Politechnika Warszawska

MSI

C5. Uczenie się przez obserwację (indukcja)

Włodzimierz Kasprzak

Zadanie 1. Klasteryzacja cech

Dany jest zbiór 15 próbek uczących 4 klas o postaci 3-wymiarowych wektorów cech. Załóżmy, że ich przynależność do klasy nie jest znana. Zastosować algorytm kwantyzacji z ustalaniem liczby klastrów – wykonać 2 iteracje tego algorytmu. Kwantyzacja korzysta z klasteryzacji k-średnich.

Oszacować wyniki wykonania 2 iteracji algorytmu kwantyzacji.

(Z uwagi na całkowite wartości cech mogą wystąpić zerowe wariancje – wtedy przyjąć wartość wariancji za równą 0.25).

Porównać przynależność próbek do klastrów z ich rzeczywistą klasą, podaną w ostatniej kolumnie tabelki.

Indeks	0	1	2	*Klasa*
1	16	27	10	1
2	16	27	12	1
3	18	29	8	1
4	17	28	8	1
5	18	29	8	1
6	27	43	5	2
7	31	48	7	3
8	23	35	7	4
9	22	34	5	4
10	28	36	6	4
11	29	45	6	4
12	27	43	7	2
13	28	44	5	3
14	28	44	7	3
15	28	44	8	3

Rozwiązanie 1 (1)

1) Jeden klaster.

Środek masy (reprezentant klastra)

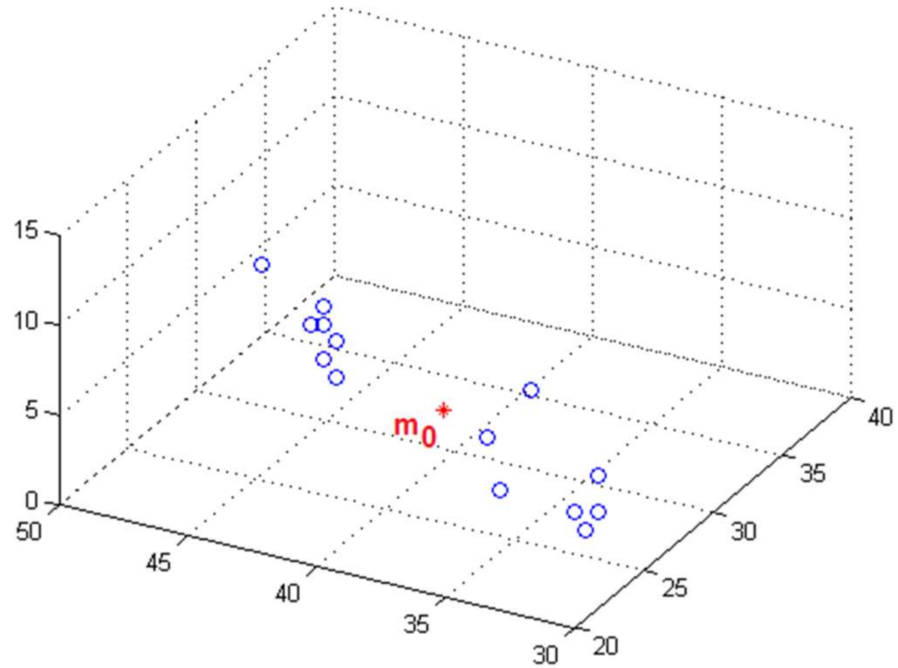
i wartości min-max dla próbek:

$m_0 = [23.7333 \quad 37.0667 \quad 7.2667];$

Min = [16 27 5]

Max= [31 48 12]

Odchyl. std. = [5.4046 7.6855 1.9074]



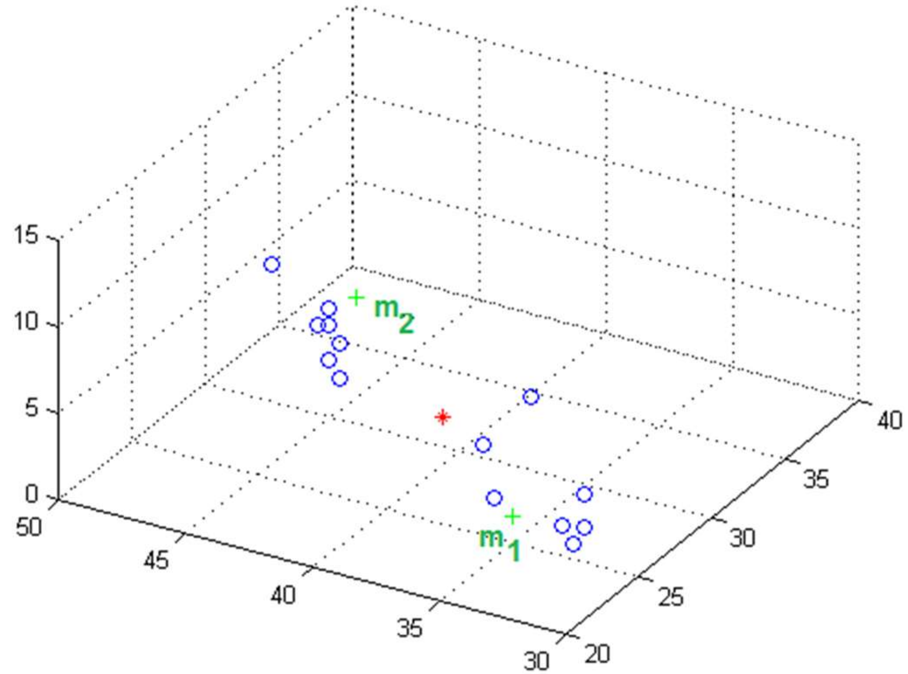
2) Dwa klastry. Inicjalizacja reprezentantów:

$m_1 = m_0\text{-}(m_0 - \text{Min})/\ 2 =$

$[19.8667 \quad 32.0333 \quad 6.1333]$

$m_2 = m_0\text{+}(\text{Max} - m_0)/\ 2 =$

$[27.3667 \quad 42.5333 \quad 9.6333]$



Rozwiązanie 1 (2)

2) Dwie klasy (c.d.)

Iterujemy:

- Klasyfikuj próbki
- Modyfikuj reprezentantów klastrów

Iteracja 1

$m_1(1) = [18.5714 \quad 29.8571 \quad 8.2857]$

$m_2(1) = [28.2500 \quad 43.3750 \quad 6.3750]$

Iteracja 2

Brak zmian → Stop

3) Analiza

Kl. 1: 7 próbek, std duże → podziel na 2

Kl. 2: 8 próbek, std duże → podziel na 2

Index	0	1	2	Class Iter 1	ClassIter 2
1	16	27	10	1	1
2	16	27	12	1	1
3	18	29	8	1	1
4	17	28	8	1	1
5	18	29	8	1	1
6	27	43	5	2	2
7	31	48	7	2	2
8	23	35	7	1	1
9	22	34	5	1	1
10	28	36	6	2	2
11	29	45	6	2	2
12	27	43	7	2	2
13	28	44	5	2	2
14	28	44	7	2	2
15	28	44	8	2	2

Rozwiązanie 1 (3)

4) Cztery klastry

<u>Inicjalizacja</u>

$m_{1n} = m_1 - (m_1 - \mathrm{Min})/\,2 =$

$[17.2857 \quad 28.4286 \quad 6.6429]$

$m_{2n} = m_1 + (m_0 - m_1)/\,2 =$

$[21.1524 \quad 33.4619 \quad 7.7762]$

$m_{3n} = m_2 - (m_2 - m_0)/\,2 =$

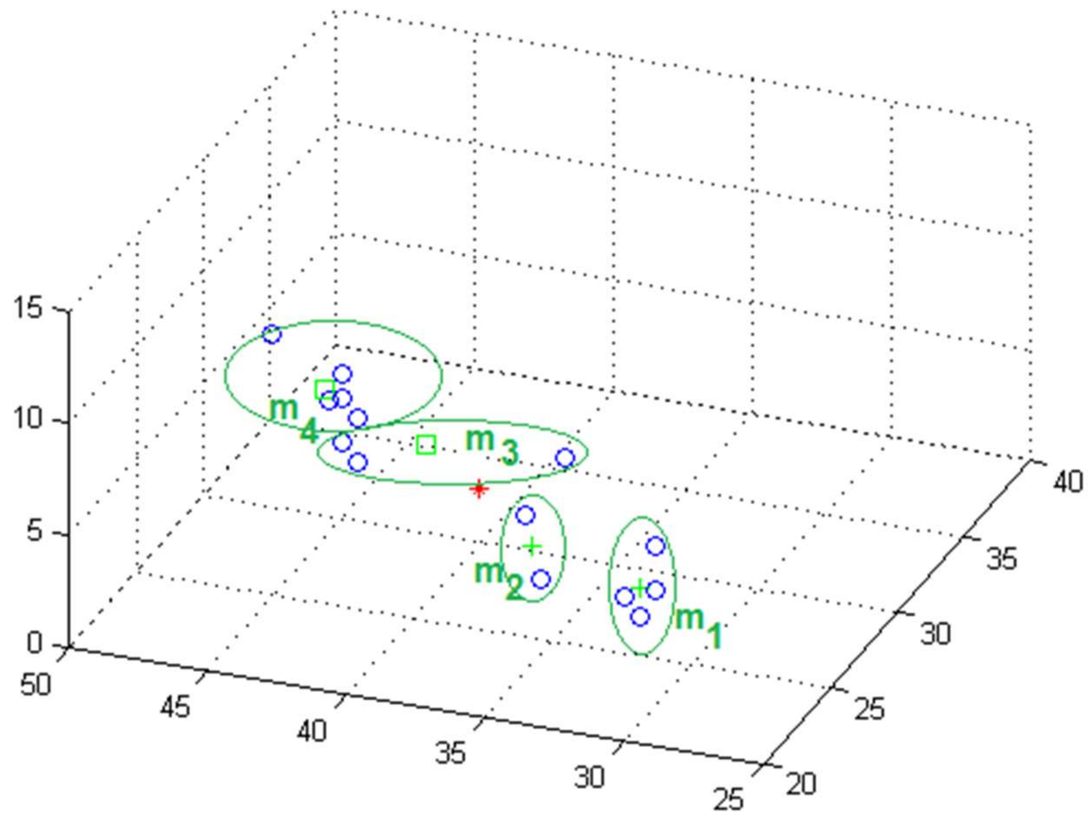
$[25.9917 \quad 40.2208 \quad 6.8208]$

$m_{4n} = m_2 + (\mathrm{Max} - m_2)/\,2 =$

$[\,29.6250 \quad 45.6875 \quad 9.1875]$

<u>Iteracja</u>

- Klasyfikuj próbki
- Modyfikuj reprezentantów



Rozwiązanie 1 (4)

4) Cztery klastry (c.d.)

