Budowa modeli klasyfikacyjnych o skośnych warunkach

Marcin Michalak
(Marcin.Michalak@polsl.pl)

III spotkanie
Polskiej Grupy Badawczej Systemów Uczących Się
Wrocław, 17–18.03.2014

Outline

1. Wstęp
   - Dwa podejścia
   - Ilustracja graficzna
2. „Inne rozwiązania"
3. „Nasze rozwiązania"
   - ORG (*ORG*1.0)
   - PCA ORG (*ORG*2.0)
   - PCA ORG (*ORG*2.1)
   - CHIRA
4. ADAA

Dwa podejścia

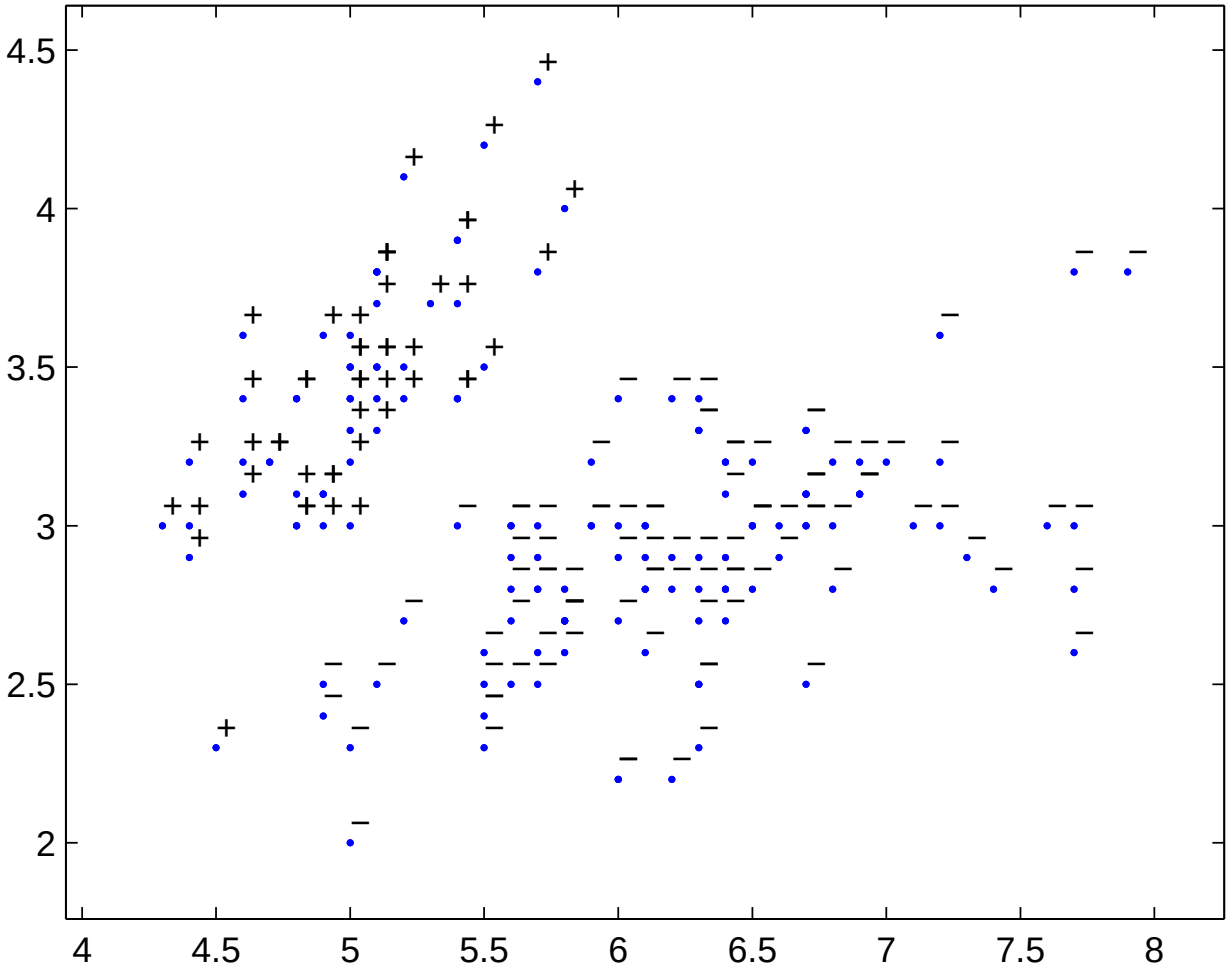
Podejście jawne

- drzewa decyzyjne
- systemy regułowe
- liniowe SVM

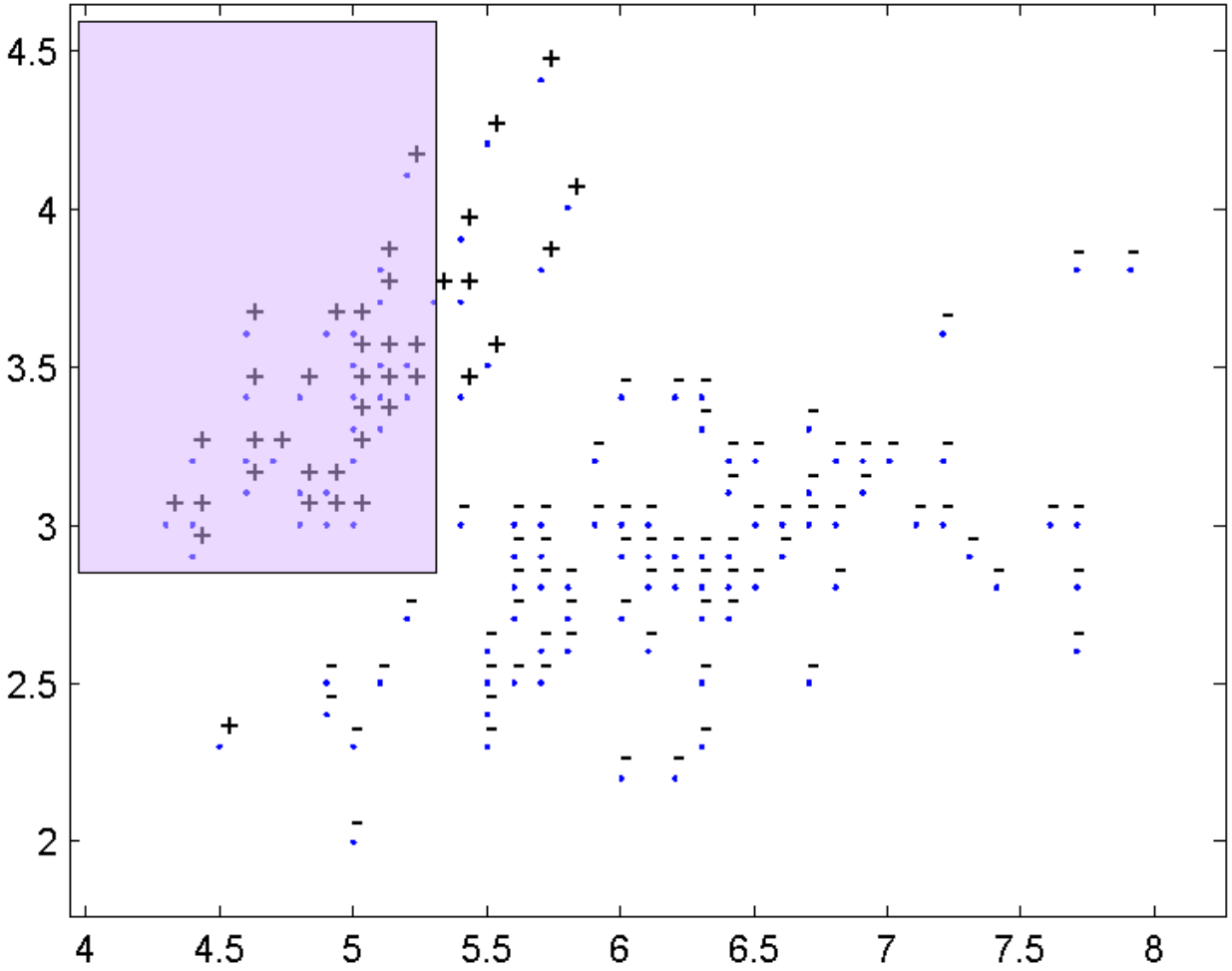
Podejście ukryte

- sztuczne sieci neuronowe
- nieliniowe SVM
- GLM

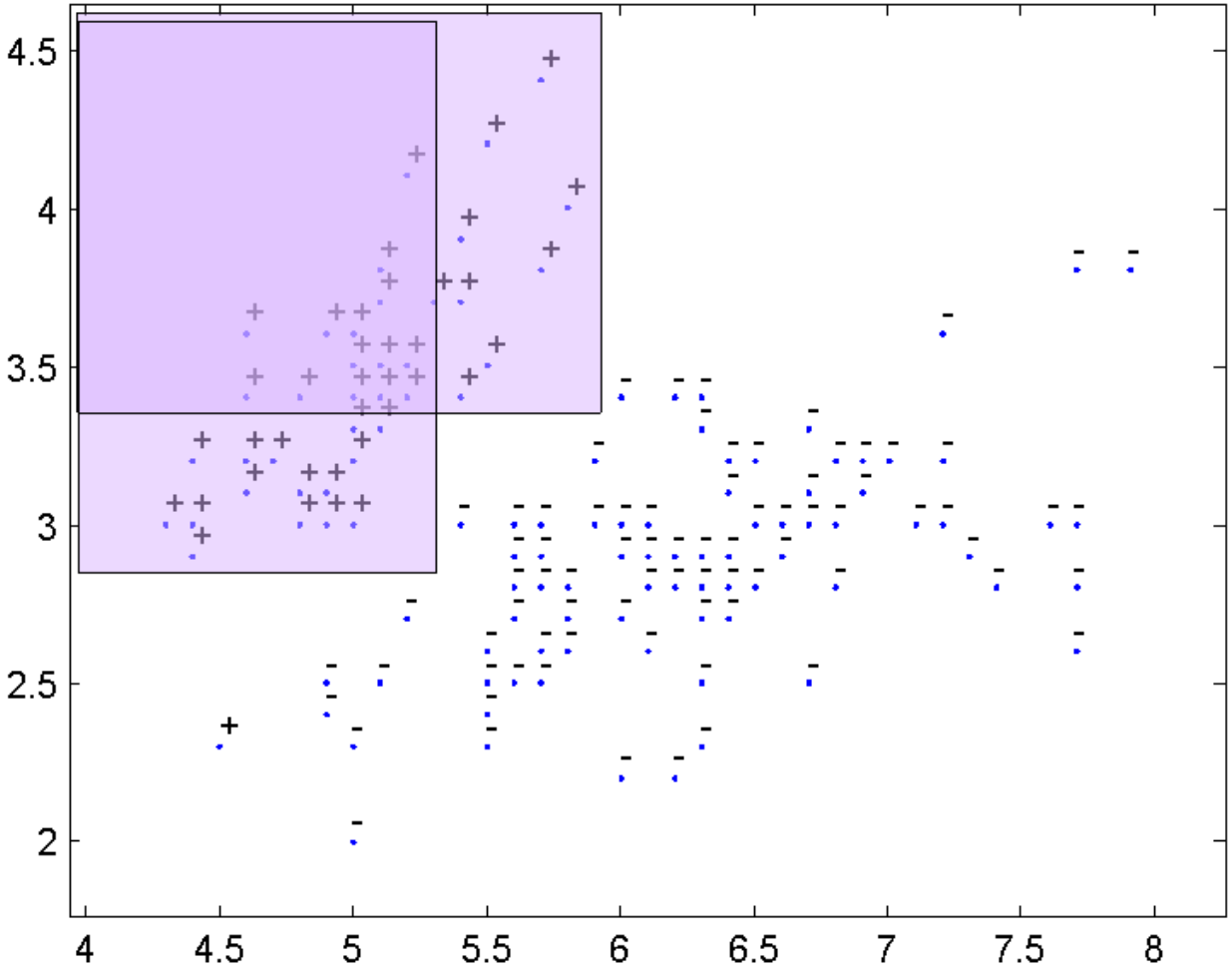
Rysunek: Dwuwymiarowy, dwuklasowy zbiór danych.



