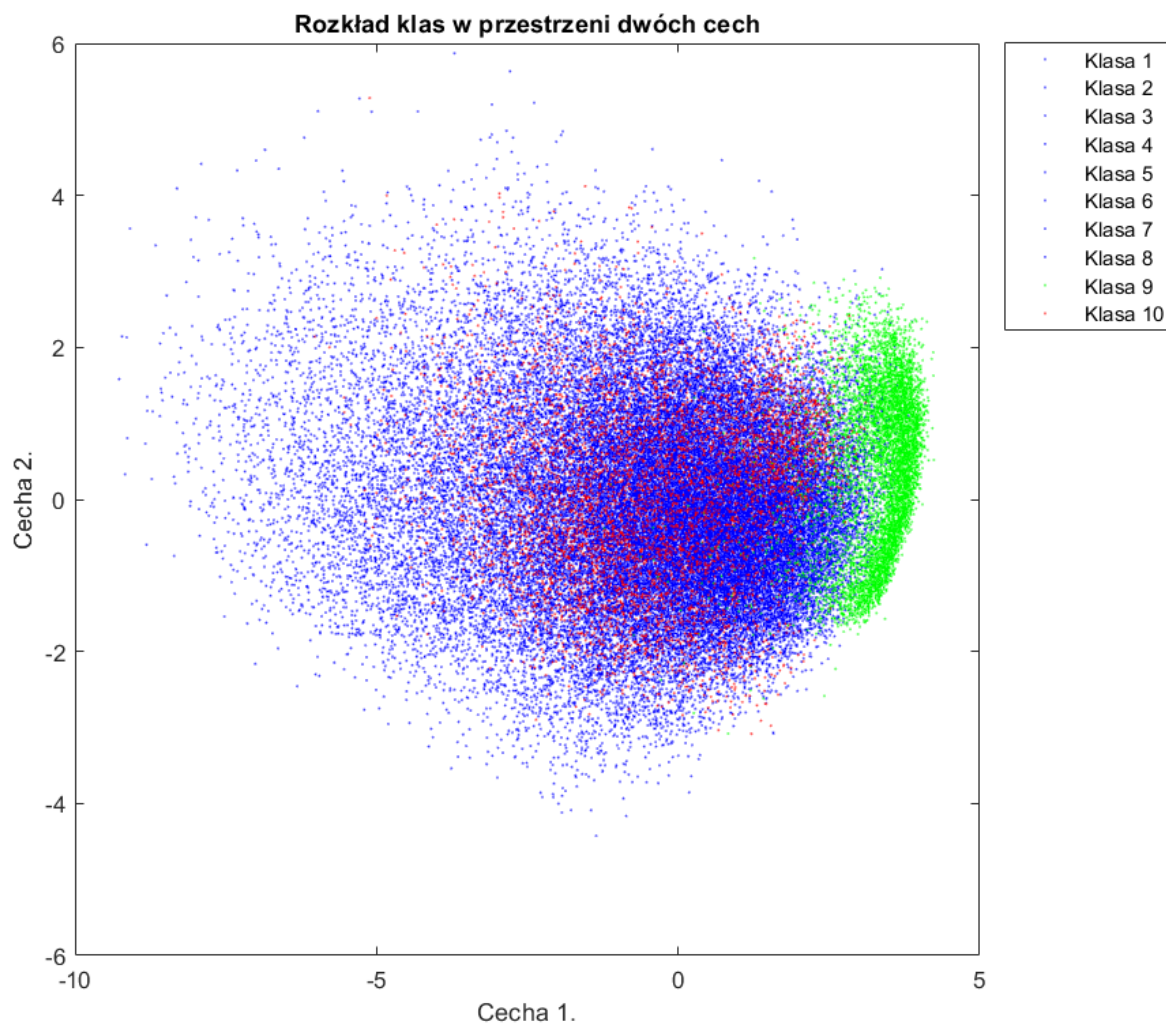
	Poprawnie sklasyfikowane	Błędnie sklasyfikowane	Niesklasyfikowane
Zbiór uczący	0.9438	0.0250	0.0312
Zbiór testujący	0.9260	0.0375	0.0365

Wnioski

Z zebranych i przedstawionych powyżej danych wynika, że grupa klasyfikatorów one vs one daje lepsze wyniki niż one vs rest – zaobserwować można niewielki wzrost liczby próbek błędnie sklasyfikowanych oraz znaczny wzrost liczby próbek z brakiem klasyfikacji – wynika to ze struktury kompanii klasyfikatorów one vs rest – w łatwy sposób powstają w nich obszary, w których jedna próbka należy do kilku klasyfikatorów, wówczas nie da się określić, do której klasy należy.