<u>Iteracja 1</u>

Liczba próbek w klastrze=[5 2 4 4]

$m_1(1)$ =[17.0 28.0 9.2]

$m_2(1)$ =[22.5 34.5 6.0]

$m_3(1)$ =[27.5 41.5 5.75]

$m_4(1)$ =[29.0 45.25 7.0]

<u>Iteracja 2</u>

Liczba próbek(klaster)= [5 2 3 5],

…

$m_3(1)$ =[27.333 40.667 6.0]

$m_4(1)$ =[28.8 45.0 6.6]

<u>Iteracja 3:</u> brak zmian → stop

Index	0	1	2	Class Iter 1	ClassI ter 2
1	16	27	10	1	1
2	16	27	12	1	1
3	18	29	8	1	1
4	17	28	8	1	1
5	18	29	8	1	1
6	27	43	5	3	3
7	31	48	7	4	4
8	23	35	7	2	2
9	22	34	5	2	2
10	28	36	6	3	2
11	29	45	6	4	2
12	27	43	7	3	3
13	28	44	5	3	4
14	28	44	7	4	4
15	28	44	8	4	4

Zadanie 2. Uczenie się drzewa decyzyjnego – DTL i kryterium zysku informacji

Wykonać uczenie drzewa decyzyjnego metodą DTL (Quinlan ID3), w którym wybór testu (atrybutu) dokonuje się na podstawie przyrostu (zysku) informacji, dla podanego obok zbioru uczącego *P*.

Przeprowadzić niezbędne obliczenia.

Próbka x	Atrybut (Test) a1	Atrybut (Test) a2	Atrybut(Test) a3	Klasa c
x_1	1	1	1	1
x_2	1	1	2	1
x_3	1	1	3	1
x_4	2	2	2	1
x_5	2	2	3	1
x_6	1	2	3	2
x_7	1	2	2	2
x_8	1	2	3	2
x_9	2	1	1	2
x_{10}	2	2	1	2

Rozwiązanie 2 (1)

Początkowa entropia: H(P)= H([5/10, 5/10]) = 1.0

1) **<u>Obliczenia dla korzenia drzewa P</u>**

Obliczamy entropię pozostałą dla zbioru P po zastosowaniu każdego z atrybutów (log oznacza logarytm o podstawie 2)

<u>Dla atrybutu a1:</u>

Dla a1=1 mamy 6 elementów (3 są klasy „1" i 3 klasy „2"), dla a1=2 mamy 4 elementy (2 są klasy „1" i 2 klasy „2")

$H_{\textit{zostało}}(P \mid a1) = 6/10\ (-3/6 \log (3/6) - 3/6 \log (3/6)) +$

$+\ 4/10\ (-2/4 \log(2/4) - 2/4 \log(2/4)) = 6/10 * 1 + 4/10 * 1 = 1$

<u>Dla atrybutu a2:</u>

Dla a2=1 mamy 4 elementy (z czego 3 są klasy „1" zaś 1 klasy „2") oraz dla a2=2 mamy 6 elementów (z czego 2 są klasy „1" zaś 4 są klasy „2)

$H_{\textit{zostało}}(\ P \mid a2) = 4/10\ (-3/4 \log (3/4) - 1/4 \log (1/4)) +$

$+\ 6/10\ (-2/6 \log(2/6) - 4/6 \log(4/6)) = 0.8755$

Rozwiązanie 2 (2)

<u>Dla atrybutu a3 :</u>

Mamy 3 elementy „1” (z czego 1 jest klasy „1” zaś 2 klasy „2”), 3 elementy „2” (z czego 2 są klasy „1” zaś 1 klasy „2”), 4 elementy „3” (z czego 2 są klasy „1” zaś 2 są klasy „2”)

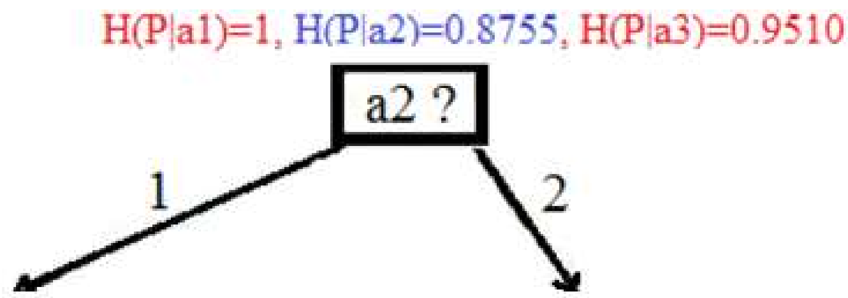
$H_{\textit{zostało}}$ (P | a3) = 3/10 (-1/3 log (1/3) - 2/3 log (2/3)) + 3/10 (-2/3 log(2/3) - 1/3 log(1/3)) + 4/10 (-2/4 log(2/4) - 2/4 log(2/4)) = 0.9510

<u>Podsumowując, w korzeniu drzewa</u> uzysk informacji (zmniejszenie niepewności) wynieść może:

ZI(P | a1) = 1- 1 = 0.000

ZI(P | a2) = 1- 0.8755 = 0.1245

ZI(P | a3) = 1- 0.9510 = 0.049



Wniosek: Największy uzysk daje zastosowanie w korzeniu drzewa atrybutu a2.

Rozwiązanie 2 (3)

<u>2) Węzły poziomu II</u>

Określamy uzysk przy zastosowaniu pozostałych atrybutów a1 i a3 w węzłach (a2=1), (a2=2) powstałych dla wszystkich możliwych wartości atrybutu a2 w korzeniu drzewa:

<u>Węzeł (a2=1) - pozostały 4 elementy</u>

Dla atrybutu a1 mamy 3 elementy „a1=1" (klasy „1" są 3 i klasy 2 jest 0), oraz 1 elementy „a1=2" (z czego 0 jest klasy 1 i 1 jest klasy „2")

$H_{\textit{zostało}}$ (a2=1 | a1) = 3/4 (-3/3 log 3/3) + 1/4 (-2/2 log 2/2) = 0

W zasadzie można już opuścić obliczenia dla a3, gdyż nie będzie większego zysku niż dla a1, ale mimo to sprawdźmy.

Rozwiązanie 2 (4)

Dla atrybutu a3 – mamy 2 elementy „1” (z czego 1 klasy „1” oraz 1 klasy „2”), 1 element „2” (klasy 1) i 1 element „3” klasy „1”

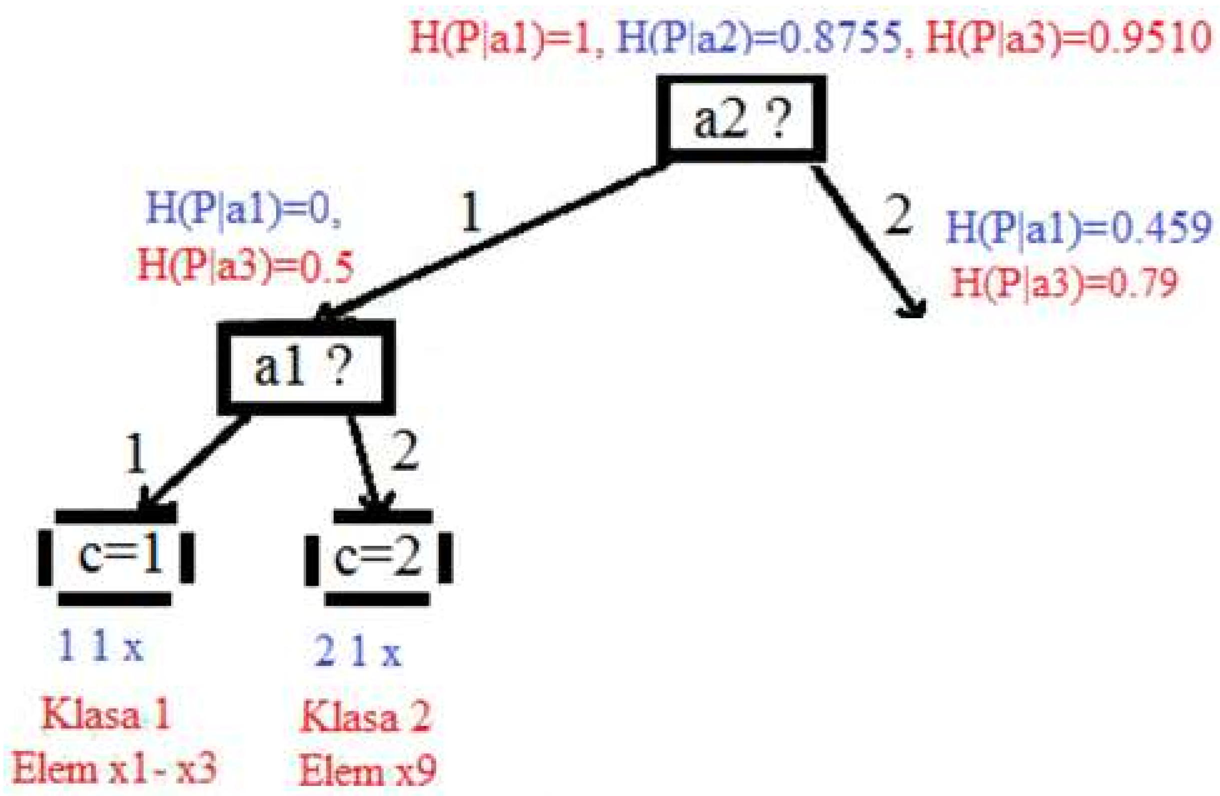
$H_{\textit{zostało}}$ (a2=1 | a3) = 2/4(-1/2 log(1/2) - 1/2 log (1/2)) + 1/4 (-1 log 1) + ¼ (-1 log 1) = 2/4 2*(-1/2 * (-1)) + 0 + 0 = 2/4 * 1 = 2/4 = 0.5

Wniosek: Wybieramy w węźle (a2=1) atrybut a1.