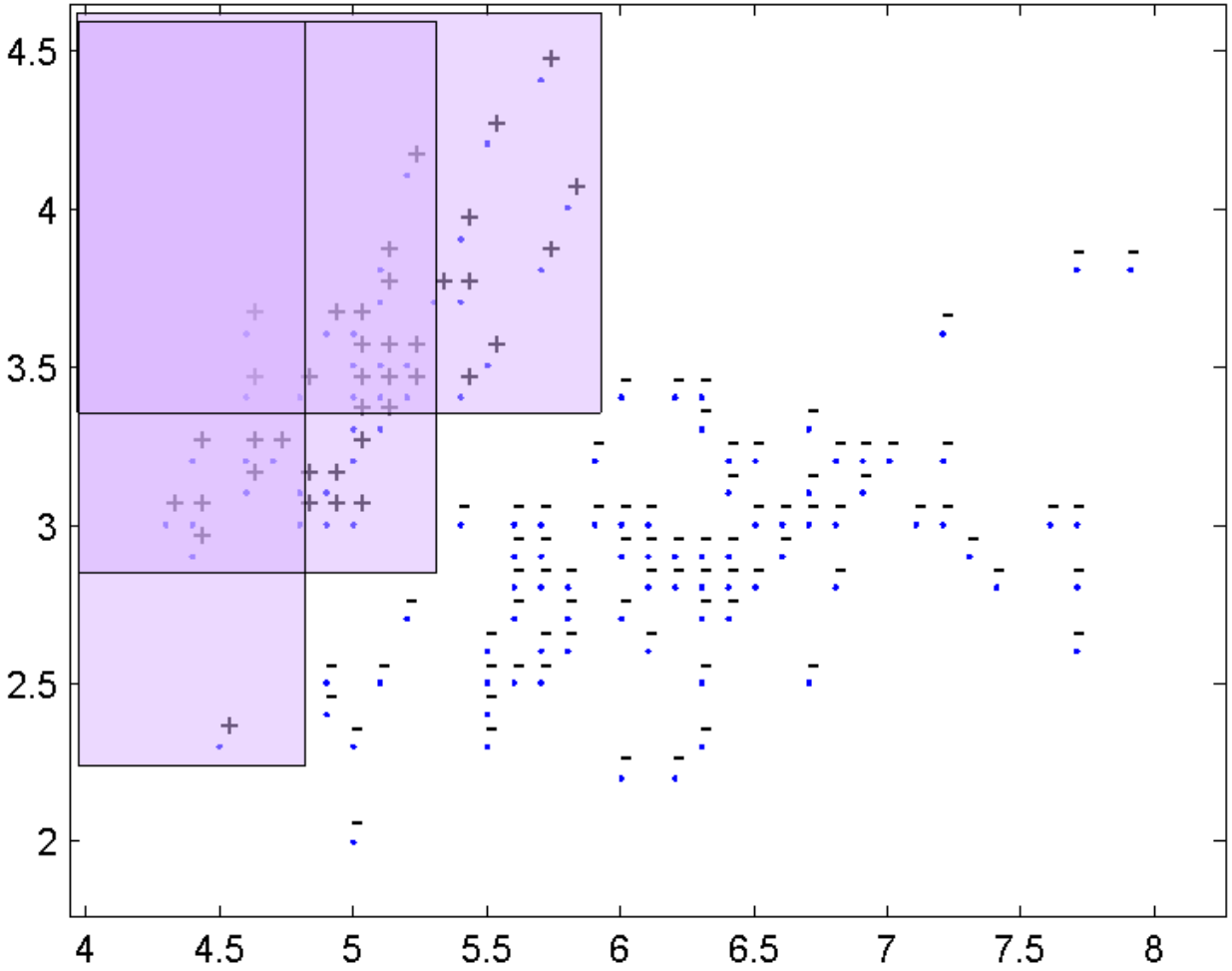
Rysunek: Dwuwymiarowy, dwuklasowy zbiór danych.



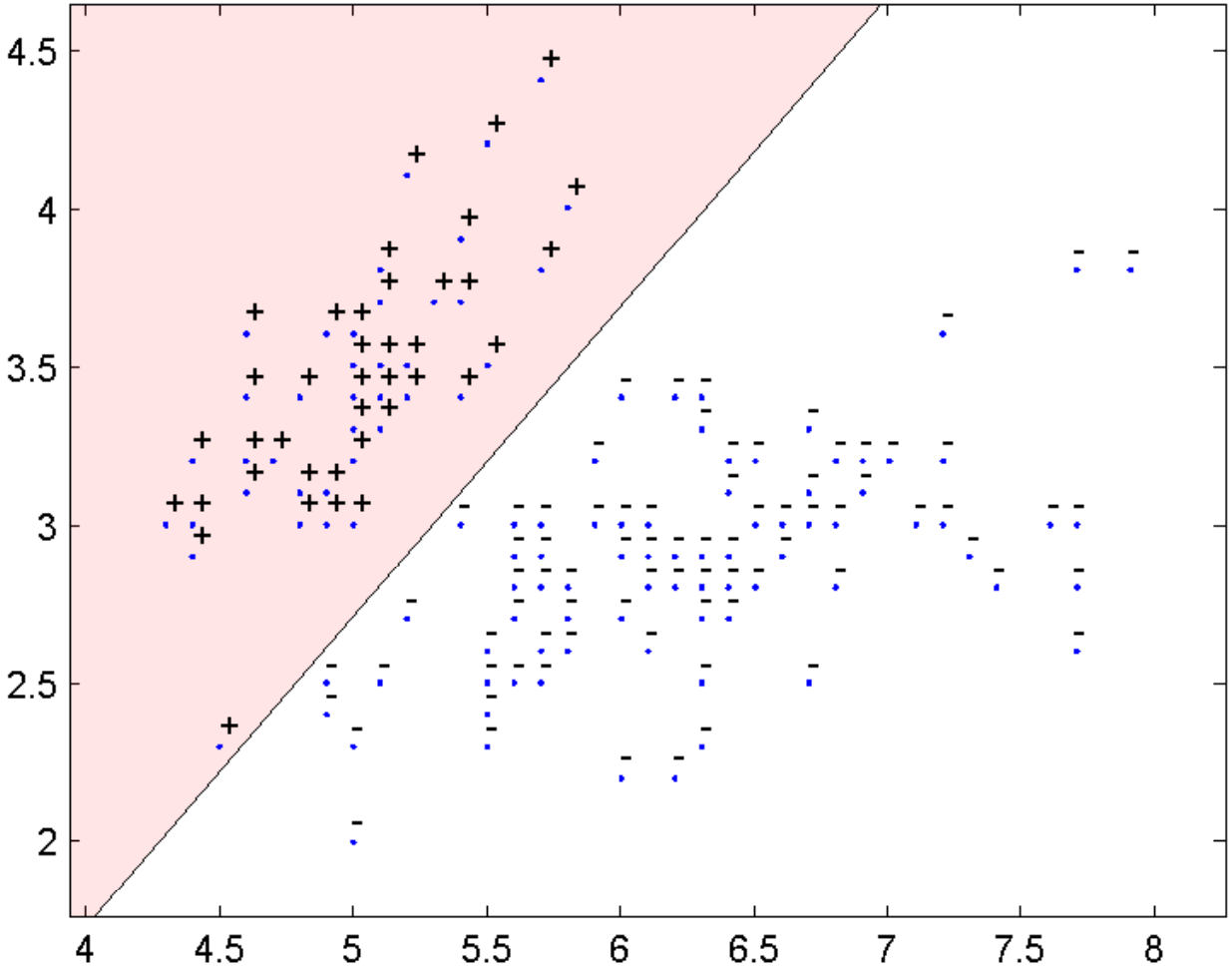
Rysunek: Dwuwymiarowy, dwuklasowy zbiór danych.



Rysunek: Dwuwymiarowy, dwuklasowy zbiór danych.



Rysunek: Dwuwymiarowy, dwuklasowy zbiór danych.



Ekstrapolacja SVM

Barakat N., Diederych J.

Learning-based Rule-extraction from Support Vector Machines. In 14th International Conference on Computer Theory and Applications ICCTA 2004 Proceedings, Alexandria, Egypt, 107-112, 2004

- Zbuduj SVM.
- Wygeneruj nowe przypadki losowo i zaklasyfikuj je zgodnie z SVM.
- Z „gęstszego” zbioru wygeneruj reguły decyzyjne.

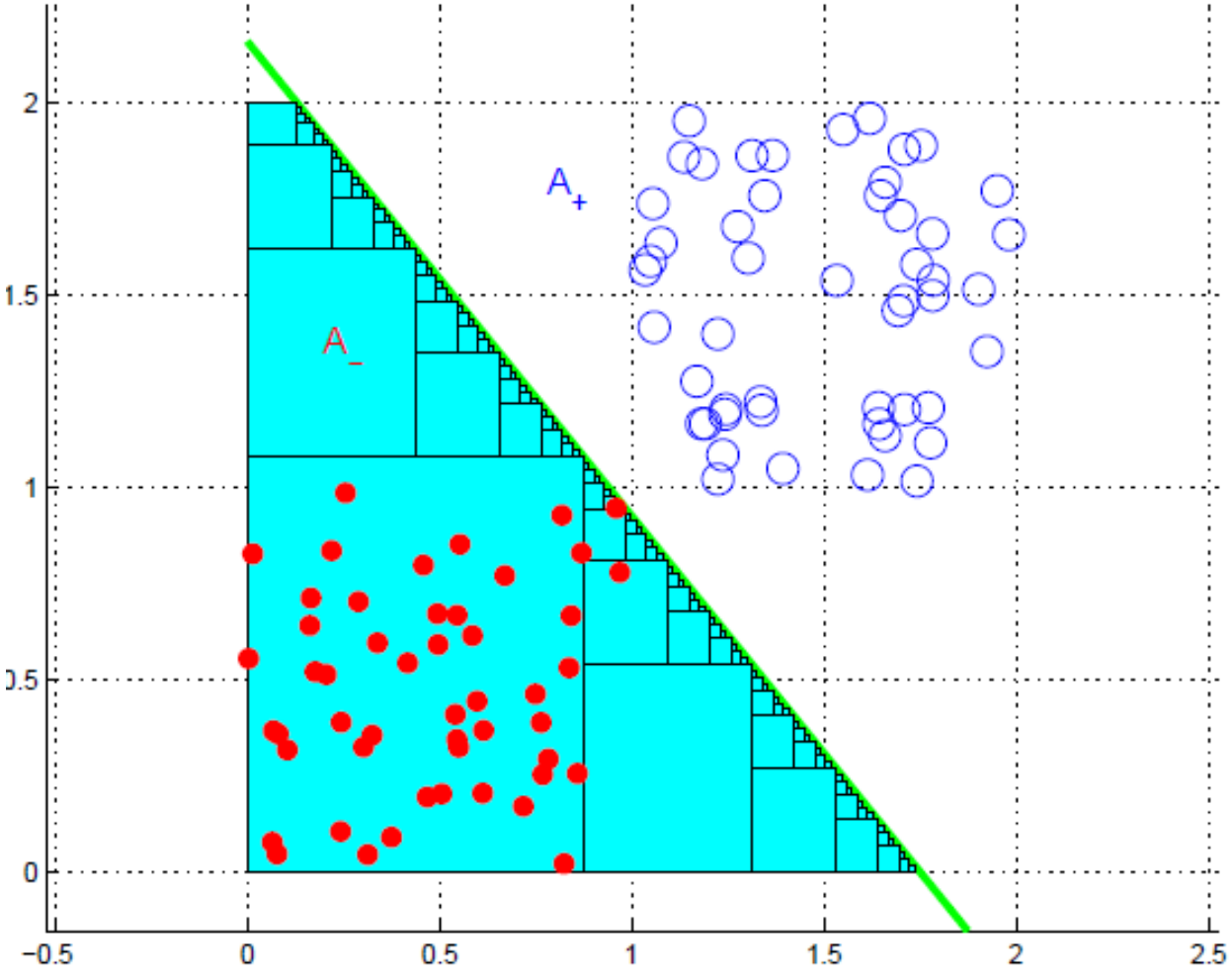
Pokywanie obszaru pod liniowym SVM

Fung G., Sandilya S., Rao R.

Rule Extraction from Linear Support Vector Machines. In Proceedings of the 11th ACM SIGKDD international Conference on Knowledge Discovery in Data Mining, 32-40, 2005

- Zbuduj liniowy SMV
- Pokrywaj obszar „pod" („nad") rozłącznymi hiperprostopadłościanami o jednym wierzchołku na OSH

Pokywanie obszaru pod liniowym SVM



ITER + MINERVA

Huysmans J., Baesens B., Vanthienen J.

ITER: an Algorithm for Predictive Regression Rule Extraction, LNCS 4081:270-279, 2006

Huysmans J., Setiono R., Baesens B., Vanthienen J.