Rozszerzenie liczby cech poprzez dodanie iloczynów cech powoduje zwiększenie poprawności klasyfikatorów o ok. 2%. Jednocześnie wiąże się to jednak z dużo większą liczbą obliczeń, co znacznie wydłuża czas działania klasyfikatora (zarówno przy wyznaczaniu samych klasyfikatorów, jak i przy późniejszej klasyfikacji próbek).

W przypadku klasyfikacji bez rozszerzania cech, najlepiej sklasyfikowana została klasa 2., a najgorzej 6. Oznacza to, że średnio klasa 2. Była najlepiej separowalna od pozostałych, a klasa 6. – najgorzej. Potwierdzają to obserwacje położenia próbek konkretnych klas na wykresach:

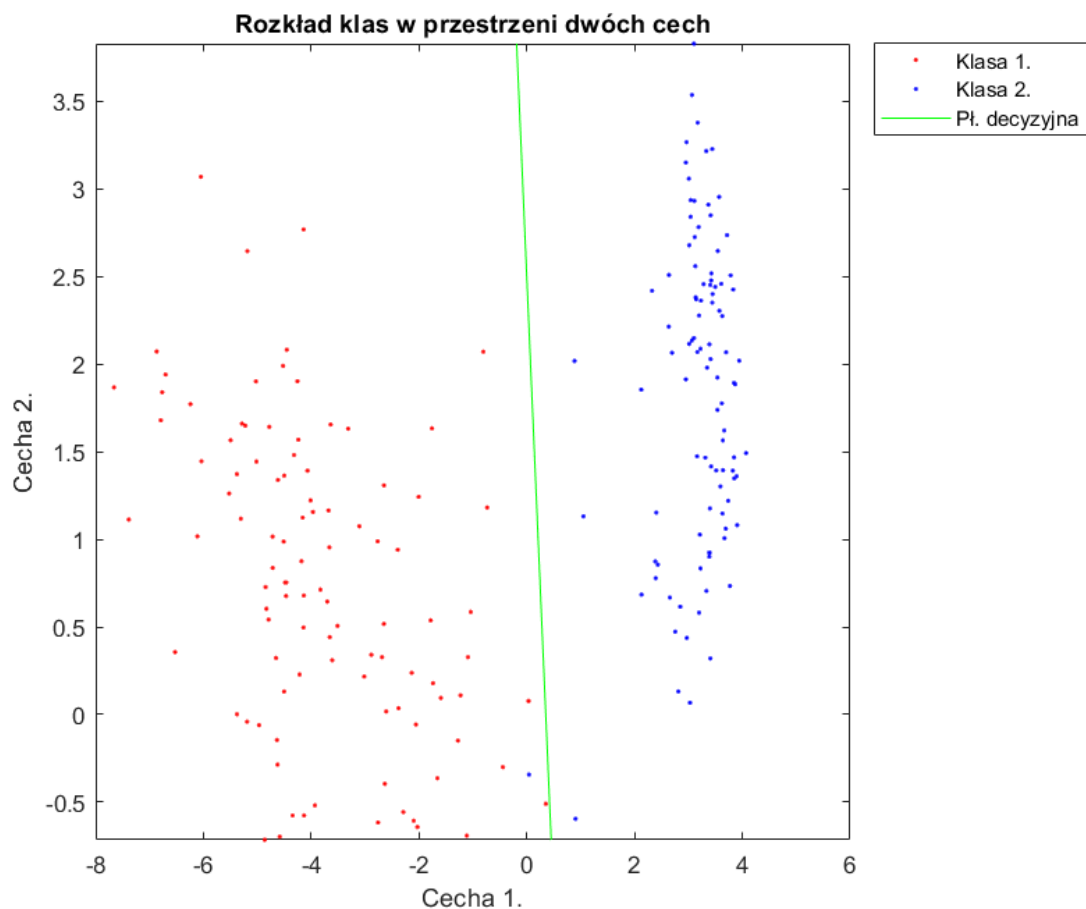


Klasa 2. – zielone punkty, klasa 6. – czerwone punkty.