Rozwiązanie 2 (4)

Na poziomie III z węzła (a2=1) wychodzą już tylko dwa liście (a2=1, a1=1) i (a2=1, a1=2) (nie będą dalej analizowane),

gdyż zawierają próbki wyłącznie jednej klasy – dla a1=1 jest to klasa 1, a dla a1=2 - klasa 2.



Rozwiązanie 2 (5)

- <u>Węzeł (a2=2) - jest 6 elementów</u>

Dla atrybutu a1 mamy 3 elementy „1” (wszystkie klasy 2), i 3 elementy „2” (z czego 1 jest klasy „1” oraz 2 elementy klasy „2”)

- $H_{zostało}$(a2=2 | a1) = 3/6 (-1 log 1) + 3/6 (-1/3 log (1/3) -2/3 log (2/3)) = 3/6 * 0 + 3/6 * (1/3 * 1.585 + 2/3 * 0.585)= ½ * 0.91835 = 0.45917

Dla atrybutu a3 mamy 1 element „1” klasy „1” oraz 2 elementy „2” z czego 1 jest klasy „1” i 1 klasy„2” oraz 3 elementy „3” (1 klasy 1 i 2 klasy 2)

$H_{zostało}$ (a2=2 | a3) = 1/6 (-1 log 1) + 2/6 (-1/2 log(1/2) – 1/2 log (1/2) + 3/6 (-1/3 log(1/3) – 2/3 log (2/3))= 0 + 2/6 * 1 + ½ * 0.91835 = 1/3 = 0.79

<u>Wniosek: </u>atrybut a1 prowadzi do mniejszej entropii – daje większy uzysk – wybieramy a1 na węzeł poziomu II.

Rozwiązanie 2 (6)

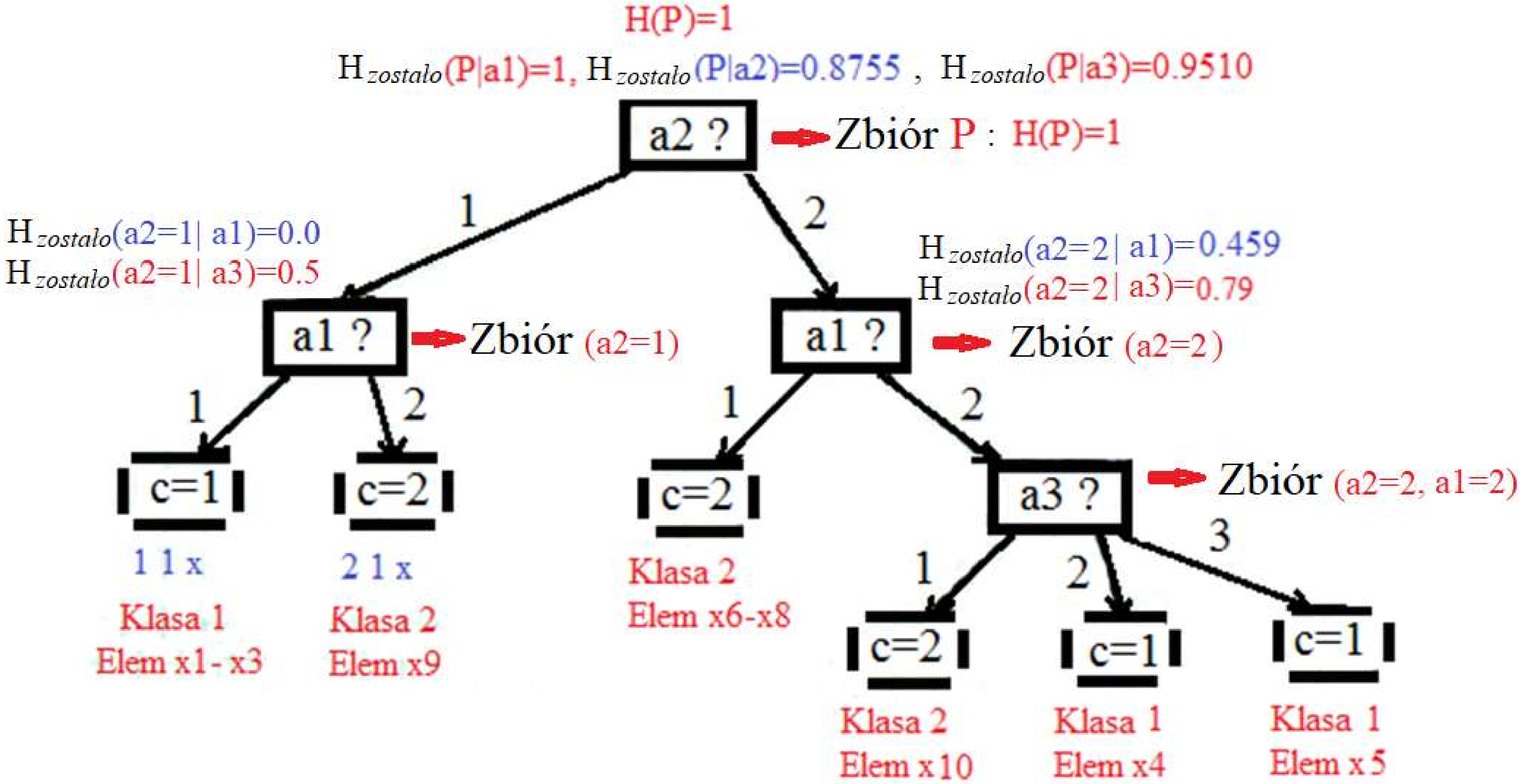
W tej gałęzi drzewa na poziomie III pozostaje już jedynie atrybut a3.

Jednak dla gałęzi (a2=2, a1=1), nie ma on znaczenia – wszystkie pozostałe tu elementy x6-x8 są klasy 2.

Jedynie w gałęzi (a2=2, a1=2) dodany zostaje węzeł z testem a3. Oddziela on od siebie pozostałe elementy x4, x5, x10.

Rozwiązanie 2 (7)

Powstało drzewo decyzyjne o postaci:



Zadanie 3. Uczenie się drzewa decyzyjnego – DTL i iloraz zysku informacji

Wykonać uczenie drzewa decyzyjnego metodą DTL (Quinlan ID3), w którym wybór testu (atrybutu) dokonuje się na podstawie ilorazu zysku informacji, dla podanego obok zbioru uczącego P. Przeprowadzić niezbędne obliczenia.

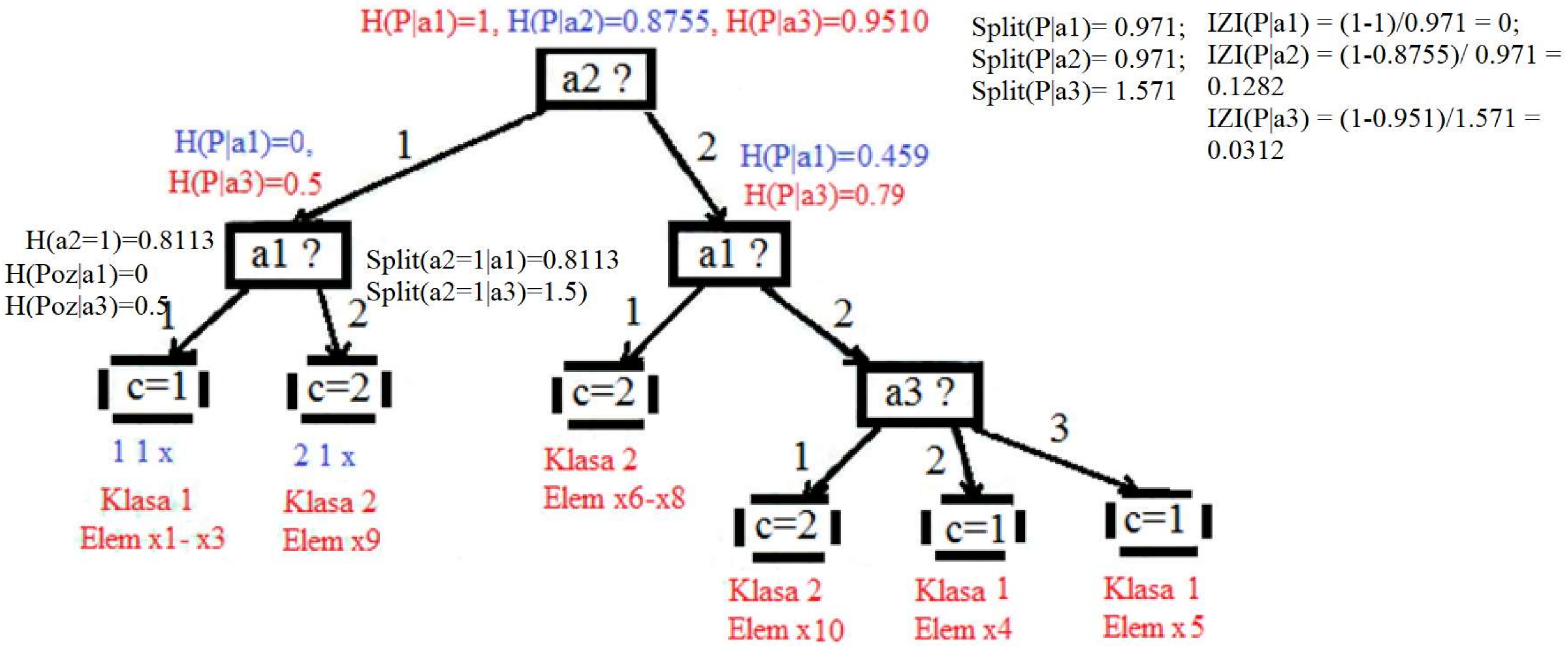
Porównać wynik z wynikiem zadania 2

Próbka x	Atrybut (Test) a1	Atrybut (Test) a2	Atrybut(Test) a3	Klasa c
x_1	1	1	1	1
x_2	1	1	2	1
x_3	1	1	3	1
x_4	2	2	2	1
x_5	2	2	3	1
x_6	1	2	3	2
x_7	1	2	2	2
x_8	1	2	3	2
x_9	2	1	1	2
x_{10}	2	2	1	2

Rozwiązanie 3

Musimy jawnie uwzględnić entropię każdego węzła w drzewie decyzyjnym, gdyż potrzebne jest porównanie zysku informacji normalizowanego entropią wartości każdego z atrybutów dla danego węzła (Split(węzeł | atrybut)).