Minerva: Sequential Covering for Rule Extraction, IEEE Transactions On Systems, Man, and Cybernetics-Part B: Cybernetics, 38(2):299-309, 2008

ITER + MINERVA

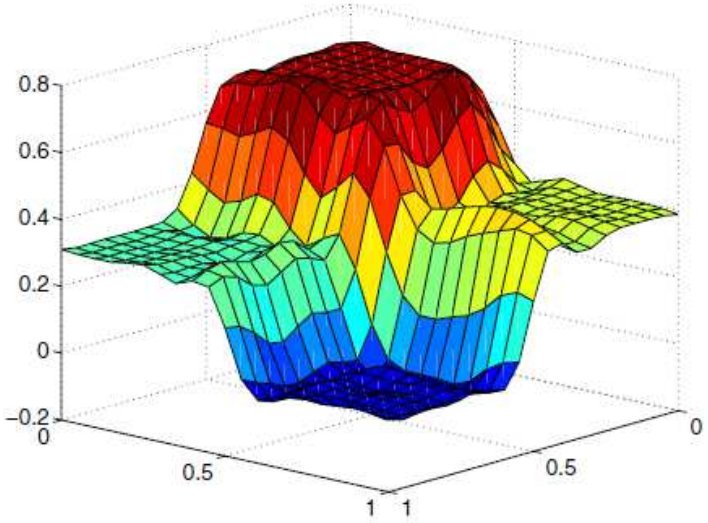
ITER

- „pedagocical regression rule extractor”,
- bazuje na dowolnym regresorze
- iteracyjnie zwiększa pokrycie hiperprostopadłościanami

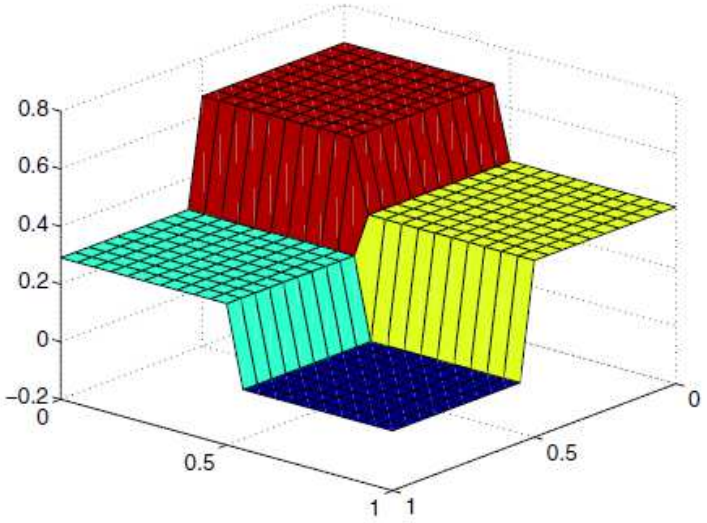
MINERVA

- zapobiega pokrywaniu się reguł, ale nadal nie wykrywa „skośności”

ITER + MINERVA

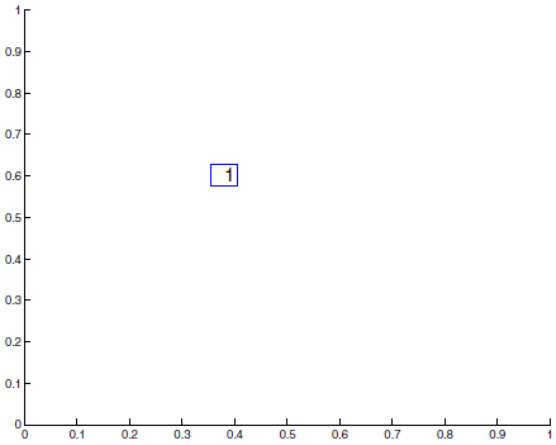


(a) LS-SVM Model

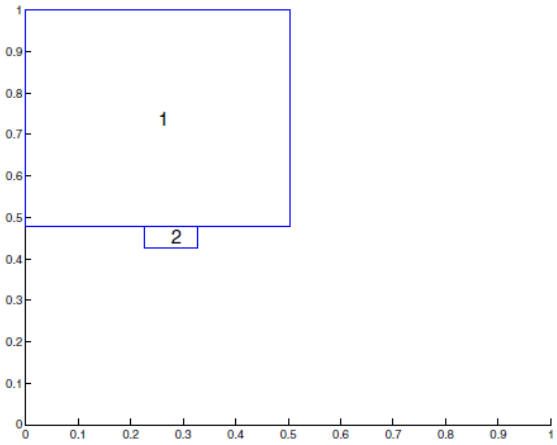


(b) Extracted Rules (threshold=0.2)

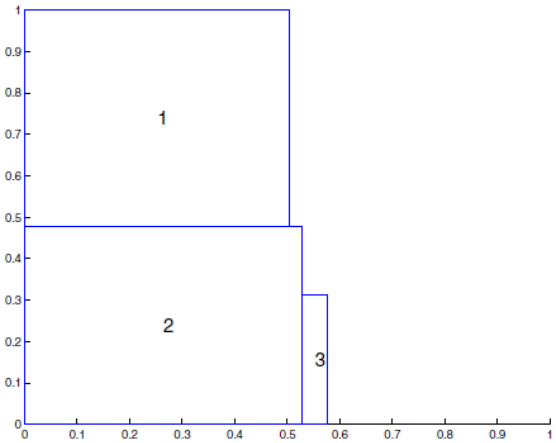
ITER + MINERVA



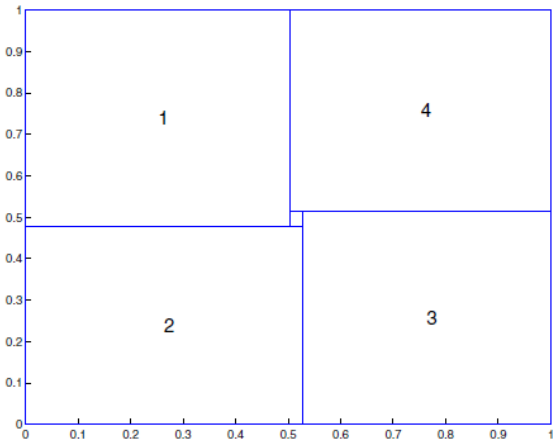
(c) Initial Cube (After 2 Updates)



(d) After 25 Updates



(e) After 50 Updates



(f) After 84 Updates

SVM+Prototypes

Nunez H., Angulo C., Catala A.

Rule Extraction from Support Vector Machines, proc. of European Symposium on Artificial Networks, 107-112, 2002

- Zbuduj SVM (znajdź wektory podpierające)
- Podziel klasy na podklasy (znajdź prototypy podklas)
- para SV + prototyp definiuje regułę (elipsoida albo hiperprostopadłościan)

SVM+Prototype

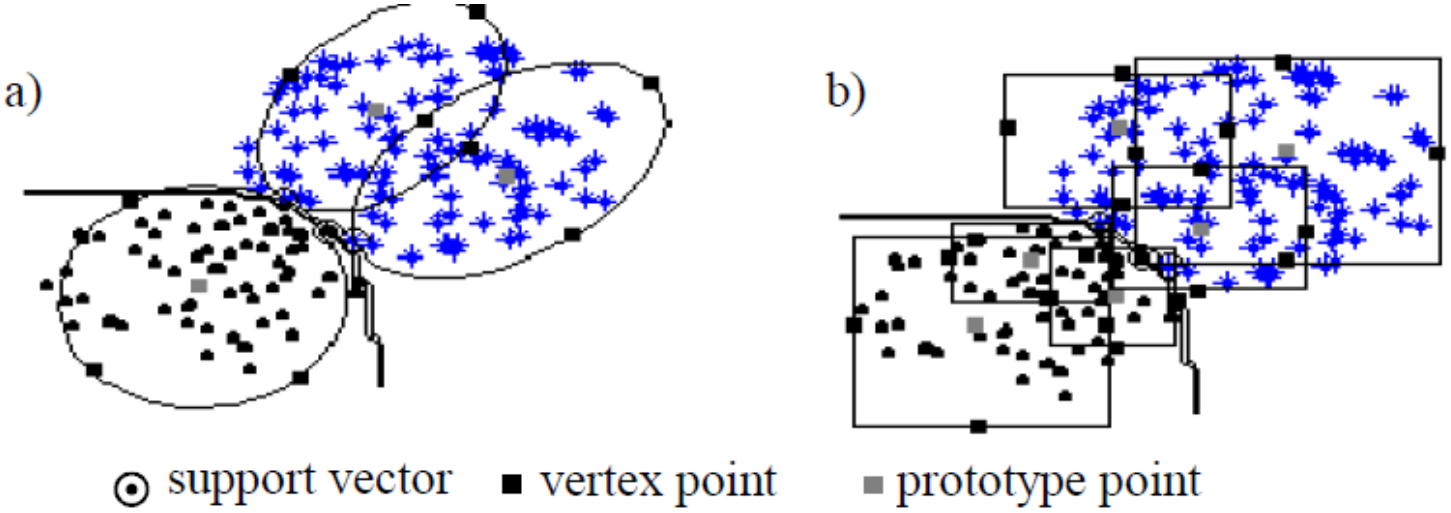
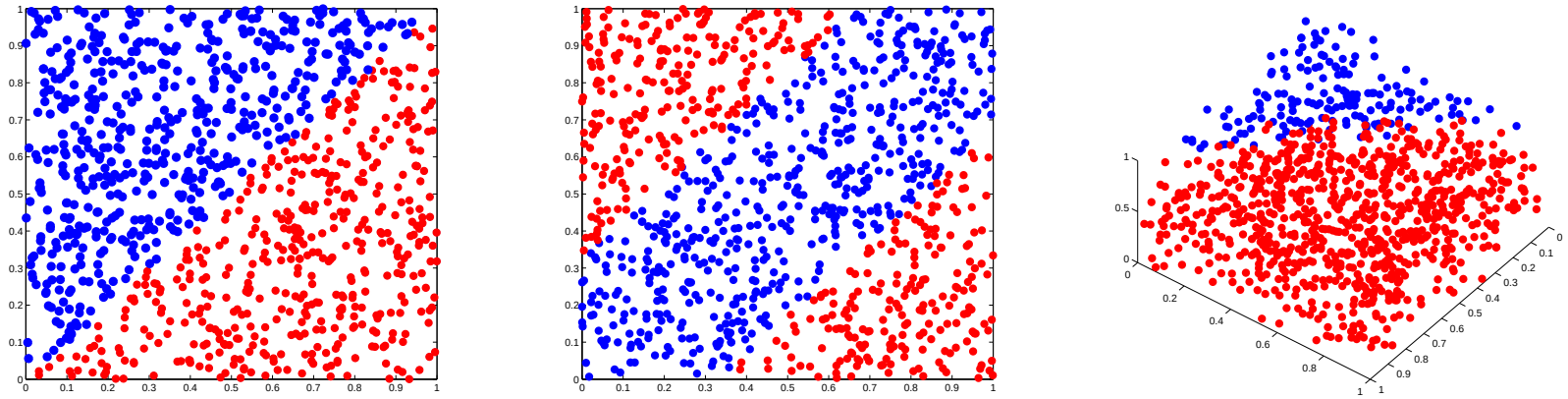


Figure 2. a) Ellipsoids for equation rules. b) Hyper-rectangles for interval rules.

„Nasze rozwiązania”

- CHIRA
- ORG (ORG1.0)
- PCA ORG (ORG2.0)
- Advanced PCA ORG (ORG2.1)

Skośne zbiory danych



Rysunek: Trzy zbiory skośnych danych: 2d (z lewej); d2d (pośrodku); 3d (z prawej).

Oblique Rules Generator

Publikacja

Michalak M., Sikora M., Ziarnik P.: ORG - Oblique Rules Generator, In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2012, Part II. LNCS, vol. 7268, pp. 152–159, (2012)

Główna idea

Kolejne skośne warunki (hiperpłaszczyzny) reguły poszukiwane metodą siatkowego przeszukiwania przestrzeni parametrów warunku.

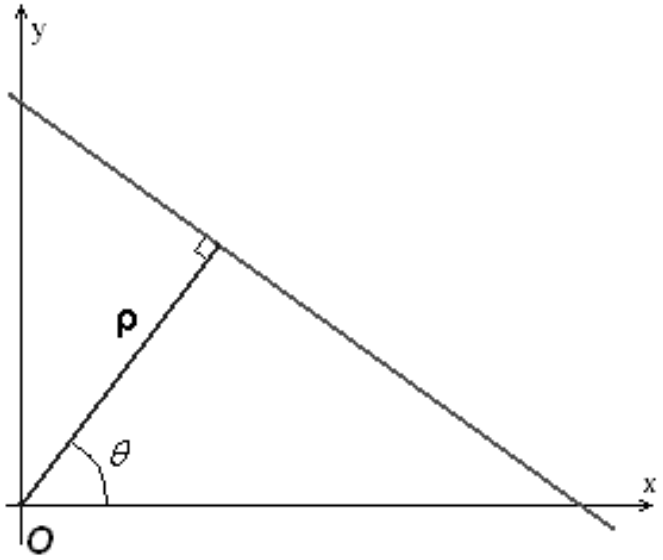
Założenia

Równanie normalne hiperpłaszczyzny:

$$\alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_n x_n - \rho = 0$$

Warunek kosinusów kierunkowych:

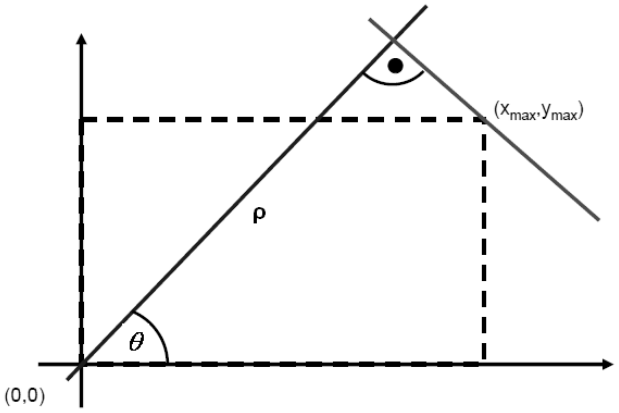
$$\alpha_1{}^2 + \alpha_2{}^2 + \ldots + \alpha_n{}^2 = 1$$



Założenia (c.d.)

