

**Problem eliminacji
nieprzystających elementów
w zadaniu rozpoznania wzorca**

Marcin Luckner

Wydział Matematyki i Nauk Informatycznych
Politechnika Warszawska

Elementy nieprzystające

$$\phi_\alpha(x) = i.$$



$$\exists x \in X \forall i \phi_\alpha(x) = i \wedge x \notin C_i$$

Definicja odrzucania

Klasyfikacja

$$\phi: X \xrightarrow{\alpha} \{1, 2, \dots, n\}$$

$$\phi_{\alpha}(x) = i.$$

$$\phi_{\alpha}(x, \bigcup_{j=1}^n L_j) = i$$

Odrzucanie

$$R: X \xrightarrow{\beta} \{1, 2, \dots, n\} \cup \{0\}$$

$$R(x) = ?$$

Błędy

$$E_i = \sum_{C_i} p(\phi_{\alpha}(x) = j|x)p(x) \text{ dla } j \neq i.$$

$$E_i = \sum_{C_i} \frac{R(x)}{j} p(\phi_{\alpha}(x) = j|x)p(x) \text{ dla } j \neq i$$

$$RE_i = \sum_{C_i} \frac{i - R(x)}{i} p(\phi_{\alpha}(x) = i|x)$$

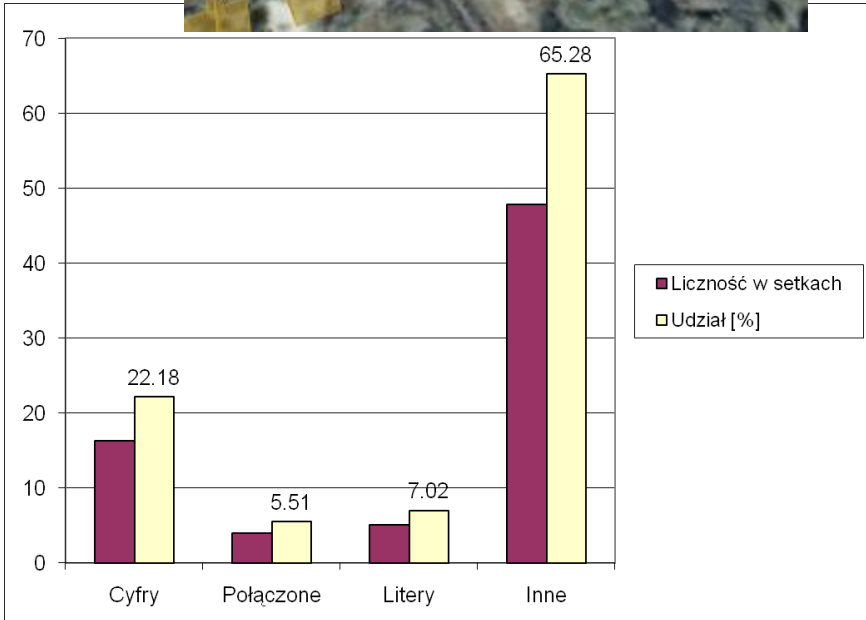
Praktyczne zastosowania

Miasto Stołeczne Warszawa

Miasto Stołeczne WARSZAWA

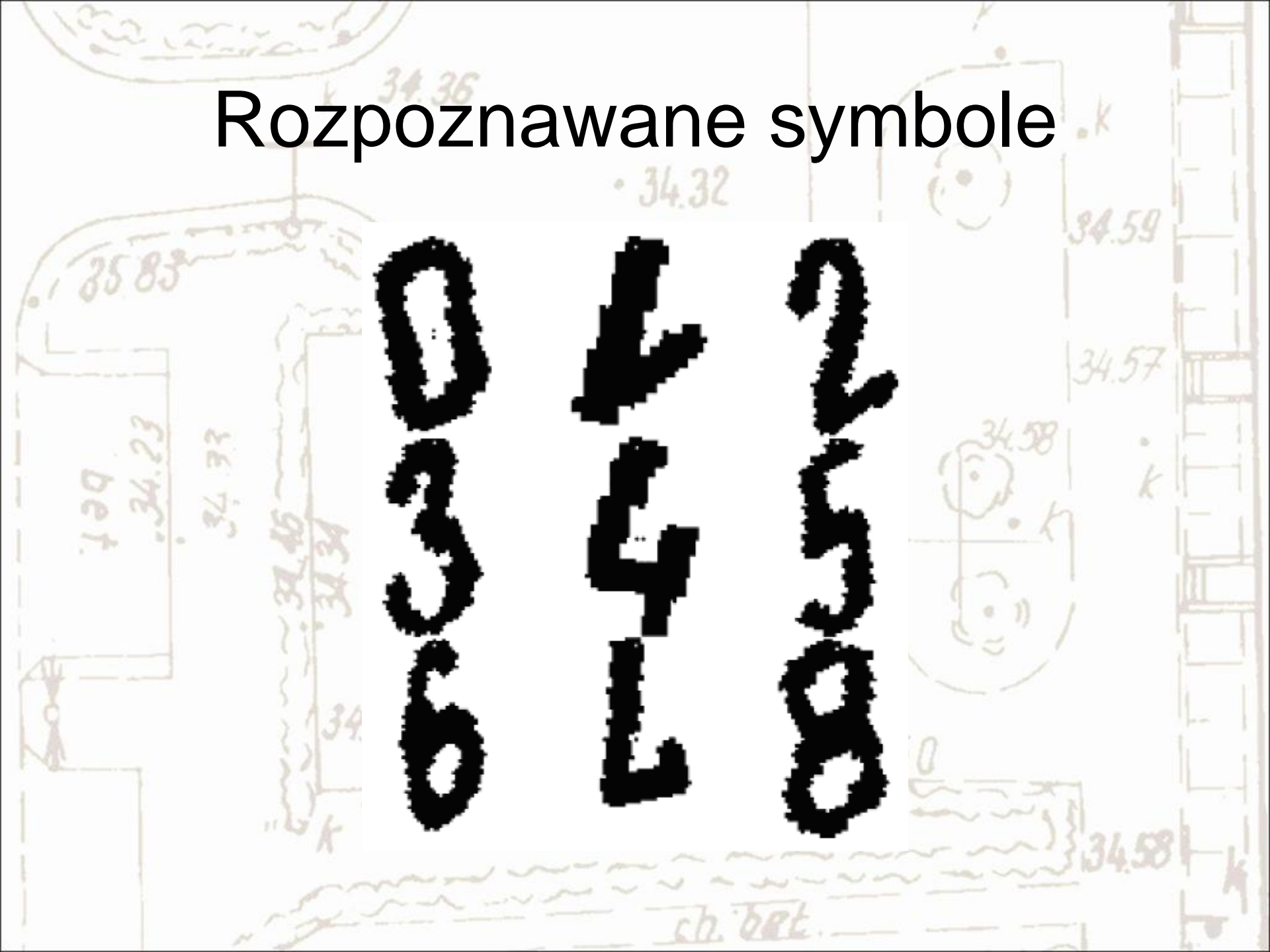


Biuro Geodezji i Katastru
Urzędu Miasta Stołecznego Warszawy
ul. Marszałkowska 77/79
00-668 Warszawa



Rozpoznawane symbole

0 1 2
3 4 5
6 7 8



Wektor cech

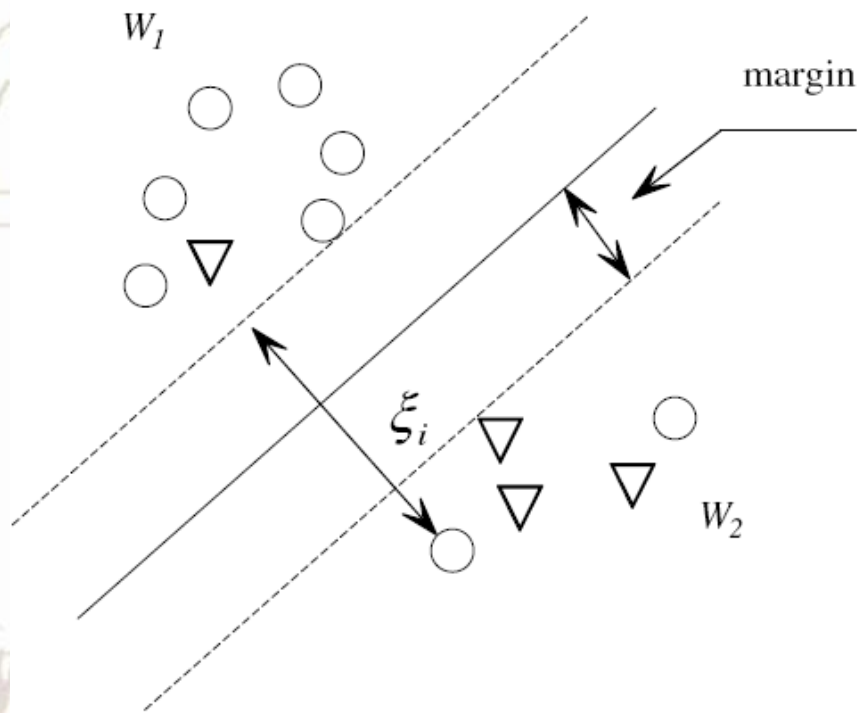
- Cechy z przedziału $[0, 1]$ (197)
 - Proporcje (1)
 - Skierowanie
 - Maksymalna długość (4)
 - Lokalne (64)
 - Udział procentowy (32)
 - Gęstość (16)
 - Tranzycje (16)
 - Offset (32)
 - Histogramy (32)