Okaże się, że drzewo decyzyjne w porównaniu z rozwiązaniem poprzedniego zadania nie ulegnie zmianie.



Zadanie 4. Klasyfikator ML i Bayesa

Nad przestrzenią cech $\Re^2$ ustalić należy (graficznie i analitycznie) rozkłady 3 klas, uwzględniając następujących 10 próbek uczących:

c_k=(x, y)	(1; 2,5)	(0.5; 3)	(0; 3,5)	(2; 3)	(4; 4)	(3; 5)	(1; 1)	(2; 0)	(3; 1)	(2; 2)
Klasa Ω	1	1	1	2	2	2	3	3	3	3

(A) Podać wyniki uczenia dla 3 klasyfikatorów:

1) klasyfikatora geometrycznego minimalnej odległości,
2) klasyfikatora ML (największej wiarygodności)
3) klasyfikatora Bayesa.

(W klasyfikatorach ML i Bayesa przyjąć rozkłady Gaussa prawdopodobieństw warunkowych a priori)

(B) Podać wyniki klasyfikacji próbki (2; 2,5) w powyższych klasyfikatorach

Rozwiązanie 4 (1)

A)

<u>1) Klasyfikator geometryczny</u>

Centroidy (μ_i) obszarów klas:

Klasa „i”	μ_i
1	(0.5; 3)
2	(3; 4)
3	(2; 1)

<u>2) Klasyfikator ML</u> .

Rozkłady warunkowe apriori: $p(\mathbf{c} \mid \text{klasa\_i}) = N(\mu_i\ , \Sigma_i\ ),\ i = 1,2,3$

Klasa „i”	μ_i	**Σ (2D) lub σ_r (1D)**	**Σ^{-1} (macierz odwrotna)**
1	(1/2; 3)	1-wymiarowy: σ_r^2 = 1/3 2D: $\Sigma = \begin{bmatrix} 1/6 & -1/6 \\ -1/6 & 1/6 \end{bmatrix}$	UWAGA: det(Σ)=0 ! Zastosować rozkład 1D zmiennej r: exp(-1/2 (r^2)/σ_r^2)
2	(3, 4)	1-wymiarowy: σ_r^2= 4/3 2D: $\Sigma = \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix}$	det(Σ) = 0.3333 $\Sigma^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$
3	(2; 1)	1-wymiarowy: σ_r^2= 1 2D: $\Sigma = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}$	det(Σ) =0.25 $\Sigma^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

Rozwiązanie 4 (2)

Dla klasy 1 rozkład 2D jest zdegenerowany – wyznacznik macierzy Σ_1 wynosi zero - musimy zastąpić rozkład 2D rozkładem 1D nowej zmiennej „$r(\boldsymbol{c})$" – odległością punktu $\boldsymbol{c}$ od centroidy (środka masy) punktów danej klasy. W celu uczciwego porównywania wyników klasyfikacji dla trzech klas określimy 1-wymiarowe rozkłady warunkowe a priori dla wszystkich trzech klas:

$$r_i(c) = \left\| c - \mu_i \right\|, \; i = 1,2,3$$

$$r_i([x, y]) = \sqrt{(x - \mu_i)^2 + (y - \mu_i)^2}$$

Otrzymamy rozkłady warunkowe a priori zmiennej r:

$$p_i(r \mid \mu_i, \sigma_i) = N(0, \sigma_i), \; i = 1,2,3$$

$$N(0, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{r^2}{2\sigma^2} \right)$$

Wyznaczamy wartości $r^2(\boldsymbol{c}_k)$:

$\boldsymbol{c}_k$=(x, y)	(1; 2.5)	(0.5; 3)	(0; 3.5)	(2; 3)	(4; 4)	(3; 5)	(1; 1)	(2; 0)	(3; 1)	(2; 2)
μ_i	(0.5; 3)	(0.5; 3)	(0.5; 3)	(3; 4)	(3; 4)	(3; 4)	(2; 1)	(2; 1)	(2; 1)	(2; 1)
r_k^2	0.5	0	0.5	2	1	1	1	1	1	1

Rozwiązanie 4 (3)

Stąd wariancje rozkładów dla trzech klas wynoszą:

$$\sigma_1^2(r) = \frac{1}{3}(0.5 + 0 + 0.5) = \frac{1}{3}$$

$$\sigma_2^2(r) = \frac{1}{3}(2 + 1 + 1) = \frac{4}{3}$$

$$\sigma_3^2(r) = \frac{1}{4}(1 + 1 + 1 + 1) = 1$$

3) Klasyfikator Bayesa

Dodatkowo do definicji rozkładów warunkowych a priori, jak dla klasyfikatora ML powyżej, potrzebny jest rozkład a priori prawdopodobieństwa klas:

$$p(\text{klasa}) = [3/10,\ 3/10,\ 4/10]\ .$$

Rozwiązanie 4 (4)

B) Klasyfikacja próbki, $\boldsymbol{c} = (2, 2.5)$

1) Klasyfikator geometryczny z miarą euklidesową:

- Klasa 1: $r_1^2(\boldsymbol{c}) = |(2, 2.5) - (0.5, 3)| = 1.5^2 + 0.5^2 = 2.25 + 0.25 = 2.5$
- Klasa 2: $r_2^2(\boldsymbol{c}) = |(2, 2.5) - (3, 4)| = 1 + 1.5^2 = 3.25$
- Klasa 3: $r_3^2(\boldsymbol{c}) = |(2, 2.5) - (2, 1)| = 0 + 1.5^2 = 2.25$

Reguła decyzyjna:

$$\arg\min_i r_i^2 = \arg\min\,[2.5, 3.25, 2.25] = 3$$

Zwycięża klasa 3.

2) Klasyfikator ML.

Reguła decyzyjna:

$$\arg\max_i p_i(r_i(c)\,|\,(0,\sigma_i))$$

$$\arg\max_i \frac{1}{\sigma_i\sqrt{2\pi}}\cdot\exp\left(-\frac{r_i^2}{2\sigma_1^2}\right) = \frac{1}{\sqrt{2\pi}}\arg\max_i \frac{1}{\sigma_i}\cdot\exp\left(-\frac{r_i^2}{2\sigma_1^2}\right)$$

Dla klasy 1: $0.02351 \cdot 1.732 = 0.040719$

gdzie $\exp\left(-\frac{2.5}{2/3}\right) = \exp(-3.75) \cong 0.02351$ $\quad \frac{1}{\sigma_1} = \frac{1}{\sqrt{1/3}} \cong 1.732$

Rozwiązanie 4 (5)

Dla klasy 2: $0.2956 \cdot 0.866 = 0.2559896$

gdzie $\exp\left(-\dfrac{3.25}{2 \cdot 4/3}\right) = \exp(-1.21875) \cong 0.2956$ $\qquad \dfrac{1}{\sigma_2} = \dfrac{1}{\sqrt{4/3}} \cong 0.866$

Dla klasy 3: $0.3246 \cdot 1 = 0.3246$

gdzie $\exp\left(-\dfrac{2.25}{2 \cdot 1}\right) = \exp(-1.125) \cong 0.3246$ $\qquad \dfrac{1}{\sigma_r} = \dfrac{1}{\sqrt{1}} = 1$

Stąd:

$$\arg\max_i\,[0.04071, 0.25599, 0.32460] = 3$$

<u>Zwycięża klasa 3.</u>

3) Klasyfikator Bayesa

Uwzględniamy dodatkowo do ML rozkład a priori klas: $p(i) = [0.3, 0.3, 0.4]$

Stąd:

$$\arg\max_i\,[0.3 \cdot 0.04071, 0.3 \cdot 0.25599, 0.4 \cdot 0.32460] = 3$$

<u>Zwycięża klasa 3.</u>

Zadanie 5. Klasyfikator k-NN

Nad przestrzenią cech $\Re^2$ ustalić należy (graficznie) rozkłady 4 klas, uwzględniając następujących 11 próbek uczących:

(x,y)	(1;2,5)	(1; 3)	(1;3,5)	(2; 3)	(4; 3)	(2; 2)	(1; 1)	(2; 0)	(3; 1)	(3,5;1,5)	(4; 1,5)
Klasa	1	1	1	2	2	3	3	3	3	4	4

- (A) Narysować diagram Voronoia uzyskany w procesie uczenia klasyfikatora według k=4 sąsiadów (4-NN).
- (B) Podać wynik klasyfikacji wektora cech (2; 2,5) w tym klasyfikatorze.

Zadanie 6. Klasyfikator SVM

A) Zdefiniować problem uczenia klasyfikatora SVM dla podanego niżej zbioru 2-wymiarowych cech i dwóch klas,

B) Podać spodziewane w tym przypadku rozwiązanie w graficzny sposób,

C) Rozwiązać problem analitycznie lub algorytmicznie i podać dokładne rozwiązanie.