Ograniczenia wartości kosinusów – naturalne; Ograniczenie maksymalnej odległości, przy założeniu, że wszystkie punkty mają dodatnie współrzędne:

$$\rho_{\max}(\theta_{opt}) = x_{\max}\cos\left(\operatorname{arc\,tg}\frac{y_{\max}}{x_{\max}}\right) + y_{\max}\sin\left(\operatorname{arc\,tg}\frac{y_{\max}}{x_{\max}}\right)$$



Algorytm

```
Train ← All, RCnt ← 0, Rules ← ∅
while card(Rules) < R do
    Rule ← ∅, Train_Rule ← Train
    while card(Rule) < C do
        bestCond ← ∅
        bestCondQ ← 0
        for each cell ∈ Grid do
            cond ← cell
            condQ ← Q(cond)
            if condQ > bestCondQ then
                bestCondQ ← condQ
                bestCond ← cond
            end if
        end for each
        Rule_temp ← Rule ∪ bestCond
        if Q(Rule_temp) > Q(Rule) then
            Rule ← Rule_temp
        end if
        Train_Rule ← Train_Rule \ under(bestCond)
        if Q(Rule) = 1 then
            break
        end if
        if card(Train_Rule) = 0 then
            break
        end if
    end while
    Train_Rule ← Train_Rule \ covered(Rule)
    Rules ← Rules ∪ Rule
end while
```

Podsumowanie

Cechy

- Poszukiwanie skośnych warunków odbywa się metodą siatkową.
- Budowa reguły odbywa się metodą wspinaczki.

Wady

- Bardzo duża złożoność obliczeniowa.
- Ocena jakości skośnych warunków jest globalna.
- Tylko niektore skośne warunki brane są pod uwagę (siatka).

PCA ORG (*ORG*2.0)

Publikacja

Michalak M., Nurzyńska K.: PCA Based Oblique Decision Rules Generating, In: Tomassini, M., Antonioni, A., Daolio, F., Buesser, B. (Eds.): ICANNGA 2013, LNCS, vol. 7824, pp. 198–207, (2013)

Główna idea

- opis klasy to opis jej podklas, wyznaczonych metodą grupowania,
- opis podklasy to hiperprostopadłościan, o krawędziach wyznaczonych przez PCA dla podklasy.

Założenia i ich spełnianie

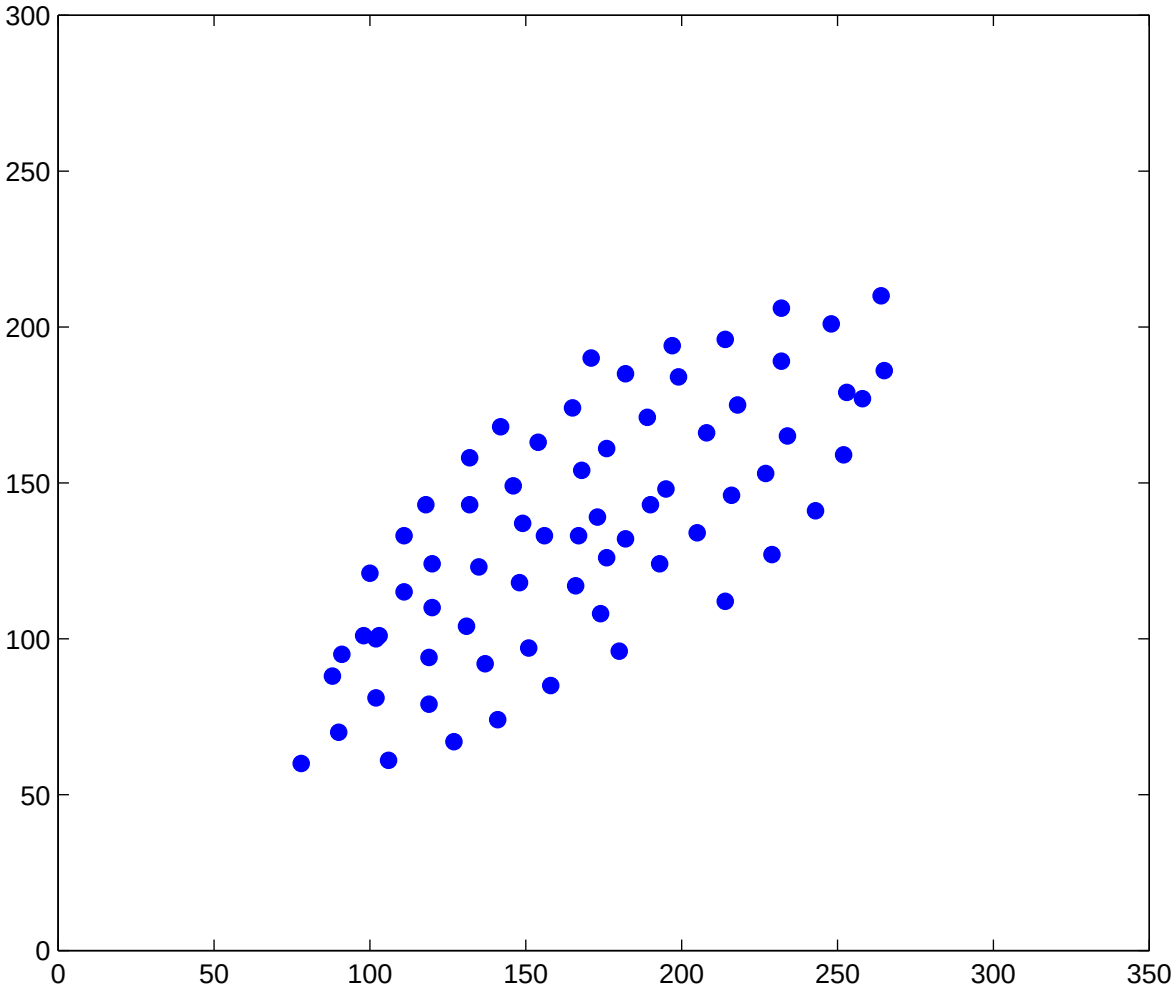
Założenia wstępne:

- Reguły opisują lokalne zależności w danych.
- Poszukiwanie skośnych warunków odbywa się lokalnie.
- Mniejsza złożoność obliczeniowa niż *ORG*1.0.

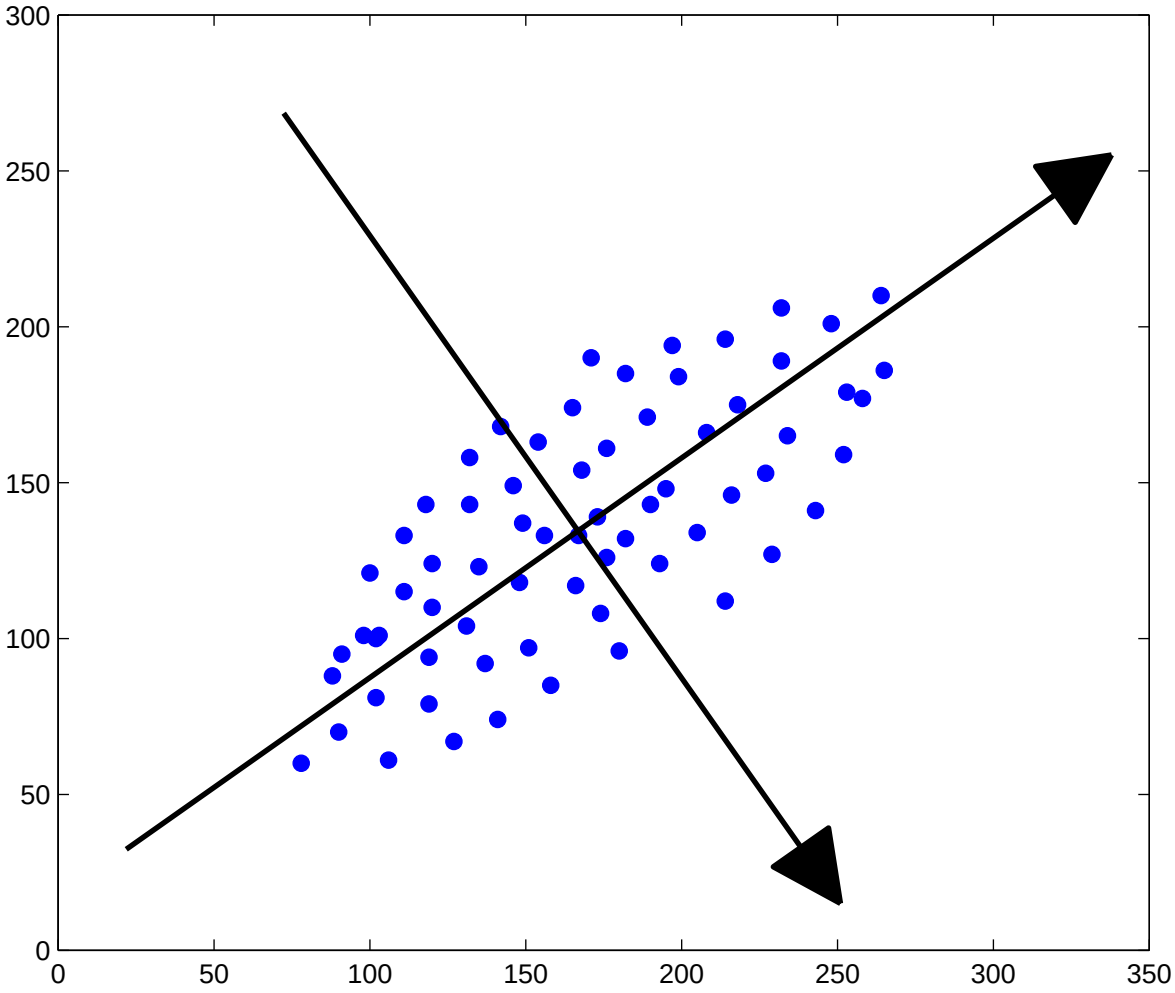
Spełnienie założeń

- Ad. 1: Podklasy wygenerowane algorytmem k-means.
- Ad. 2 i 3: ściany hiperprostopadłościanu są równoległe do kierunków wyznaczonych przez *PCA* dla podklasy.