Oczekiwane cechy klasyfikatora

- Główne
 - Wysoka skuteczność
 - Wielowymiarowość
 - Możliwość adaptacji metod odrzucania
- Drugorzędne
 - Łatwość implementacji
 - Przyzwoita prędkość działania

Support Vector Machine

[Vapnik 98]



$$f(x) = \sum_{i=1}^M a_i y_i K(x_i, x) + b$$

Liniowa:

$$x^T x_i,$$

Wielomianowa:

$$(\gamma x^T x_i + c)^{\text{stopień}},$$

RBF:

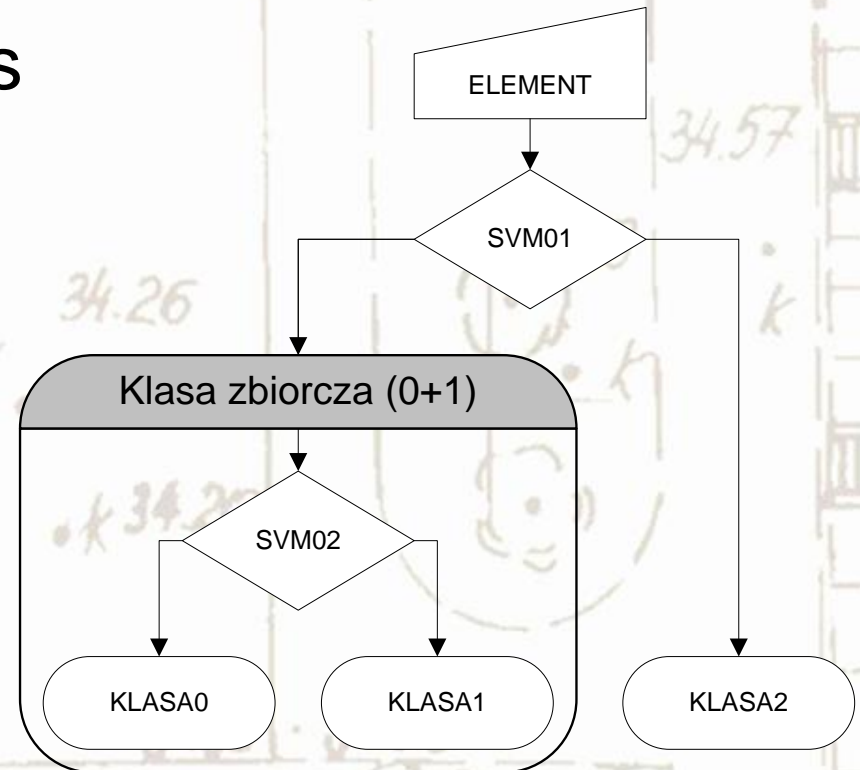
$$\exp(-\gamma |x - x_i|^2),$$

Sigmoidalna:

$$\text{tgh}(\gamma x^T x_i + c).$$

Drzewo SVM

- Umożliwia rozpoznawanie wielu klas
- SVM dzielą dane na klasy i klasy zbiorcze
- Liście odpowiadają przypisaniu klasy



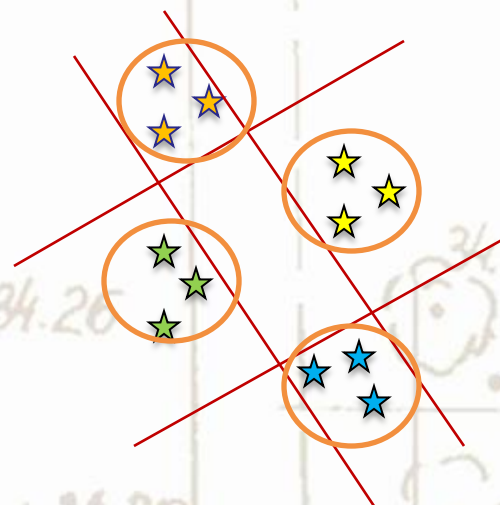
Klasyfikacja wielowymiarowa

[Hsu 02]



Jeden kontra wszyscy

Implementacja drzewiasta [Han 07]



Jeden kontra jeden

DAGSVM [Platt 00]

Wady metod

- Jeden kontra wszyscy
 - Nieoptymalny podział między poszczególnymi klasami
 - Utrudniony podział liniowy



- Jeden kontra jeden
 - Znaczna liczba podziałów
 - Zbędne podziały



Metody nie analizują rozkładu klas w przestrzeni danych.