Zadany jest następujący zbiór próbek uczących:

Indeks „i”	1	2	3	4	5	6
$C_i = (c_{i1}, c_{i2})$	(2; 2)	(1; 4)	(1; 6)	(2; 1)	(4; 2)	(5; 2)
Klasa Ω_k =	Ω_1	Ω_1	Ω_1	Ω_2	Ω_2	Ω_2

Rozwiązanie 6 (1)

A) Postać pierwotna („prymalna”) problemu optymalizacji dla procesu uczenia liniowego klasyfikatora SVM.

Minimalizacja kwadratowej funkcji celu: $\frac{1}{2}\mathbf{a}^T\mathbf{a}$

przy ograniczeniach liniowych: $y_i(\mathbf{a}^T \cdot \mathbf{c}_i + a_0) \geq 1\,;\; i = 1,\ldots,N$

Gdzie $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ – wektor normalny do hiperpłaszczyzny.

Szukane parametry: wektor normalny $\mathbf{a}$ i wyraz wolny a_0 .

W tym zadaniu konkretnie: $n = 2$, h=3 → $\hat{\mathbf{a}} = (a_0, a_1, a_2)$, $\mathbf{a} = (a_1, a_2)$

N=6 próbek: $i = 1, 2, \ldots 6$.

Optymalizacja oznacza, że przy zadanych ograniczeniach maksymalizowany jest też „margines rozwiązania”: $\tau = 1 / |\mathbf{a}|$.

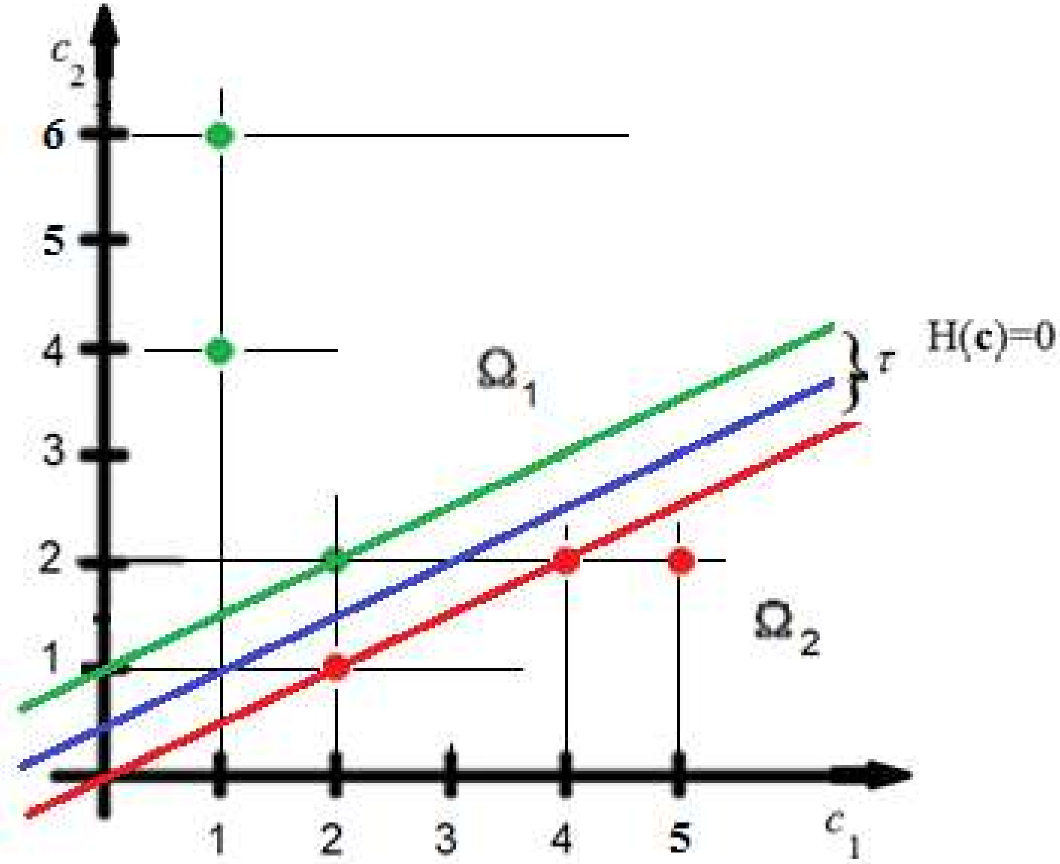
Hiperpłaszczyzna – w tym przypadku prosta skierowana – separuje obszary 2 klas w przestrzeni cech: $\mathrm{H}(\mathbf{c})$: $a_0 + \mathbf{a}^\mathrm{T}\mathbf{c} = 0$, gdzie wektor cech $\mathbf{c}$=(c_1, c_2).

Rozwiązanie 6 (2)

B) Rozwiązanie graficzne

Poszukujemy 3 wektorów wspierających: 2 z nich należą do jednej klasy a 1 do klasy drugiej. Wyznaczają one dwie wzajemnie równoległe proste wspierające: $H_{+1}(c) = 1$, i $H_{-1}(c) = -1$. Pomiędzy tymi prostymi nie leżą żadne inne próbki uczące.

Odstęp pomiędzy prostymi jest największy spośród możliwych rozwiązań tego typu.



Rozwiązanie 6 (3)

<u>C) Rozwiązanie analityczne – wykorzystamy rozwiązanie ogólnego problemu programowania kwadratowego</u>

Np. funkcja **quadprog** w MATLAB-ie służy do rozwiązania ogólnego problemu optymalizacji kwadratowej o postaci:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^T \mathbf{H}\mathbf{x} + \mathbf{f}^T \mathbf{x}$$

przy ograniczeniach: $\mathbf{A}\mathbf{x} \le \mathbf{b}$.

Szukany jest wektor $\mathbf{x}$.

Aby wyrazić problem uczenia klasyfikatora SVM jako specyficzne **quadprog** należy zdefiniować macierze $\mathbf{H}$, $\mathbf{A}$ i wektory $\mathbf{f}$, $\mathbf{b}$,

dla wywołania funkcji: **x = quadprog**($\mathbf{H}, \mathbf{f}, \mathbf{A}, \mathbf{b}$) .

Problem uczenia klasyfikatora SVM przedstawimy jako minimalizację funkcji o postaci:

$$\frac{1}{2}[\mathbf{a}\ a_0]\begin{pmatrix} \mathbf{I}_n & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \mathbf{a} \\ a_0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \\ a_0 \end{pmatrix}$$

przy czym:

$$\mathbf{x} = \begin{pmatrix} a_1 \\ a_2 \\ a_0 \end{pmatrix} \qquad \mathbf{H} = \begin{pmatrix} \mathbf{I}_n & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{f} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Rozwiązanie 6 (4)

Ograniczenia dla n=2 i N=6 przyjmują postać:

$$\begin{pmatrix} y_1 & 0 & 0 & 0 \\ 0 & y_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & y_N \end{pmatrix} \begin{pmatrix} {}^1c_1 & {}^1c_2 & 1 \\ {}^2c_1 & {}^2c_2 & 1 \\ \cdots & \cdots & \cdots \\ {}^Nc_1 & {}^Nc_2 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_0 \end{pmatrix} \geq \begin{pmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{pmatrix}$$

gdzie

$$\mathbf{A} = -\begin{pmatrix} y_1 & 0 & 0 & 0 \\ 0 & y_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & y_N \end{pmatrix} \begin{pmatrix} {}^1c_1 & {}^1c_2 & 1 \\ {}^2c_1 & {}^2c_2 & 1 \\ \cdots & \cdots & \cdots \\ {}^Nc_1 & {}^Nc_2 & 1 \end{pmatrix} \qquad \mathbf{b} = -\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Podstawiając próbki uczące otrzymamy konkretnie:

$$\mathbf{A} = -\begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & -1 & & \\ & & & & -1 & \\ & & & & & -1 \end{pmatrix} \begin{pmatrix} 2 & 2 & 1 \\ 1 & 4 & 1 \\ 1 & 6 & 1 \\ 2 & 1 & 1 \\ 4 & 2 & 1 \\ 5 & 2 & 1 \end{pmatrix} \qquad \mathbf{A} = -\begin{pmatrix} 2 & 2 & 1 \\ 1 & 4 & 1 \\ 1 & 6 & 1 \\ -2 & -1 & -1 \\ -4 & -2 & -1 \\ -5 & -2 & -1 \end{pmatrix} = \begin{pmatrix} -2 & -2 & -1 \\ -1 & -4 & -1 \\ -1 & -6 & -1 \\ 2 & 1 & 1 \\ 4 & 2 & 1 \\ 5 & 2 & 1 \end{pmatrix}$$

Rozwiązanie 6 (5)

Rozwiązanie: $\mathbf{x}$ = [-1, 2, -1], tzn. $a_1 = -1$, $a_2 = 2$, $a_0 = -1$

Równanie prostej separującej: $H(\mathbf{c})$: $-1 + (-1)\, c_1 + 2\, c_2 = 0$

Margines wynosi: $\tau = \frac{1}{\|\mathbf{a}\|} = \frac{1}{\sqrt{5}} \approx 0.447$

Zadanie 7. SVM – wybór i porównanie rozwiązań

Zakładamy zbiór uczący dla klasyfikatora SVM w postaci:

j	1	2	3	4	5	6
${}^{j}\mathbf{c} = ({}^{j}c_1, {}^{j}c_2)$	(0; 1)	(1; 3)	(2; 6)	(2; 1)	(3; 3)	(4; 4.5)
Class	**1**	**1**	**1**	**-1**	**-1**	**-1**

Założmy hipotezy istnienia dwóch zbiorów wektorów nośnych klasyfikatora SVM:

$\boldsymbol{A} = \{(0, 1), (1, 3), (3, 3); \quad \boldsymbol{B} = \{(3, 3), (4, 4.5), (0, 1)\}.$