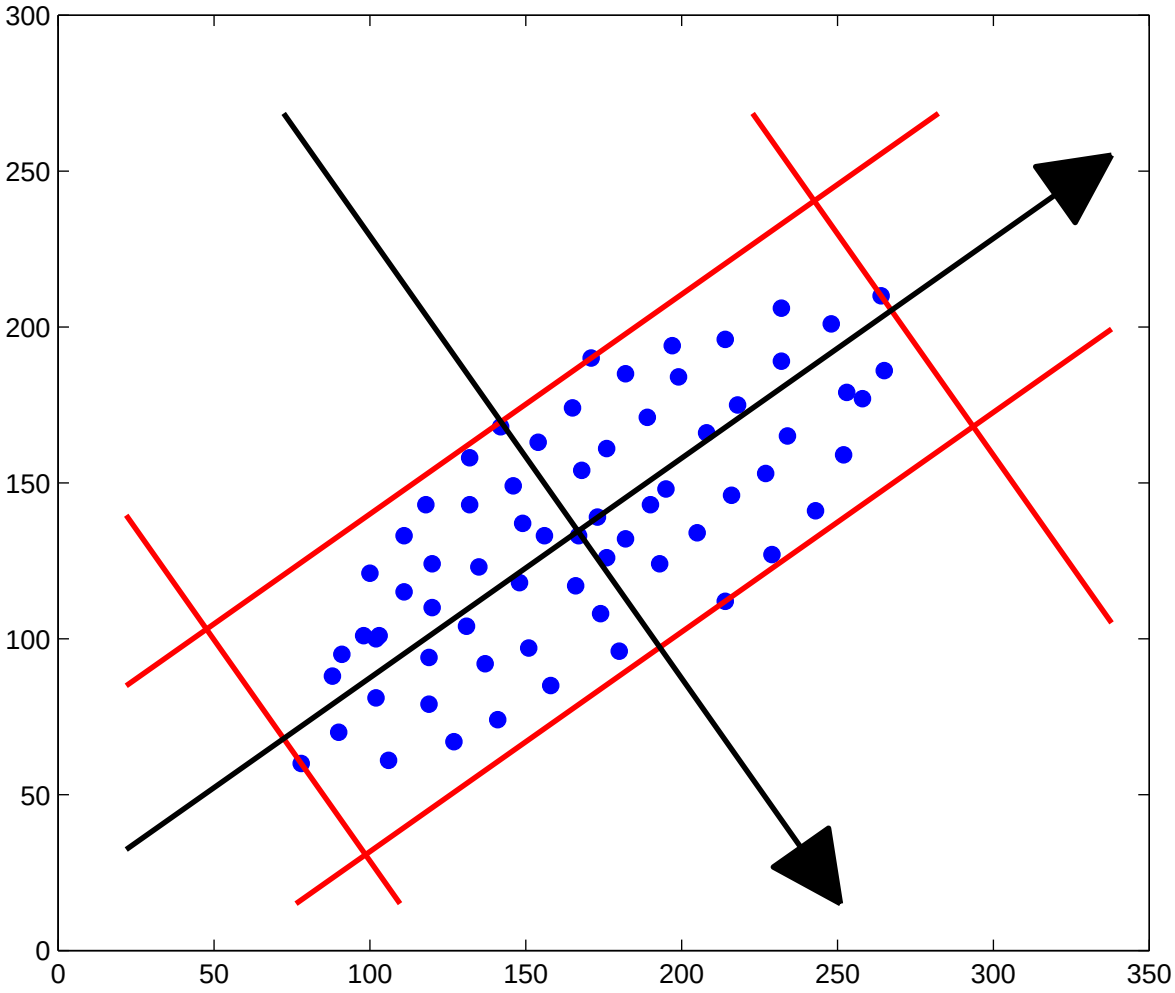
Pojedyńcza podklasa



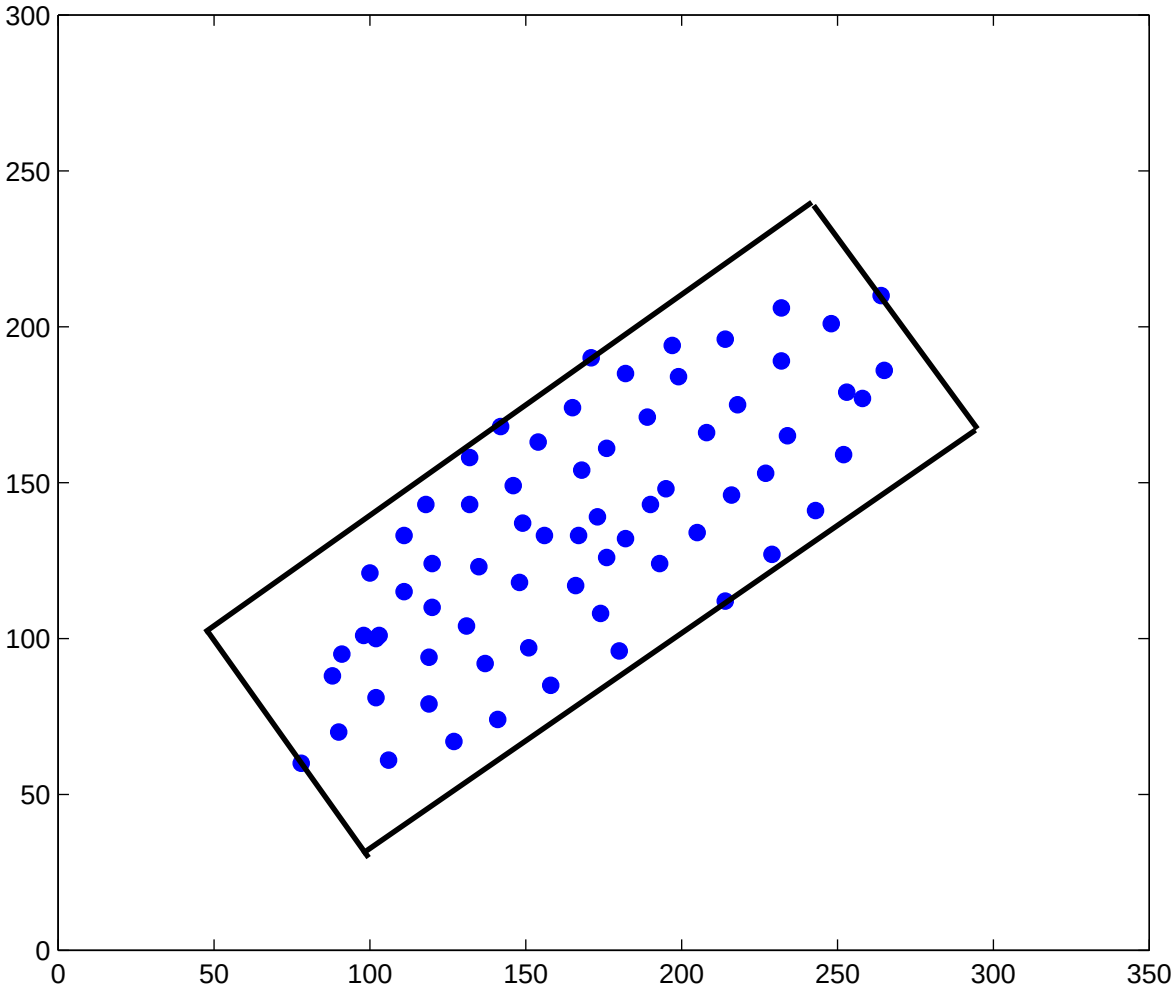
Nowe kierunki znalezione przez *PCA*



Pary skośnych warunków, równoległe do kierunków z *PCA*



Nowa, wypukła reguła decyzyjna



Silne strony

- Mała liczba reguł przypadająca na każdą klasę.
- Bardzo wysoka jakość klasyfikacji typowo skośnych zależności.

Słabe strony

- Słabe wyniki dla klas pokrywających się i wklęsłych.
- Duża liczba parametrów uzyskanych modeli.
- Brak postprocessingu uzyskanych reguł.

Advanced PCA ORG (*ORG*2.1)

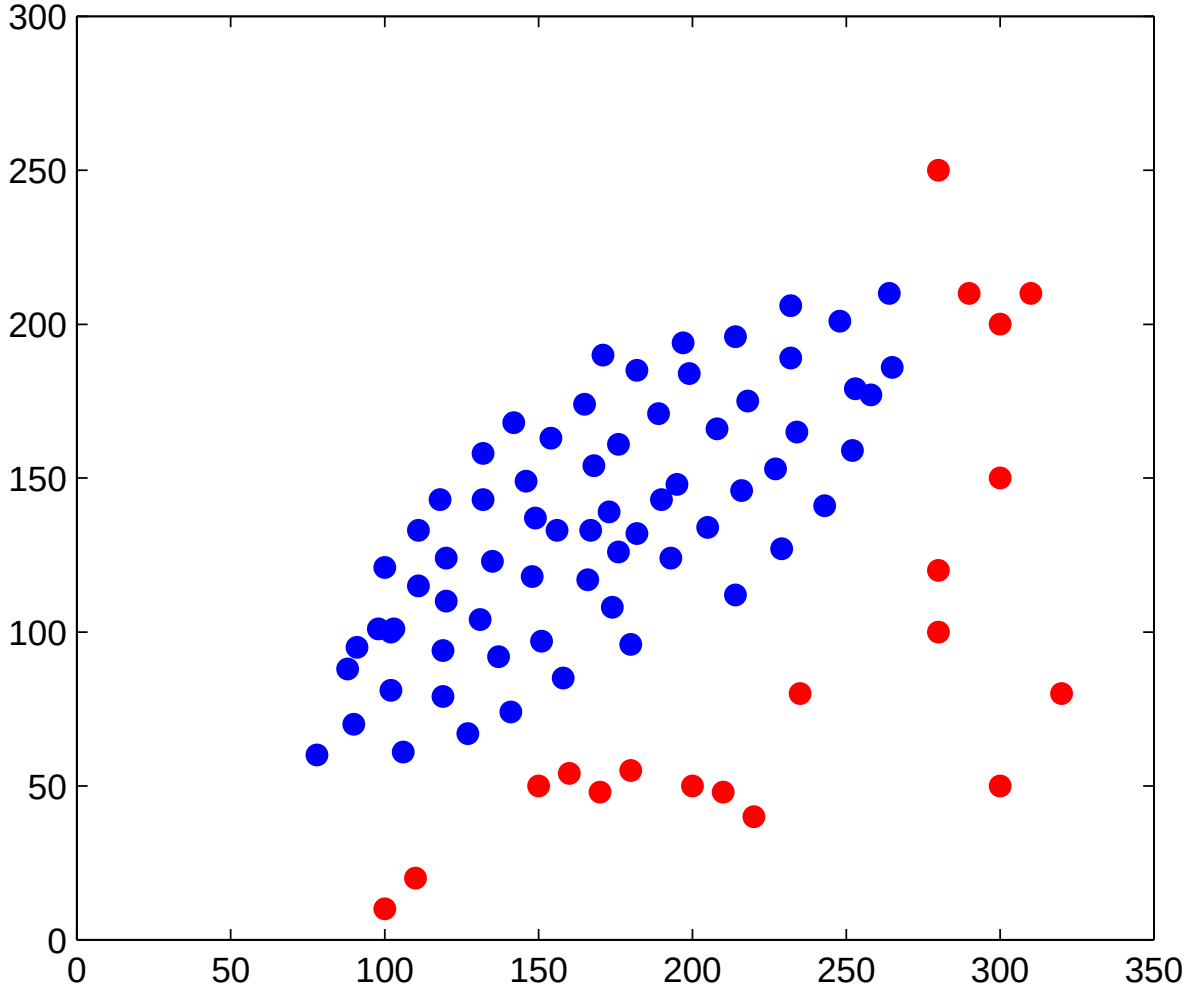
Publikacja (w druku)

Michalak M., Nurzyńska K.: Advanced Oblique Rule Generating Based on PCA, ICAISC 2014, LNCS ?? (2014)

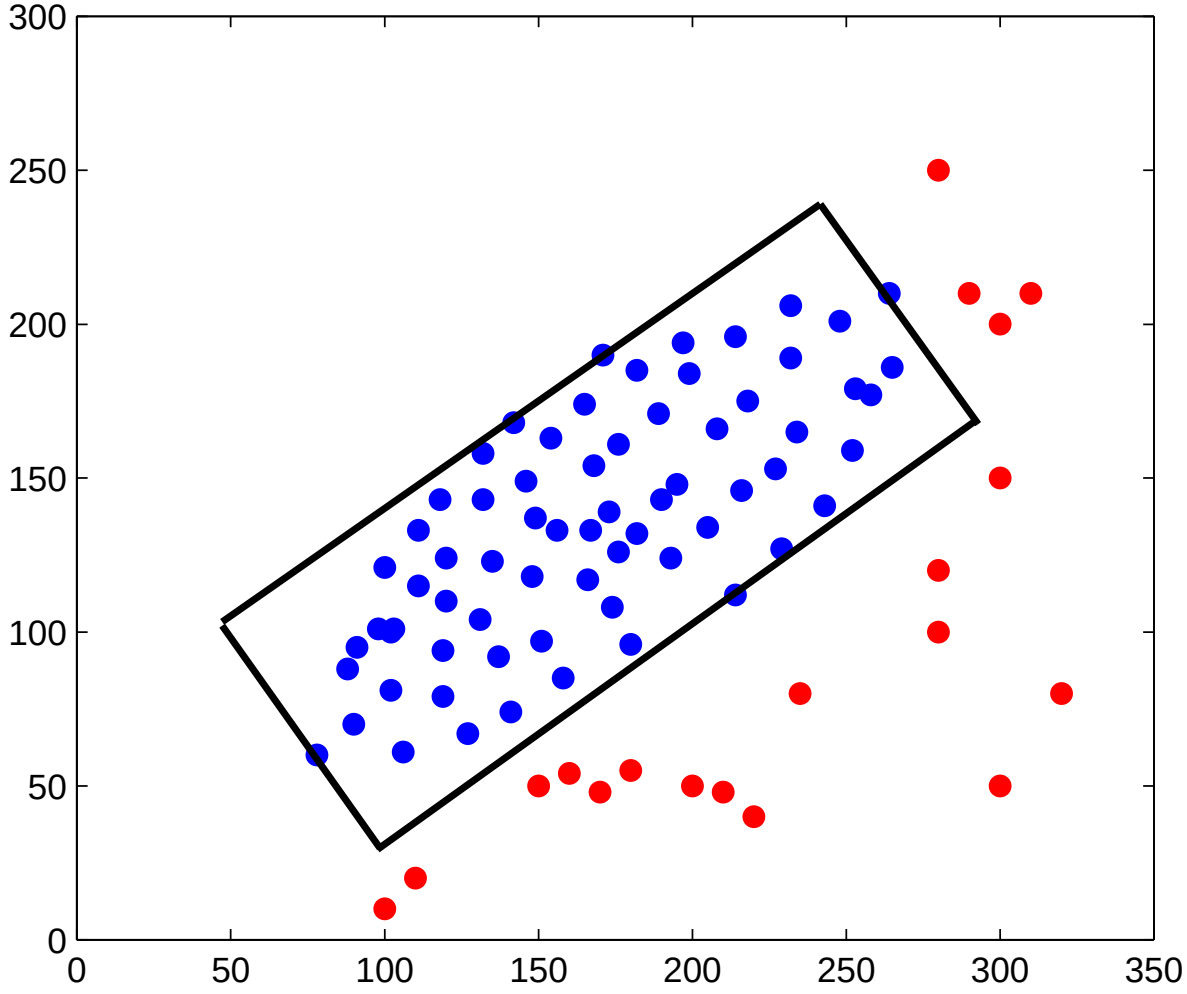
Główne idee:

- przycinanie reguł: eliminacja zbędnych warunków metodą wspinaczki;
- uogólnianie skośnych warunków: eliminacja zmiennych z warunku, jeśli jej współczynnik jest niski.