Modyfikacja algorytmu

Wprowadzoną modyfikacją była drobna zmiana działania perceptronu. Oceniając, czy próbka została błędnie sklasyfikowana, sprawdzane jest, czy iloczyn skalarny próbki i płaszczyzny są mniejsze równe 1, (a nie nieujemne). W ten sposób algorytm próbuje znaleźć taką granicę, która nie tylko dzieli prawidłowo próbki na 2 grupy, ale również jest oddalona o pewną odległość od próbek. Dzięki temu zwiększa się prawdopodobieństwo poprawnego sklasyfikowania próbek, które leżą blisko granicy obu klas.

Wyniki działania zmodyfikowanej metody preceptorowej (dla tych samych próbek, co w przypadku pierwszego testu) przedstawiono na poniższym wykresie:



Jak widać, granica decyzyjna przesunęła się w stosunku do poprzedniego testu perceptronu - wówczas znajdowała się w pobliżu kilku próbek klasy 1., obecnie znajduje się w podobnej odległości od najbliższych próbek obu klas.

Modyfikacja ta poprawia wyniki klasyfikacji dla wszystkich klasyfikatorów o ok. 1%, dla klasyfikatora OVO z 40 cechami (wybranego ze względu na dobre wyniki przy jednoczesnym niedużym nakładzie obliczeń w stosunku do OVO z rozszerzonymi cechami) otrzymano następujące wyniki:

Klasa dodatnia	Klasa ujemna	Fałszywie sklasyfikowane			Zmiana w stosunku do oryginalnych klasyfikatorów
		Do klasy ujemnej	Do klasy dodatniej	Łącznie	
1	2	6	7	13	-1
1	3	46	72	118	-15
1	4	38	37	75	-11
1	5	17	16	33	-9
1	6	84	68	152	-31
1	7	41	49	90	-17
1	8	15	21	36	-8
1	9	48	51	99	-12
1	10	21	34	55	-13
2	3	55	45	100	-19
2	4	53	61	114	0
2	5	24	16	40	-8
2	6	27	29	56	-2

2	7	11	12	23	-7
2	8	20	39	59	-8
2	9	90	154	244	26
2	10	26	28	54	-11
3	4	171	193	364	-25
3	5	111	80	191	-38
3	6	159	122	281	-29
3	7	178	91	269	-11
3	8	89	84	173	-21
3	9	231	156	387	-31
3	10	102	59	161	-16
4	5	42	19	61	-4
4	6	296	265	561	-27
4	7	49	15	64	-3
4	8	92	67	159	-25
4	9	242	198	440	-52
4	10	113	111	224	-36
5	6	59	60	119	-2
5	7	53	59	112	-22
5	8	57	71	128	-15
5	9	30	57	87	-7
5	10	208	250	458	-12
6	7	149	122	271	2
6	8	46	48	94	-9
6	9	219	284	503	-24
6	10	83	119	202	-1
7	8	4	15	19	-6
7	9	62	80	142	3
7	10	11	15	26	-8
8	9	68	53	121	-1
8	10	296	260	556	-46
9	10	127	122	249	-24

Numer klasy	1	2	3	4	5	6	7	8	9	10
Zbiór uczący	0.9726	0.9726	0.9072	0.8943	0.9411	0.8803	0.9547	0.9347	0.8754	0.8880
Wpływ modyfikacji	0.0157	0.0089	0.0066	0.0072	0.0188	0.0113	0.0127	0.0080	0.0036	0.0065
Zbiór testujący	0.9796	0.9824	0.9070	0.9158	0.9481	0.8700	0.9520	0.9212	0.8860	0.8920
Wpływ modyfikacji	0.0214	0.0141	-0.0077	0.0099	0.0235	0.0101	0.0198	0.0126	0.0113	0.0169

	Poprawnie sklasyfikowane	Błędnie sklasyfikowane	Niesklasyfikowane
Zbiór uczący	0.9231	0.0570	0.0199
Wpływ modyfikacji	0.0098	0.0019	-0.0117
Zbiór testujący	0.9265	0.0538	0.0197
Wpływ modyfikacji	0.0131	-0.0016	-0.0115

Jak widać na powyższych wynikach, uzyskano lepsze klasyfikatory w 42/45 przypadków, oraz wszystkie klasy (poza jedną dla zbioru testującego) zostały lepiej sklasyfikowane. Prowadzi to do lepszych ogólnych wyników klasyfikacji (wzrost poprawnie sklasyfikowanych o 1% dla zbioru uczącego i 1,3% dla zbioru testującego).