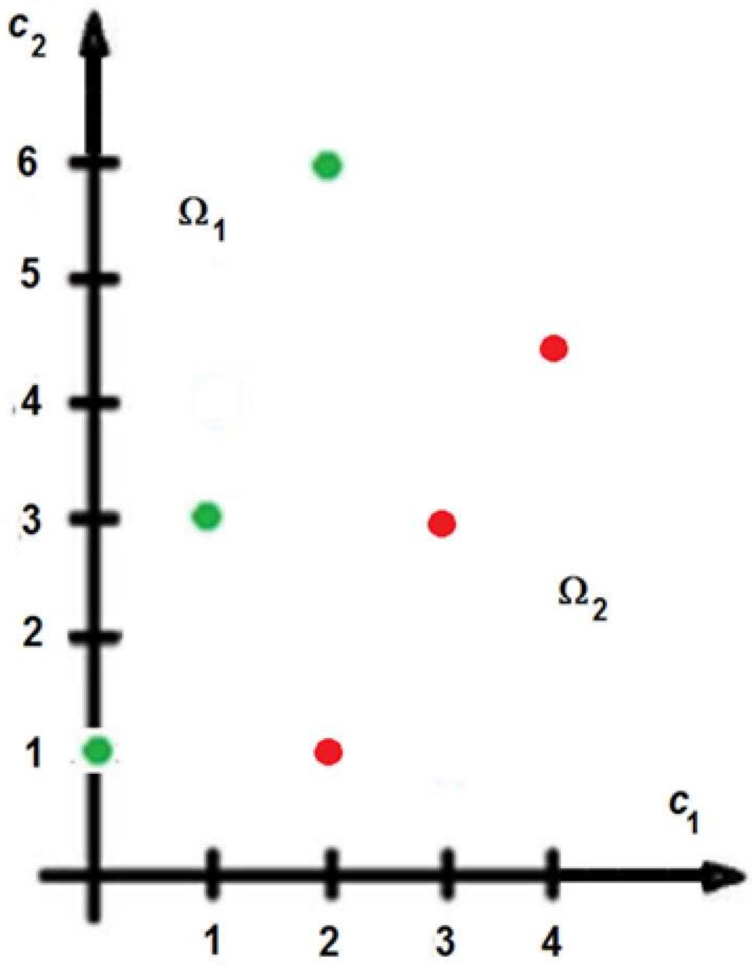
Należy sprawdzić, który z tych zbiorów stanowi potencjalne rozwiązania klasyfikatora SVM.

Zilustrować rozwiązania graficznie.

Wskazówka: należy znaleźć równania prostych separujących, sprawdzić marginesy rozwiązań i spełnianie ograniczeń.

Rozwiązanie 7 (1)

Zbiór próbek uczących:



Rozwiązanie 7 (2)

Zbiór A

1) Prosta separująca

Ograniczenia dla „wektorów nośnych” dają układ trzech równań:

$$y_i(\mathbf{a}^T \cdot \mathbf{c} + a_0) = 1\,; \quad i = 1,2,3\,.$$

$$\begin{cases} 1 \cdot \left([a_1\ a_2] \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} + a_0\right) = 1 \\ 1 \cdot \left([a_1\ a_2] \cdot \begin{bmatrix} 1 \\ 3 \end{bmatrix} + a_0\right) = 1 \\ -1 \cdot \left([a_1\ a_2] \cdot \begin{bmatrix} 3 \\ 3 \end{bmatrix} + a_0\right) = 1 \end{cases}$$

Rozwiązanie 7 (3)

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 3 \\ -1 & -3 & -3 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 3 \\ -1 & -3 & -3 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 3 \\ -1 & -3 & -3 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\boxed{\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ -1 \\ 0.5 \end{bmatrix}}$$

Rozwiązanie 7 (4)

Stąd równanie prostej skierowanej

$$H^*(\mathbf{c}) : \boxed{-1\, c_1 + \frac{1}{2} c_2 + \frac{1}{2} = 0}$$

Graficznie - linia w układzie x-y: $y = 2x - 1$

2) Margines rozwiązania

$$\boxed{d = \frac{1}{|\mathbf{a}|} = \frac{1}{\sqrt{(-1)^2 + 0.5^2}} = \frac{2}{\sqrt{5}} \approx 0.8944}$$

3) Sprawdzenie ograniczeń:

Nr	Próbka (c_1, c_2)	$a_1 c_1 + a_2 c_2 + a_0 = ?$	Powinno być	Ograniczenie spełnione?
1	(0,1)	1	$= 1$	True
2	(1,3)	1	$= 1$	True
3	(2,6)	1.5	≥ 1	True
4	(2,1)	-1	≤ -1	True
5	(3,3)	-1	$= -1$	True
6	(4, 4.5)	-1.25	≤ -1	True

Rozwiązanie 7 (6)

Zbiór B

1) Równanie prostej separującej

$$\begin{cases} -1\cdot([a_1\ a_2]\cdot\begin{bmatrix}3\\3\end{bmatrix}+a_0)=1 \\ -1\cdot([a_1\ a_2]\cdot\begin{bmatrix}4\\4.5\end{bmatrix}+a_0)=1 \\ 1\cdot([a_1\ a_2]\cdot\begin{bmatrix}0\\1\end{bmatrix}+a_0)=1 \end{cases}$$

$$\begin{bmatrix}-1 & -3 & -3\\ -1 & -4 & -4.5\\ 1 & 0 & 1\end{bmatrix}\begin{bmatrix}a_0\\a_1\\a_2\end{bmatrix}=\begin{bmatrix}1\\1\\1\end{bmatrix} \rightarrow \boxed{\begin{bmatrix}a_0\\a_1\\a_2\end{bmatrix}=\begin{bmatrix}0.2\\-1.2\\0.8\end{bmatrix}}$$

Rozwiązanie 7 (7)

Prosta H*(c): $\boxed{-1\frac{1}{5}c_1+\frac{4}{5}c_2+\frac{1}{5}=0}$

Graficznie w układzie x-y: $\boxed{y = 1.5\,x - 0.25}$

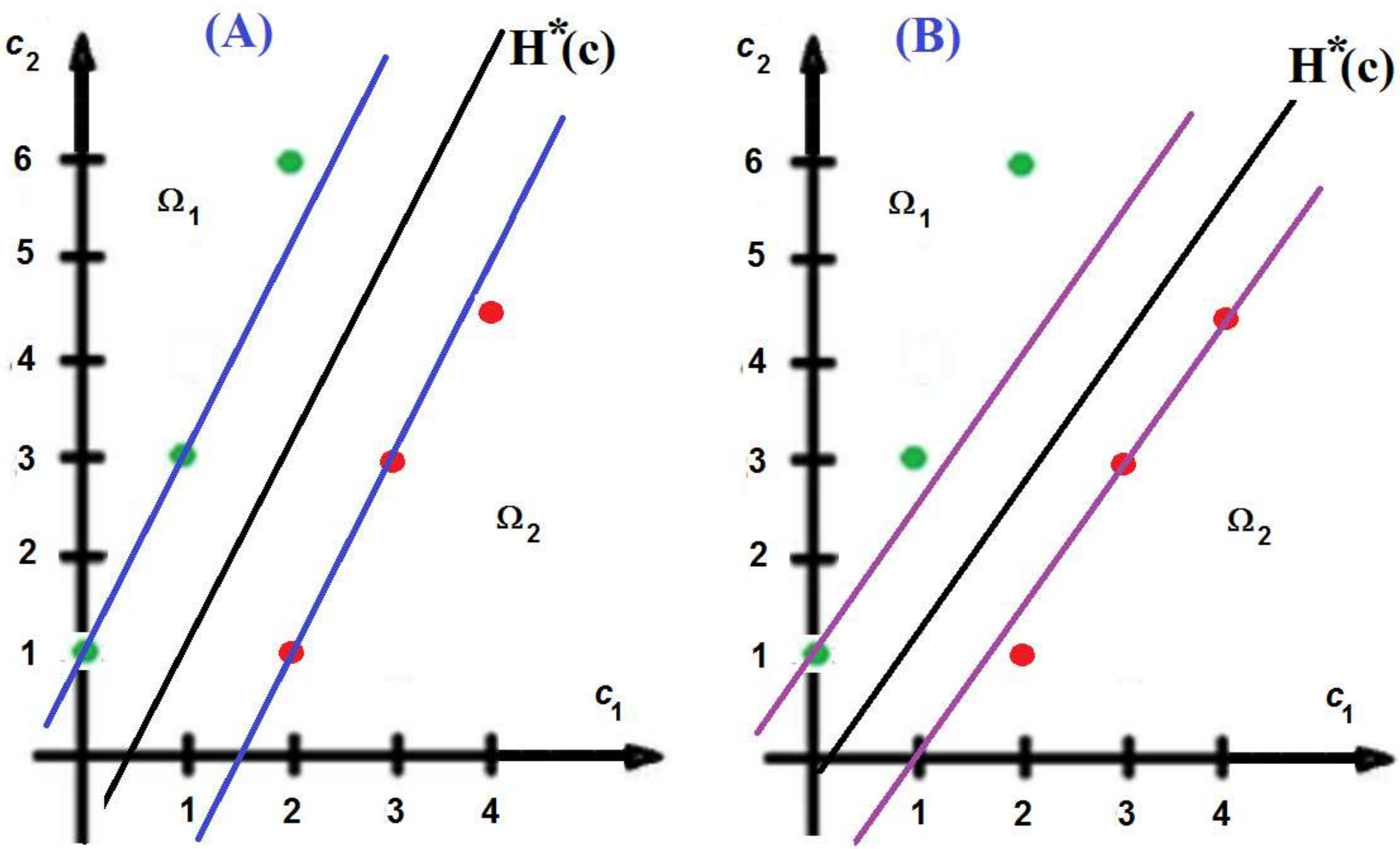
2) „Margines” rozwiązania: $$\boxed{d=\frac{1}{|\mathbf{a}|}=\frac{1}{\sqrt{\left(\frac{-6}{5}\right)^2+\left(\frac{4}{5}\right)^2}}=\frac{5}{\sqrt{52}}\approx 0.6934}$$

3) Sprawdzenie ograniczeń

Nr	Próbka (c_1, c_2)	$\boxed{a_1c_1+a_2c_2+a_0=?}$	Powinno być	Ograniczenie spełnione?
1	(0,1)	1	= 1	True
2	(1,3)	1.4	≥ 1	True
3	(2,6)	2.6	≥ 1	True
4	(2,1)	-1.4	≤ -1	True
5	(3,3)	-1	= -1	True
6	(4, 4.5)	-1	≤ -1	True

Rozwiązanie 7 (5)

Ilustracja obu hipotetycznych rozwiązań:



Rozwiązanie 7 (8)

Odpowiedź

Oba zbiory wyznaczają rozwiązania spełniające ograniczenia, ale zbiór A wyznacza rozwiązanie o większym „marginesie”.