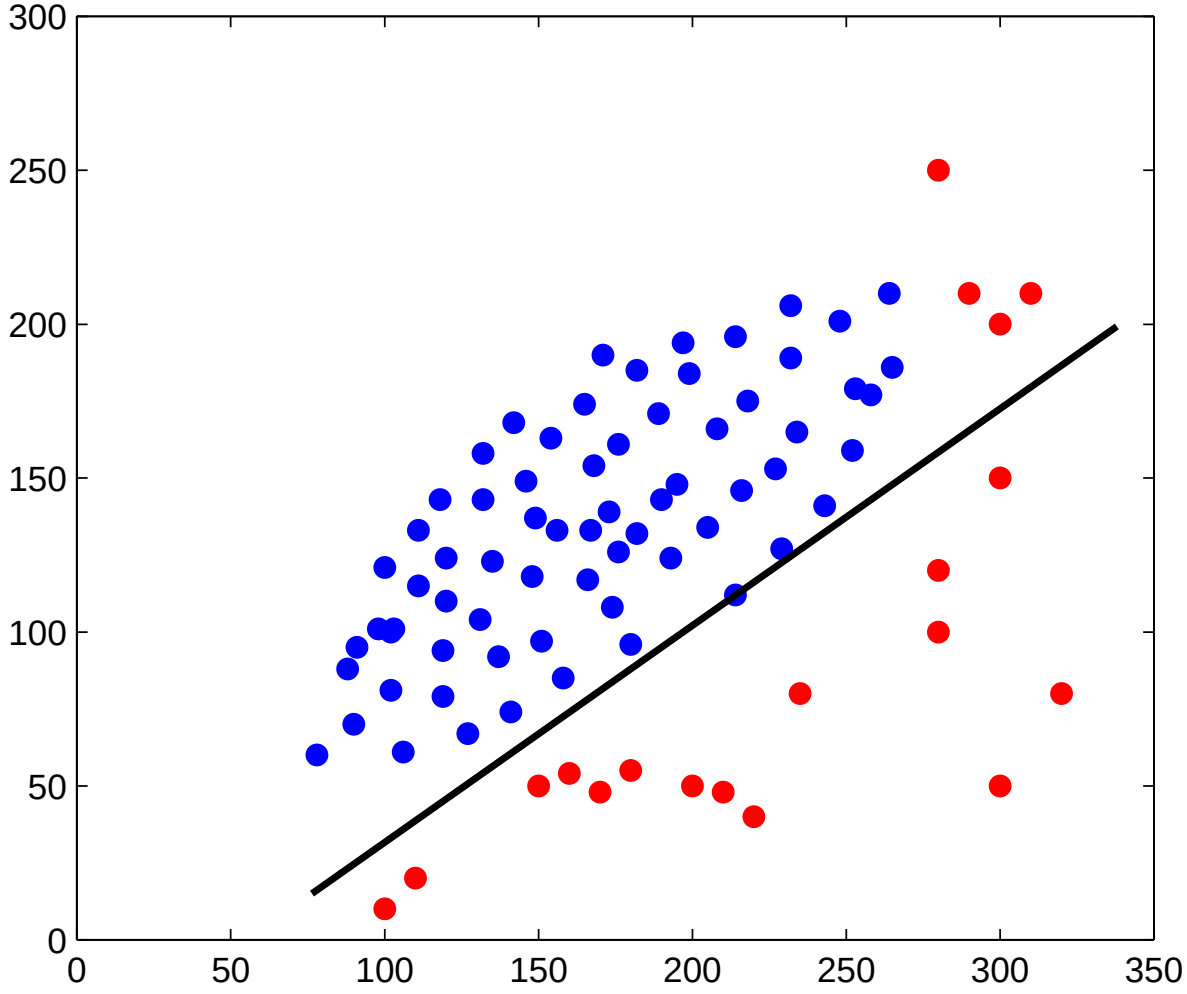
Przycinanie reguł – idea



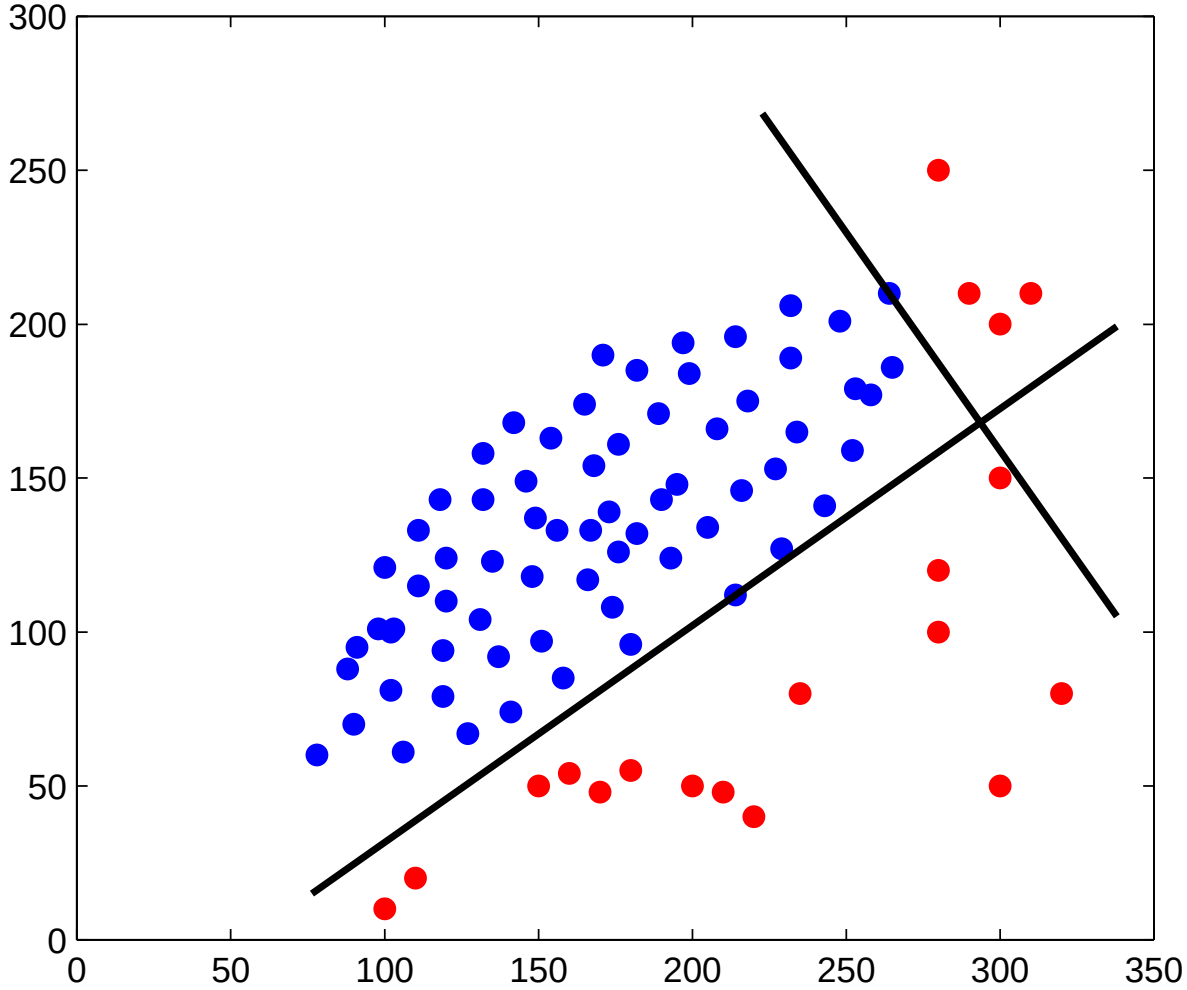
Przycinanie reguł – idea



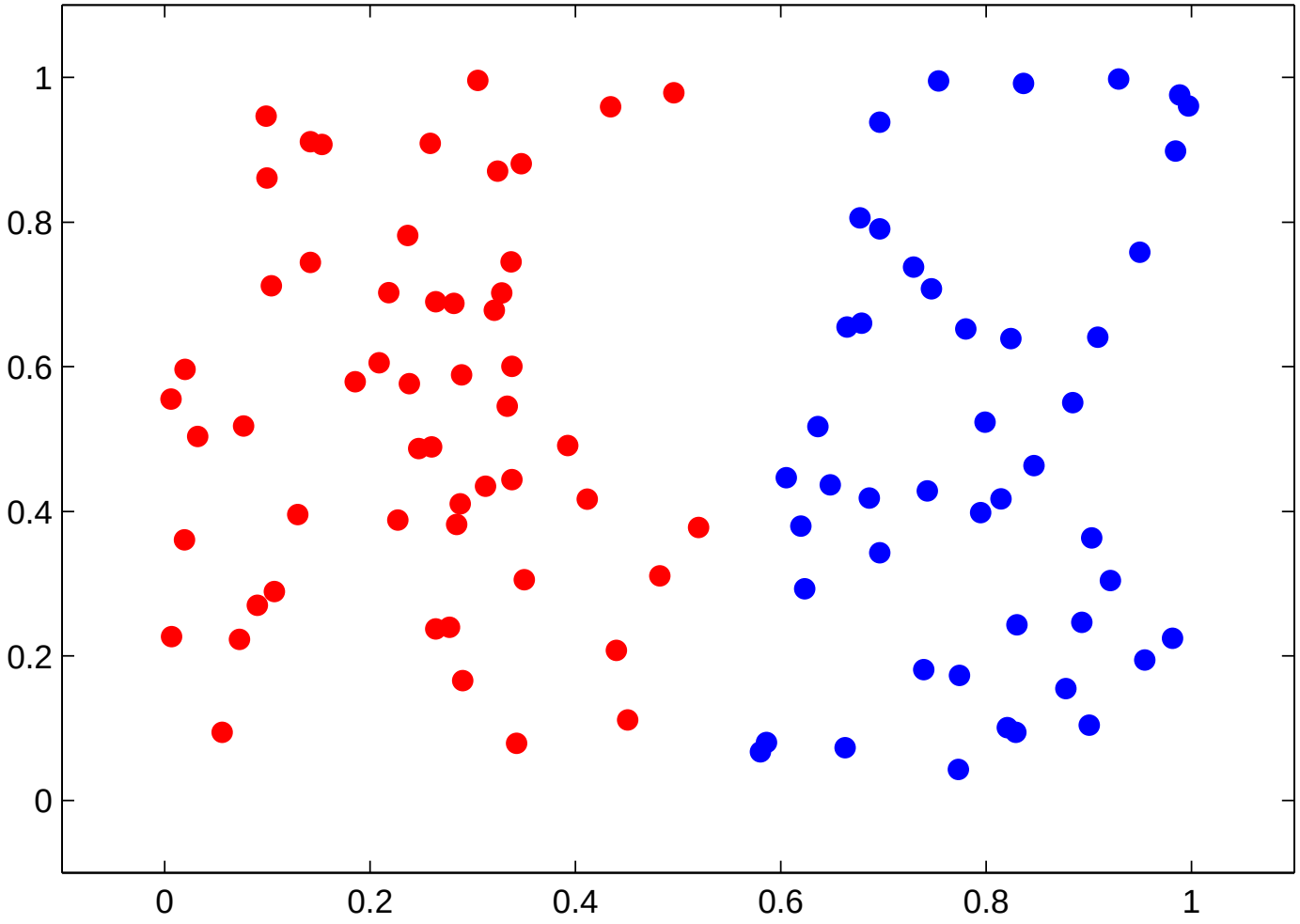
Przycinanie reguł – idea



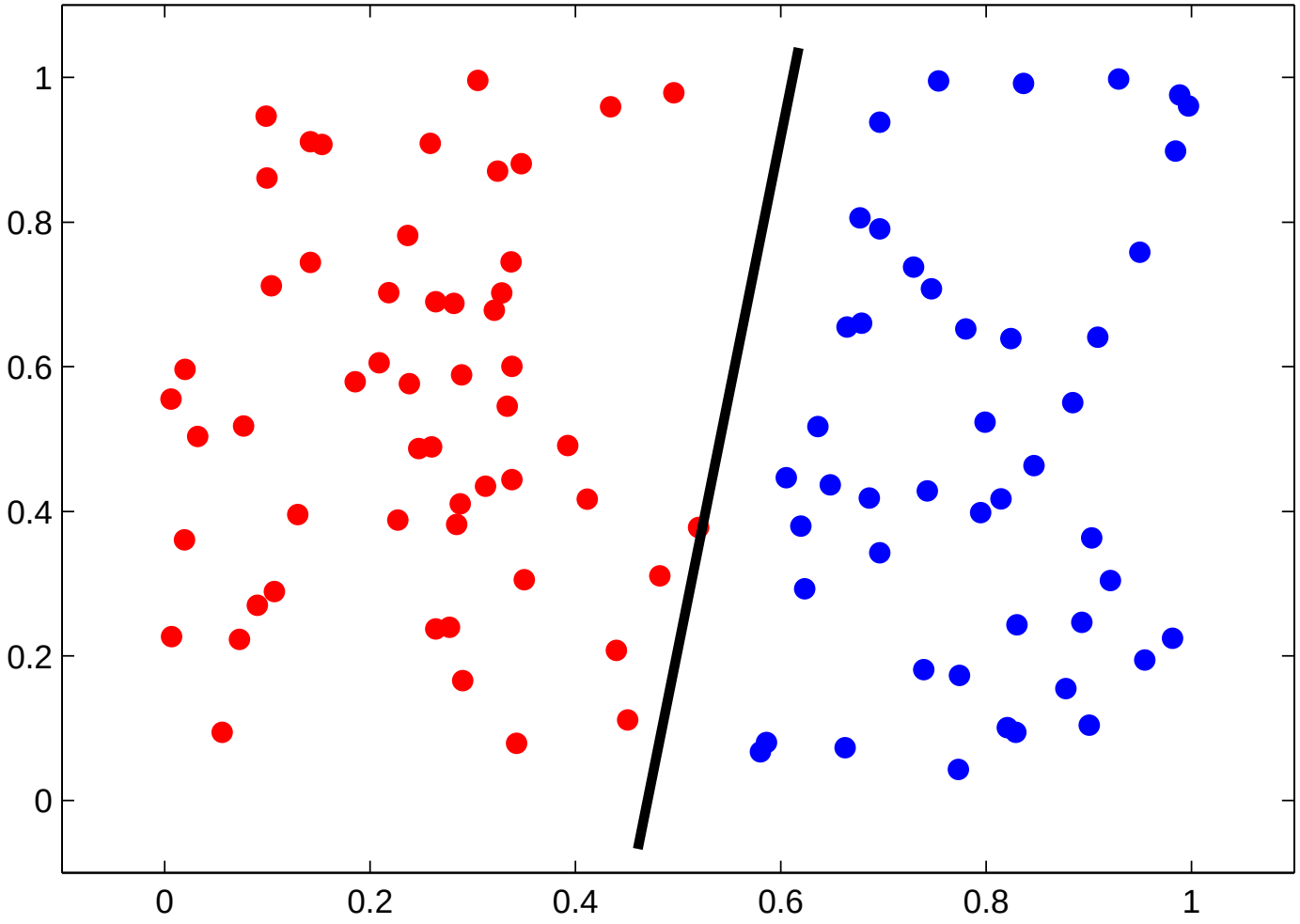
Przycinanie reguł – idea



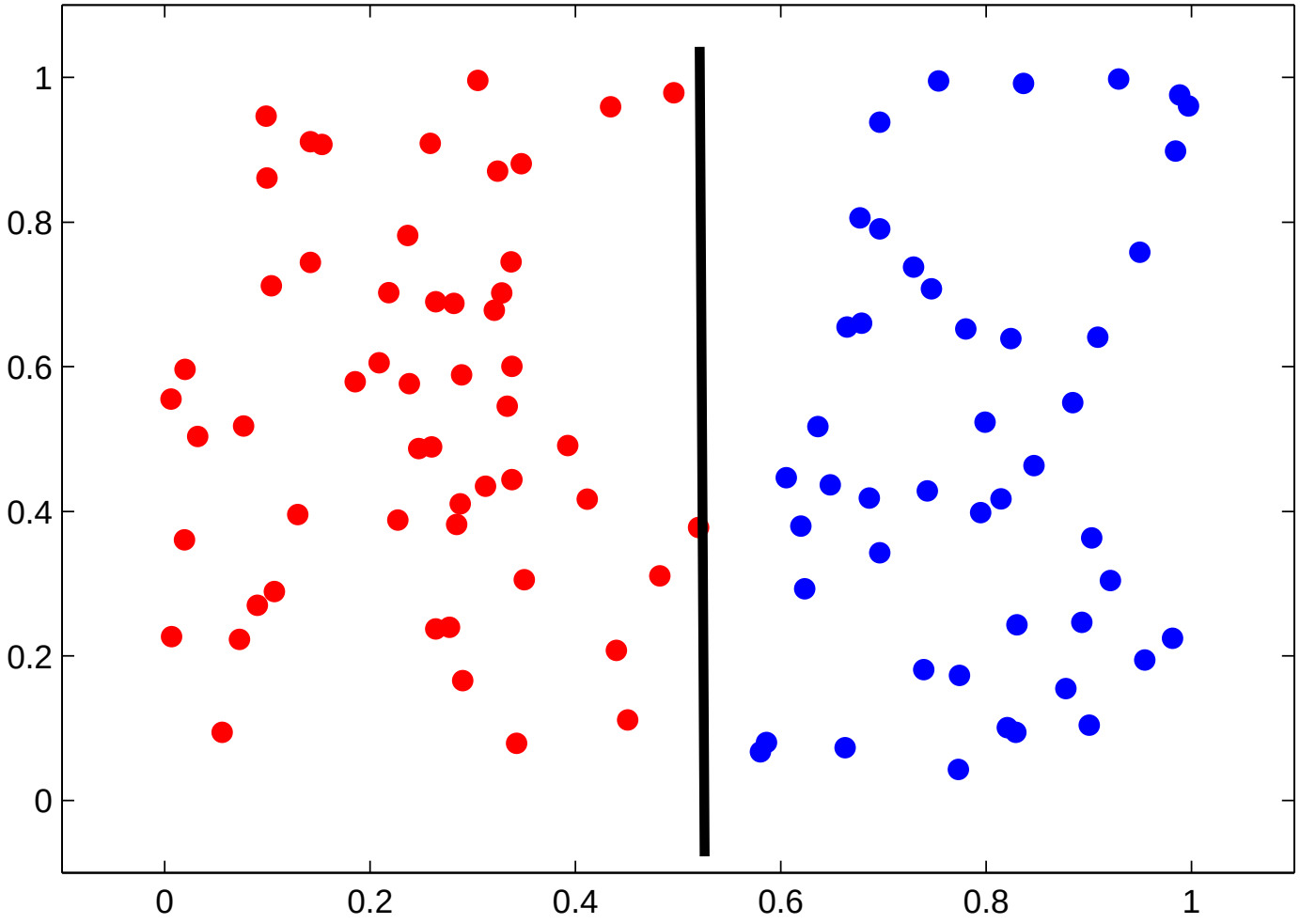
Uogólnianie reguł – idea



Uogólnianie reguł – idea



Uogólnianie reguł – idea



Porównanie wyników

dataset	accuracy avg (std)		
	ORG	*ORG2.0*	*ORG2.1*
2d	96.0(1.5)	98.4(1.0)	98.8 (0.98)
d2d	84.3(3.1)	99.0(1.1)	99.3 (0.64)
3d	98.2(1.2)	98.3(1.6)	98.7 (1.27)

dataset	avg. rules number			total number of cond. elem.		
	ORG	*ORG2.0*	*ORG2.1*	*ORG*	*ORG2.0*	*ORG2.1*
2d	2.0	2.0	3.0	3.0	16.0	8.0
d2d	3.0	2.0	2.0	6.0	16.0	6.0
3d	2.0	2.0	2.0	2.0	32.0	4.0

Porównanie wyników

dataset	accuracy avg (std)		
	ORG	*ORG2.0*	*ORG2.1*
iris	94(4.6)	84(8.4)	90.7 (7.42)
balance	92(2.4)	80(4.2)	76.6 (4.7)
Ripley	81(8.4)	65(11.6)	74.4 (10.76)
breast w.	97(1.7)	92(3.0)	89.9 (4.86)

dataset	avg. rules number			total number of cond. elem.		
	ORG	*ORG2.0*	*ORG2.1*	*ORG*	*ORG2.0*	*ORG2.1*
iris	3.1	3.0	4.5	5.2	96.0	28.4
balance	6.0	3.0	3.0	12.0	96.0	16.0
Ripley	2.0	6.6	5.2	4.0	28.8	15.0
breast w.	3.0	2.0	2.0	19.0	324.0	36.9

Porównanie wyników

dataset	average accuracy (std)		
	PART	*JRIP*	*ORG2.1*
d2d	93.6	95.5	99.3 (0.64)
3d	94.5	94.8	98.7 (1.27)
2d	96.3	95.5	98.8 (0.98)

dataset	avg. rules number			total number of. cond. elem.		
	PART	*JRIP*	*ORG2.1*	*PART*	*JRIP*	*ORG2.1*
d2d	14	9	3.0	37	18	8.0
3d	13	8	2.0	25	19	6.0
2d	10	10	2.0	22	18	4.0

Porównanie wyników

dataset	average accuracy (std)		
	PART	*JRIP*	*ORG2.1*
Ripley	85.6	85.6	74.4(10.76)
balance	83.5	81.0	76.6(4.70)
breast w.	95.5	95.6	89.8(4.86)
iris	94.0	95.3	90.7(7.42)
vehicle	71.5	69.2	73.9(4.31)

dataset	avg. rules number			total number of. cond. elem.		
	PART	*JRIP*	*ORG2.1*	*PART*	*JRIP*	*ORG2.1*
Ripley	5	2	5.2	7	1	15.0
balance	47	11	3.0	147	30	16.0
breast w.	11	6	2.0	21	10	36.9
iris	3	3	4.5	3	4	28.4
vehicle	29	16	7.9	108	40	1006.2

CHIRA

Publikacja

Sikora M., Gudyś A.: CHIRA - Convex Hull Based Iterative Algorithm of Rules Aggregation, Fundamenta Informatice 123:143-170, 2013

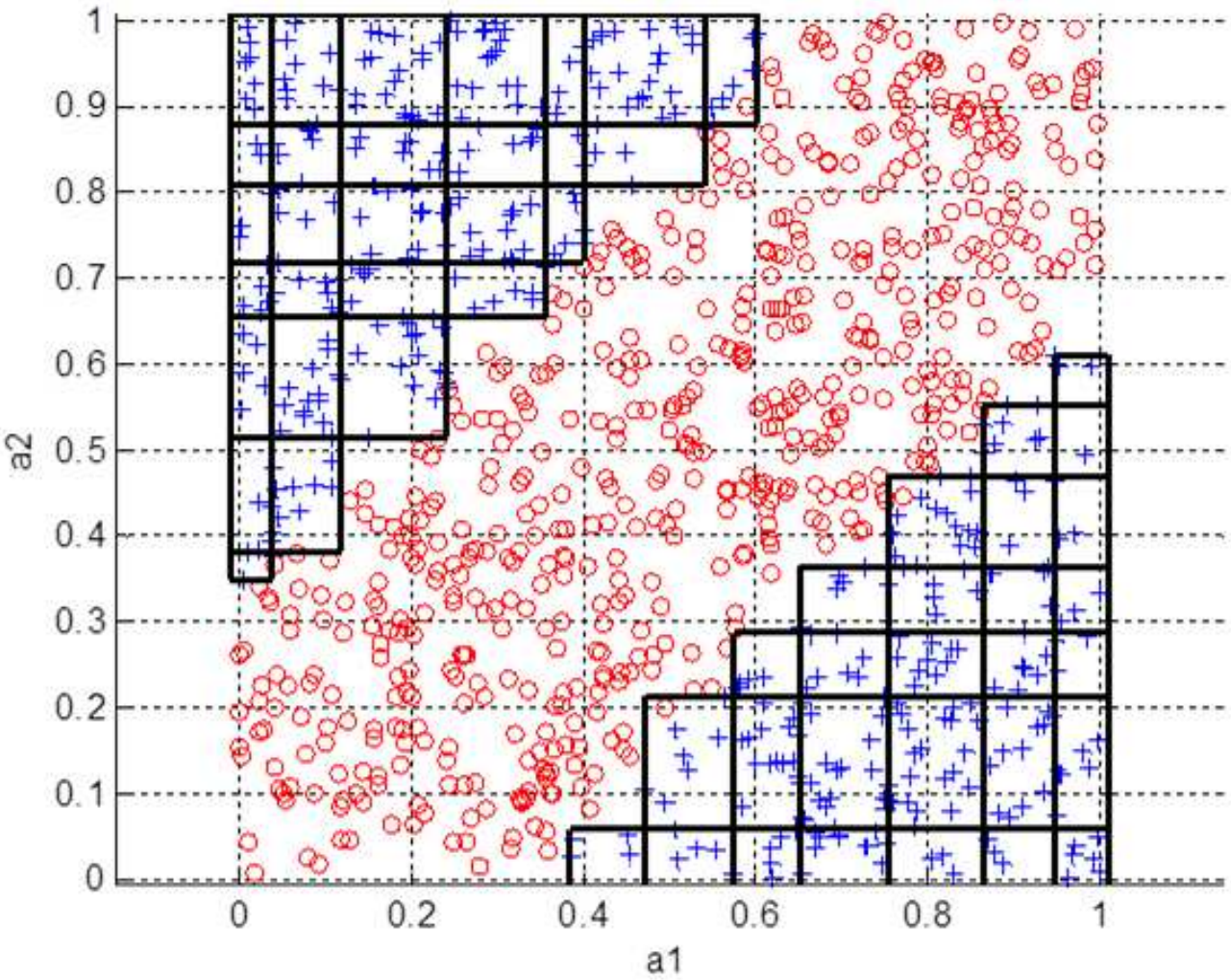
Główne idee:

- przycinanie reguł: eliminacja zbędnych warunków metodą wspinaczki;
- uogólnianie skośnych warunków: eliminacja zmiennych z warunku, jeśli jej współczynnik jest niski.