Dlatego potencjalnym rozwiązaniem jest zbiór A.

Zad. 8. Klasyfikator SVM „zakłócany szumem” – analiza możliwych rozwiązań

Zaproponować prostą metodę sprawdzania “zbiorów kandydatów na wektory nośne” dla klasyfikatora SVM, tzn. po wybraniu odpowiedniej liczby próbek ze zbioru uczącego należy znaleźć rozwiązanie (równanie hiperpłaszczyzny $H(\mathbf{c})$), wyznaczyć „margines rozwiązania”, sprawdzić spełnianie ograniczeń (ewentualnie określić i dodać wartość kary ε w przypadku niespełniania ograniczenia przez próbkę).

Kryterium uczenia klasyfikatora SVM („zakłócanego szumem”) wynosi:

$$\min_{\tilde{\mathbf{a}} \in R^{n+1}} f(\tilde{\mathbf{a}}) = \min_{\mathbf{a} \in R^n} \left(\frac{1}{2}|\mathbf{a}|^2 + \mathrm{C}\sum_{j=1}^{N} \varepsilon_j\right)$$

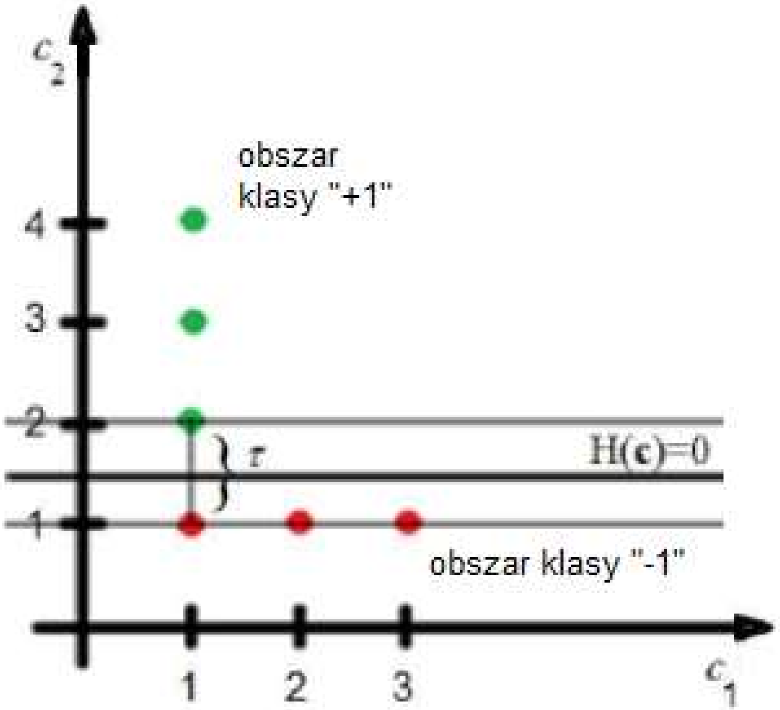
Założyć, że współczynnik kary C wynosi 1, a zbiór uczący obejmuje 6 poniższych próbek (po 3 w klasach +1 i -1), w przestrzeni cech $\Re^2$:

$\mathbf{c}=(c_1,c_2)$	(1; 2)	(1; 3)	(1; 4)	(1; 1)	(2; 1)	(3; 1)
Klasa	+1	+1	+1	-1	-1	-1

Rozw. 8. Analiza rozwiązań „wektorów nośnych” (1)

Ponieważ rozmiar przestrzeni cech, n=2, poszukujemy zbioru 3 (h=n+1) „wektorów nośnych”, z których 2 próbki należą do jednej klasy a jedna do drugiej.

A) Jeśli przyjmiemy, że dwie próbki należą do klasy „+1” to każda ich kombinacja (2 próbek z 3 możliwych) daje hiperpłaszczyznę wspierającą „+1” o równaniu $c_1 = 1$. Możliwe jest dobranie jednego z trzech punktów klasy „-1”.



Rozwiązanie 8 (2)

- Jeśli dobierzemy próbkę [1, 1] to margines rozwiązania wyniesie 0 (obie proste wspierające pokrywają się), ale wszystkie ograniczenia będą spełnione. Jednak: $\boxed{f(\tilde{\boldsymbol{a}}^{(1)}) = \infty}$
- Jeśli dobierzemy próbkę [2, 1] to H(**c**): $-2 \cdot c_1 + 0 \cdot c_2 + 3 = 0$.

Margines rozwiązania wyniesie 0.5, ale dla próbki [1, 1] nie będzie spełnione ograniczenie: H([1, 1])=1, czyli $-1 \cdot 1 = 1 - \varepsilon \Rightarrow \varepsilon = 2$.

Kara wyniesie $C \cdot \varepsilon = 2$, a funkcja celu: $\boxed{f(\tilde{\boldsymbol{a}}^{(2)}) = \frac{1}{2} \cdot 2^2 + 2 = 4}$

- Jeśli dobierzemy próbkę [3, 1] to H(c): $-1 \cdot c_1 + 0 \cdot c_2 + 2 = 0$.

Margines rozwiązania wyniesie 1.0, ale dla próbek [1, 1] i [2,1] nie będzie spełnione ograniczenie:

H([1, 1]) =1, czyli $-1 \cdot 1 = 1 - \varepsilon_1 \Rightarrow \varepsilon_1 = 2$;

H([2, 1]) =0, czyli $-1 \cdot 0 = 1 - \varepsilon_2 \Rightarrow \varepsilon_2 = 1$.

Funkcja celu: $\boxed{f(\tilde{\boldsymbol{a}}^{(3)}) = \frac{1}{2} \cdot 1^2 + 3 = 3.5}$

Rozwiązanie 8 (3)

B) Jeśli przyjmiemy, że dwie próbki należą do klasy „-1” to w tym przypadku każda ich kombinacja (2 próbek z 3 możliwych) daje prostą wspierającą „-1” o równaniu $c_2 = 1$. Możliwe jest dobranie jednego z trzech punktów klasy „+1”.

B1) Niech zbiór próbek-kandydatów jest następujący:

$^{1}\mathbf{c} = (1,2)$, klasy +1

$^{2}\mathbf{c} =(1,1)$, klasy -1

$^{3}\mathbf{c} = (2,1)$, klasy -1

Pomimo, że rozwiązanie w przypadku klasycznego liniowego SVM jest dla nas oczywiste, rozwiążemy jednak ten przypadek w sposób ogólny, dla ilustracji potrzebnego nam algorytmu, a także uwzględnimy rozwiązania nie spełniające wszystkich ograniczeń.

Rozwiązanie 8 (4)

Równanie prostej skierowanej znajdujemy analogicznie jak w zadaniu 6. Ma ono postać, H*(***c***): $0\, c_1 + 2\, c_2 - 3 = 0$

Alternatywnie możemy rozwiązać problem poprzez mnożniki Lagrange'a, korzystając z własności SVM, że hiperpłaszczyzna rozdzielająca obszary klas zadana jest sumą iloczynów skalarnych danej próbki (tu: kolejnych punktów ze zbioru kandydatów) z kandydatami na „wektory nośne":

$$d_{\Lambda,a_0}({}^1\mathbf{c}) = \sum_{j=1}^{3} \lambda_j y_j ({}^1\mathbf{c}^T\, {}^j\mathbf{c}) + a_0 = {}^1y$$

$$d_{\Lambda,a_0}({}^2\mathbf{c}) = \sum_{j=1}^{3} \lambda_j y_j ({}^2\mathbf{c}^T\, {}^j\mathbf{c}) + a_0 = {}^2y$$

$$d_{\Lambda,a_0}({}^3\mathbf{c}) = \sum_{j=1}^{3} \lambda_j y_j ({}^3\mathbf{c}^T\, {}^j\mathbf{c}) + a_0 = {}^3y$$

Niezbędne jest tu dodanie jednego równania, gdyż nieznane jest także a_0 .

Korzystamy z ograniczenia o ważonej sumie mnożników Lagrange'a równej zeru.

$$\sum_{j=1}^{3} \lambda_j y_j = 0$$

Rozwiązanie 8 (5)

Po podstawieniu danych otrzymujemy układ 4 równań z 4 niewiadomymi:

$$\lambda_1 (1,2)^T (1,2) - \lambda_2 (1,2)^T (1,1) - \lambda_3 (1,2)^T (2,1) + a_0 = 1$$

$$\lambda_1 (1,1)^T (1,2) - \lambda_2 (1,1)^T (1,1) - \lambda_3 (1,1)^T (2,1) + a_0 = -1$$

$$\lambda_1 (2,1)^T (1,2) - \lambda_2 (2,1)^T (1,1) - \lambda_3 (2,1)^T (2,1) + a_0 = -1$$

$$\lambda_1 - \lambda_2 - \lambda_3 = 0$$

Po wymnożeniu wektorów układ równań ma postać:

$$\lambda_1 5 - \lambda_2 3 - \lambda_3 4 + a_0 = 1$$

$$\lambda_1 3 - \lambda_2 2 - \lambda_3 3 + a_0 = -1$$

$$\lambda_1 4 - \lambda_2 3 - \lambda_3 5 + a_0 = -1$$

$$\lambda_1 - \lambda_2 - \lambda_3 = 0$$

Rozwiązanie układu (mnożniki Lagrange'a λ_i i a_0):

- $\lambda_1 = 2$, $\lambda_2 = 2$, $\lambda_3 = 0$, $a_0 = -3$

Rozwiązanie 8 (6)

Z warunku K-K-T, na podstawie znanych mnożników Lagrange'a wyliczamy współczynniki kierunkowe szukanej prostej, $\boldsymbol{a} = (a_1, a_2)^T$:

$$\mathbf{a} = \sum_{j=1}^{3} y_j \lambda_j {}^j\mathbf{c}$$

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 2 \end{bmatrix} - 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

Stąd równanie prostej separującej ma postać:

$$0\ c_1 + 2\ c_2 - 3 = 0$$

„Margines" tego rozwiązania:

$$d = \frac{1}{\|\boldsymbol{a}\|} = \frac{1}{\sqrt{a_1^{\ 2} + a_2^{\ 2}}} = \frac{1}{\sqrt{0^2 + 2^2}} = \frac{1}{2}$$

Sprawdzamy ograniczenia dla tego rozwiązania →

Rozwiązanie 8 (7)

Sprawdzamy ograniczenia dla tego rozwiązania:

Nr	c=(c_1, c_2)	$a_1c_1+a_2c_2+a_0=?$	Powinno być	Błąd ε
1	(1,2)	1	= 1	0
2	(1,3)	3	≥ 1	0
3	(1,4)	5	≥ 1	0
4	(1,1)	-1	= -1	0
5	(2,1)	-1	= -1	0
6	(3,1)	-1	≤ -1	0

Wartość funkcji celu:

$$\boxed{f(\tilde{\boldsymbol{a}}^{(4)}) = \tfrac{1}{2} \cdot 2^2 + 0.0 = 2.0}$$

Rozwiązanie 8 (8)

Analogicznie znajdujemy i sprawdzamy pozostałe dwa zbiory potencjalnych wektorów nośnych:

(obliczenia pominięto).

Pozostałe zbiory - próbka klasy „+1” wynosi:

- ${}^{1}c = (1,3) \rightarrow \boxed{f(\tilde{\boldsymbol{a}}^{(5)}) = \tfrac{1}{2} \cdot 1^2 + 1 = 1.5} \rightarrow$ najlepsze rozwiązanie!

lub

- ${}^{1}c = (1,4) \rightarrow \boxed{f(\tilde{\boldsymbol{a}}^{(6)}) = \tfrac{1}{2} \cdot \tfrac{2}{3}^2 + (\tfrac{2}{3} + \tfrac{4}{3}) = 2.222}$

Zadanie 9. Regresja liniowa

Wyznaczyć aproksymację liniową funkcji, y=f(x), w oparciu o poniższy zbiór próbek (x_i, y_i) $(i=1,2,\ldots,6)$.

i	1	2	3	4	5	6
x	1	1	1	1	2	3
y	2	3	4	1	1	1

Rozwiązanie 9. Regresja liniowa

Idea rozwiązania.

Rozwiązanie problemu metodą najmniejszych kwadratów (MNK) .

Poszukiwana jest funkcja liniowa o postaci parametrycznej:

$$y = f(x \mid a, b) = a\,x + b$$

Początkowy układ 6 równań dla zadanych próbek uczących:

$$a \cdot 1 + b = 2 + \varepsilon_1 \;; \quad a \cdot 1 + b = 3 + \varepsilon_2$$

$$a \cdot 1 + b = 4 + \varepsilon_3 \;; \quad a \cdot 1 + b = 1 + \varepsilon_4$$

$$a \cdot 2 + b = 1 + \varepsilon_5 \;; \quad a \cdot 3 + b = 1 + \varepsilon_6$$

Sumaryczna funkcja błędu: $\mathrm{U}(a, b) = \varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \varepsilon_4^2 + \varepsilon_5^2 + \varepsilon_6^2$

Jest to funkcja kwadratowa względem a, b. Poszukiwane minimum znajdujemy w punkcie o zerowych pochodnych cząstkowych funkcji U względem nieznanych parametrów a, b :

$$(1/2)\;(\partial \mathrm{U} / \partial a) = 0 \;; \qquad (1/2)\;(\partial \mathrm{U} / \partial b) = 0 \,.$$

Zadanie 10. Aproksymacja wielomianem

Informacja o zależności dwóch zmiennych x, y dana jest w postaci wyników N=9 niezależnych pomiarów obarczonych błędami przypadkowymi:

x = [0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0]

y = [1.02, 0.88, 0.48, 0.44, 0.33, 0.30, 0.15, 0.14, 0.12]

Dokonać aproksymacji funkcji, $y = f(x)$, wielomianem 3-go stopnia.

Idea rozwiązania

To zadanie rozwiązujemy metodą najmniejszych kwadratów (MNK). Wielomian stopnia trzeciego jest postaci:

$$y = f(x \mid a, b, c, d) = ax^3+bx^2+cx+d\ .$$

Początkowy układ 9 równań (liniowe względem a, b, c, d):

$$a \cdot 0^3 + b \cdot 0^2 + c \cdot 0 + d = 1.02 + \varepsilon_1$$

$$a \cdot 0.5^3 + b \cdot 0.5^3 + c \cdot 0.5 + d = 0.88 + \varepsilon_2$$

…

Zadanie 11. Model kNN w aproksymacji funkcji

Zastosować model pamięciowy uczenia się aproksymacji funkcji nad przestrzenią cech $\Re^2$. Zadanych jest 11 próbek uczących:

c= (x, y)	(1;2,5)	(1; 3)	(1;3,5)	(2; 3)	(4; 3)	(2; 2)	(1; 1)	(2; 0)	(3; 1)	(3,5;1,5)	(4;1,5)
f(**c**)	1	1	1	2	2	3	3	3	3	4	4

Podać wartości aproksymowanej funkcji w metodzie 4 najbliższych sąsiadów dla cech (2; 2,5), (3, 3) i (1, 2).

Idea rozwiązania dla (2; 2.5).

Znajdujemy 4 najbliższych sąsiadów tego punktu -

c= (x, y)	**(1;2,5)**	**(1; 3)**	(1;3,5)	**(2; 3)**	(4; 3)	**(2; 2)**	(1; 1)	(2; 0)	(3; 1)	(3,5;1,5)	(4;1,5)
$\lVert c - (2; 2.5) \rVert$	1	$\sqrt{1.25}$	$\sqrt{2}$	0.5	$\sqrt{4.25}$	0.5	$\sqrt{3.25}$	2.5	$\sqrt{3.25}$	$\sqrt{3.25}$	$\sqrt{5}$

Rozwiązanie 11. Model kNN w aproksymacji

Idea rozwiązania dla (2; 2.5) (c.d.).

Najbliżsi 4 sąsiedzi to: (1, 2.5), (1, 3), (2, 3) i (2, 2):

c= (x, y)	**(1;2,5)**	**(1; 3)**	(1;3,5)	**(2; 3)**	(4; 3)	**(2; 2)**	(1; 1)	(2; 0)	(3; 1)	(3,5;1,5)	(4;1,5)
$\Vert \boldsymbol{c} -$ (2; 2.5) $\Vert$	1	$\sqrt{1.25}$	$\sqrt{2}$	0.5	$\sqrt{4.25}$	0.5	$\sqrt{3.25}$	2.5	$\sqrt{3.25}$	$\sqrt{3.25}$	$\sqrt{5}$

Stosujemy wzór na aproksymację funkcji w modelu kNN (średnia wartość dla k sąsiadów) otrzymując:

$$z = f(2{,}2.5) \approx \frac{1}{4}\left(1+1+2+3\right) = \frac{7}{4}$$

Zadanie 12. Model kNN w aproksymacji funkcji

Zilustrować zasadę modelu pamięciowego w uczeniu się aproksymacji funkcji, z zastosowaniem algorytmu k sąsiadów (kNN), dla *k* = 3. Niech pamięć zawiera przykłady uczące $U = \{x_1, x_2, \ldots, x_{25}\}$, reprezentowane pojedynczą cechą *c* o postaci:

$$c_i = \varphi(x_i) = 0.006\pi \cdot i^{3/2}$$

Wartości przyjmowane dla tych przykładów przez funkcję docelową wynoszą: $f(c) = \frac{\sin c}{c}$

Obliczyć sumaryczny błąd i błąd średniokwadratowy dla aproksymacji funkcji reprezentowanej przez zapamiętane przykłady na zbiorze przykładów $T = (x_1', x_2', \ldots, x_8')$, zadanych jako:

$$\varphi(x_i') = 0.05\pi \cdot i^{4/3}$$

Rozwiązanie 12 (1)

Idea rozwiązania

Należy napisać nieduży program, który posiada następujące kroki obliczeniowe:

1. Wyznacza 25 przykładów (próbek) uczących *U:*

i	1	2	3	4	…	…	24	25
c	0.006π	0.017π	0.031π	0.048π	…	…	0.705π	0.75π
$f(c)$	0.9999	0.9995	0.9984	0.9962	…	…	0.3610	0.3001

2. Próbki te są pamiętane i reprezentują komórki Voronoia w dziedzinie funkcji (w tym przypadku jest to 1-wymiarowa dziedzina i komórki mają postać przedziałów na osi c).
3. Wyznacza 8 przykładów w zbiorze testowym *T*:

i	1	2	3	4	5	6	7	8
c'	0.05π	0.126π	0.216π	0.317π	0.427π	0.545π	0.670π	0.800π
$f(c')$		…	…	…	…	…	…	…

Rozwiązanie 12 (2)

Idea rozwiązania (c.d.)

4. Dla przykładów ze zbioru T ($c'_i \in T$) wyznacza aproksymację wartości funkcji $f'(c'_i)$ w oparciu o 3 najbliższych sąsiadów $\{c_k\}$:

$$f'(c'_i) = \frac{1}{3}\sum_{k=1}^{3} f(c_k)$$

5. Każda z 8 aproksymowanych wartości jest porównywana z rzeczywistymi wartościami funkcji podanymi w zbiorze testowym:

$$\varepsilon_i = f(c'_i) - f'(c'_i)\ , dla\ i = 1,2,\ldots,8$$

6. Sumaryczny błąd dla tych 8 aproksymowanych wartości: $\mathrm{E} = \sum_{i=1}^{8} \varepsilon_i$

7. Błąd średniokwadratowy: $\mathrm{E}_2 = \frac{1}{8}\sum_{i=1}^{8} \varepsilon_i^2$