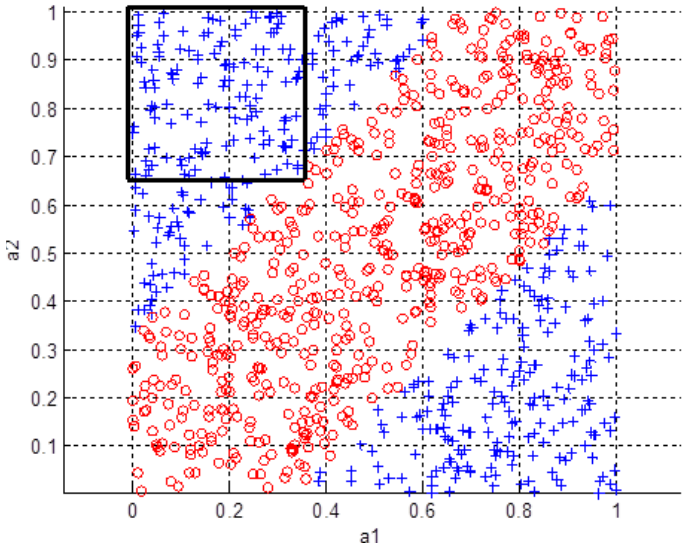
Implementacje w:

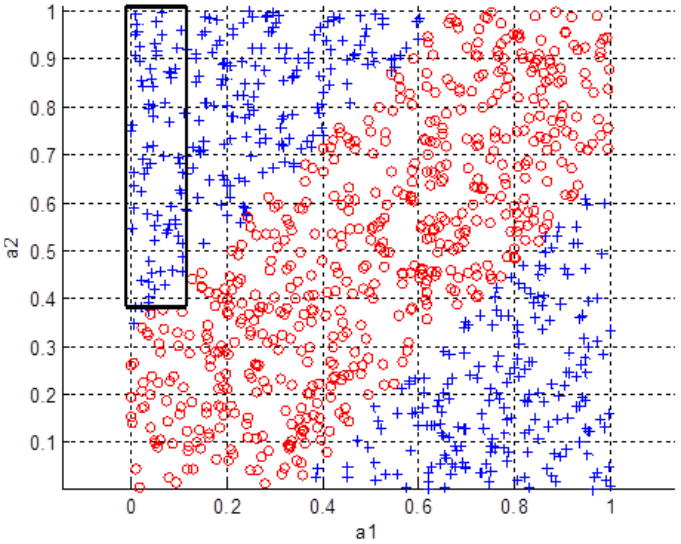
- Matlab
- Weka

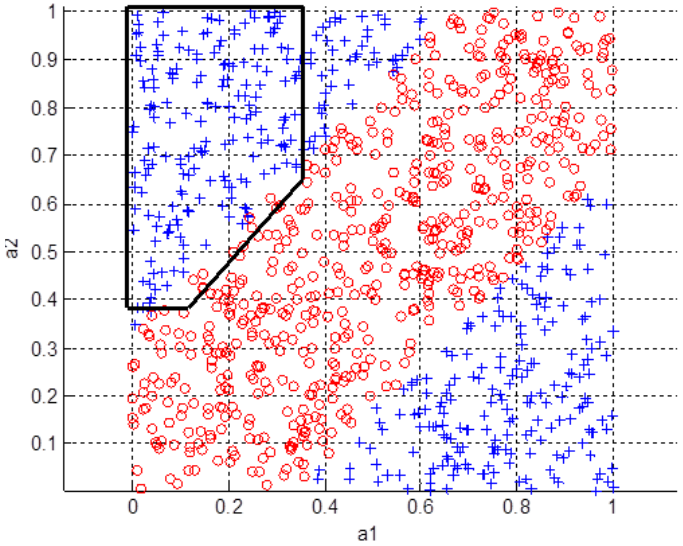
CHIRA - surowe reguły



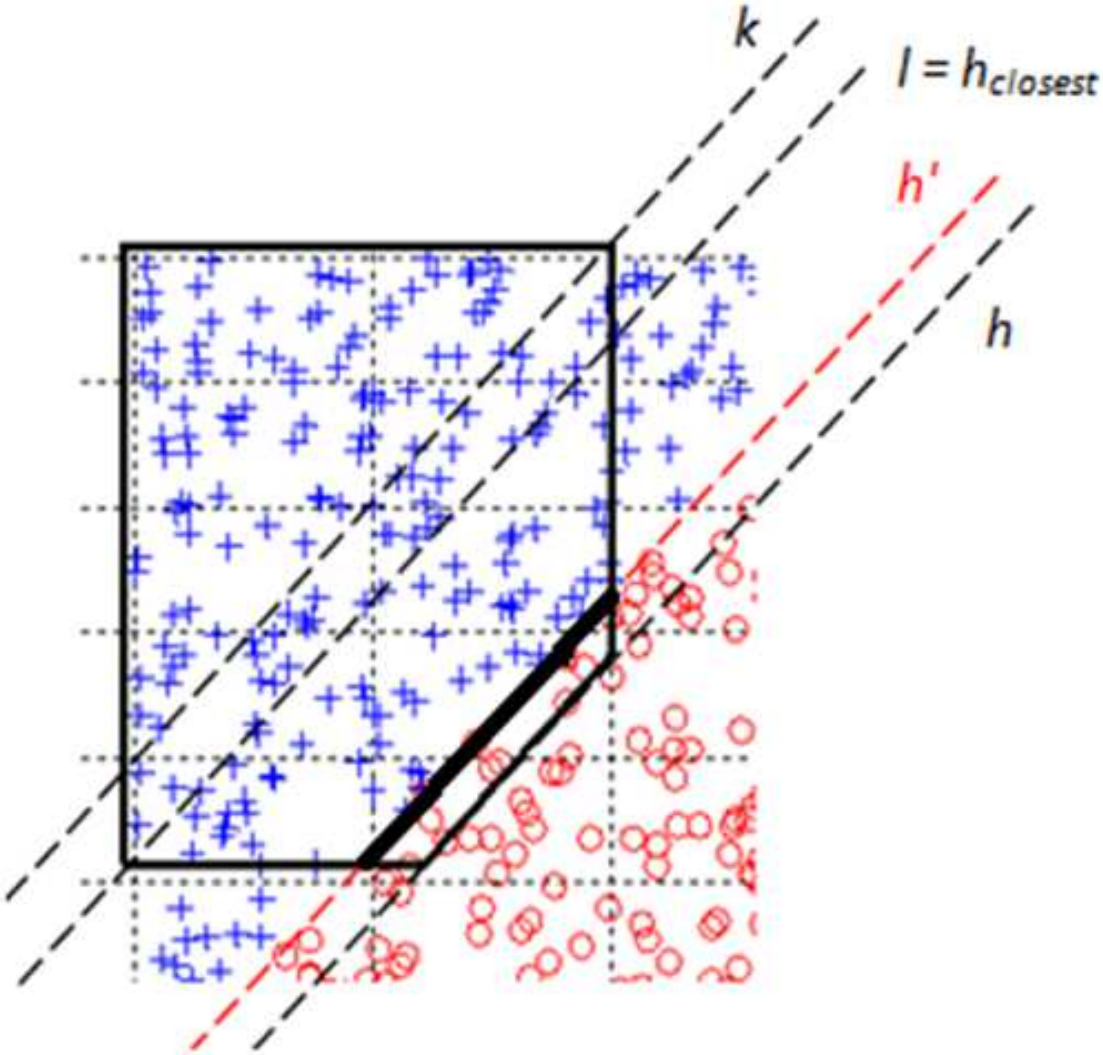
CHIRA - agregacja



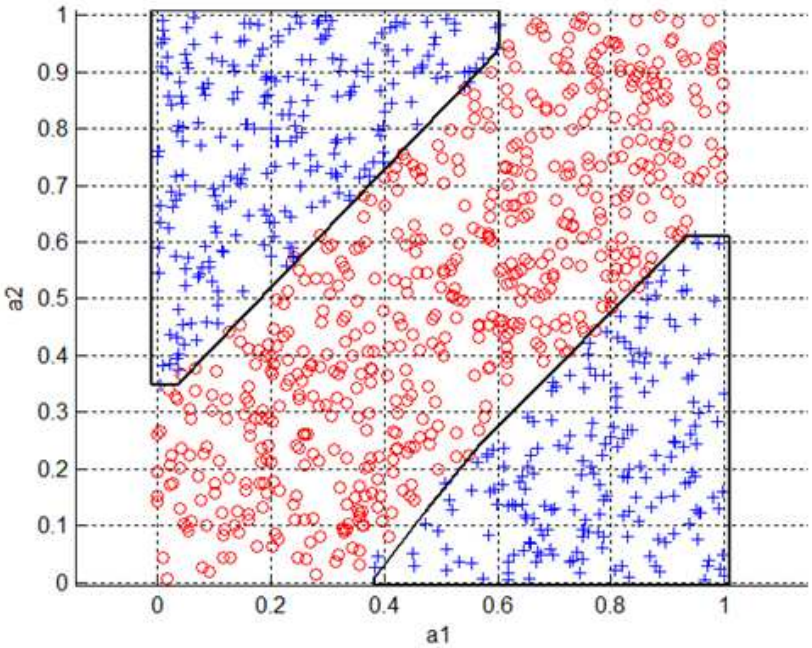


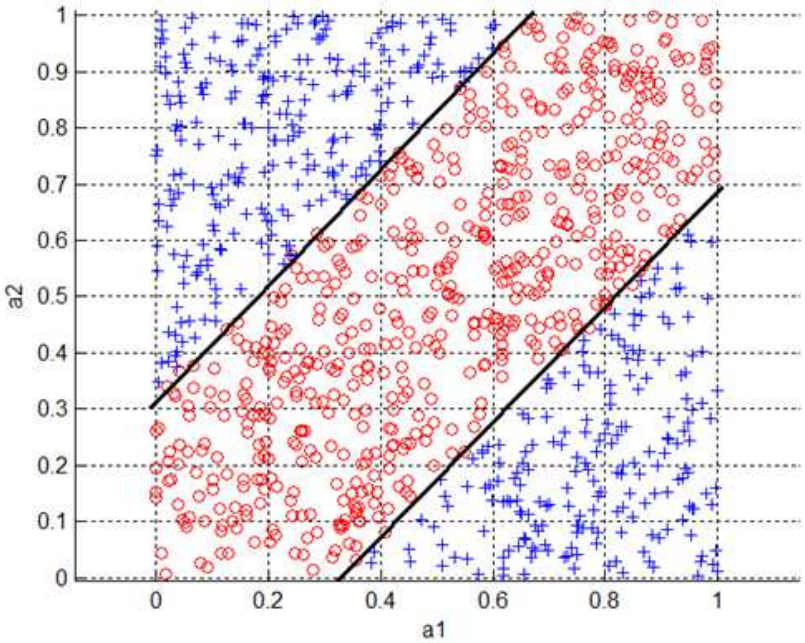


CHIRA - tuning

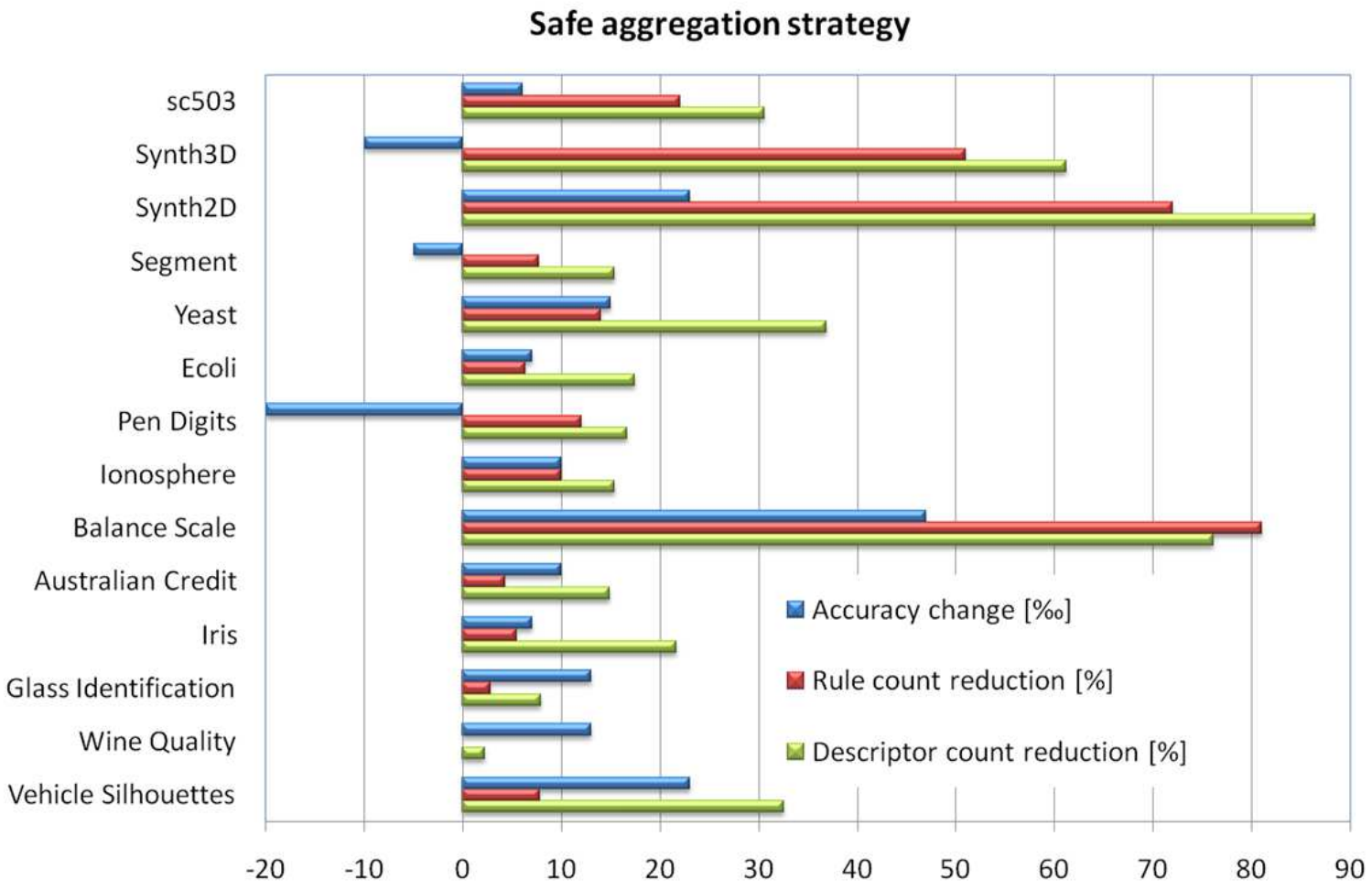


CHIRA - przycinanie





CHIRA - wyniki



Advanced Data Analysis
in Applications Group

ADAA

strona internetowa: http://adaa.polsl.pl
spotkania: czwartki ok 14:00, aula F
2. piętro wydziału Automatyki Elektroniki i Informatyki (Politechnika Śląska)
ul. Akademicka 16, 44-100 Gliwice
kontakt: dr inż. Michał Kozielski (Michal.Kozielski@polsl.pl)