Podobieństwo klas

Teza: Najbliższe klasy są do siebie najbardziej podobne

Metryka:

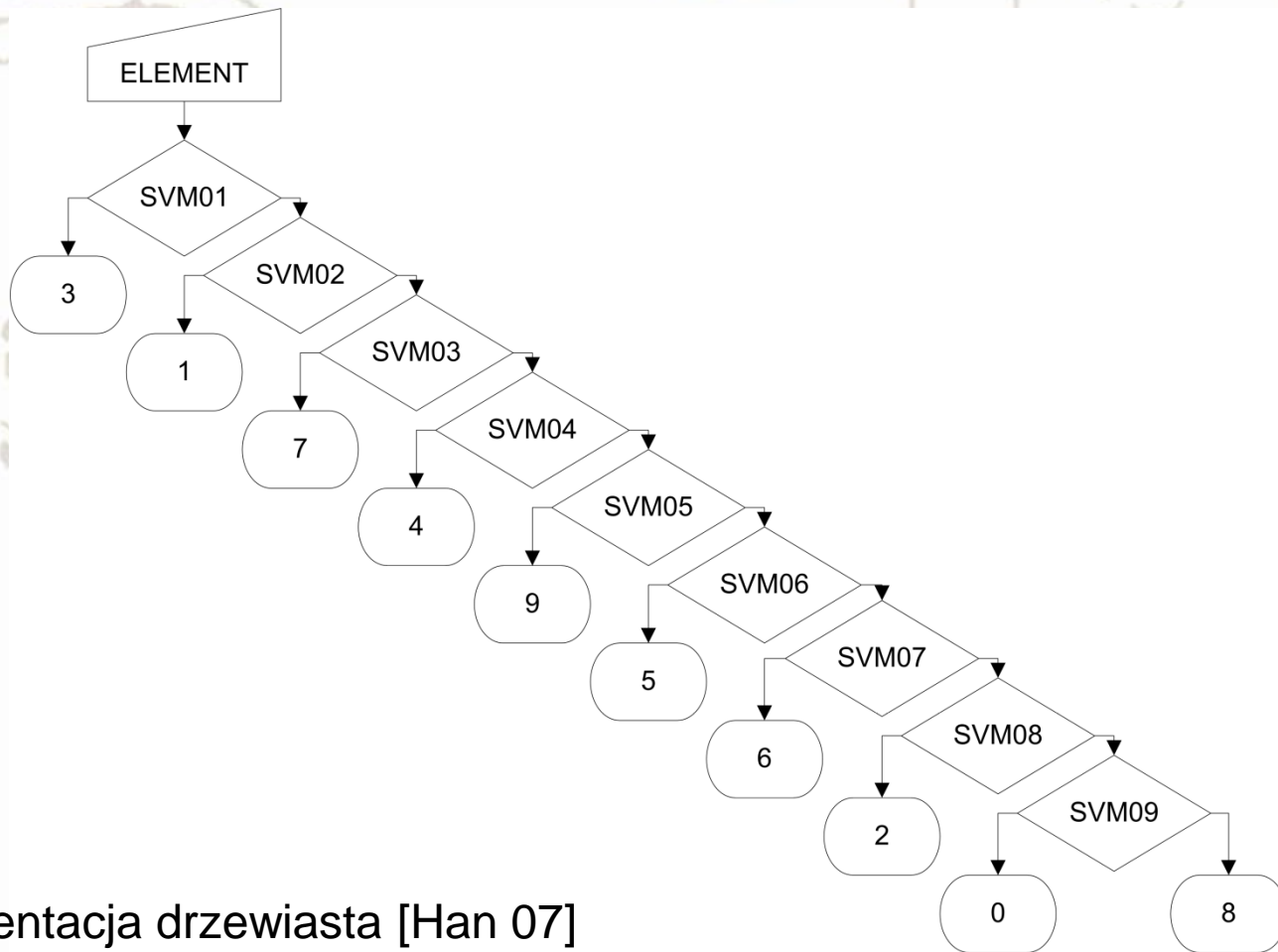
Euklidesowa:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2},$$

Odległość między klasami:

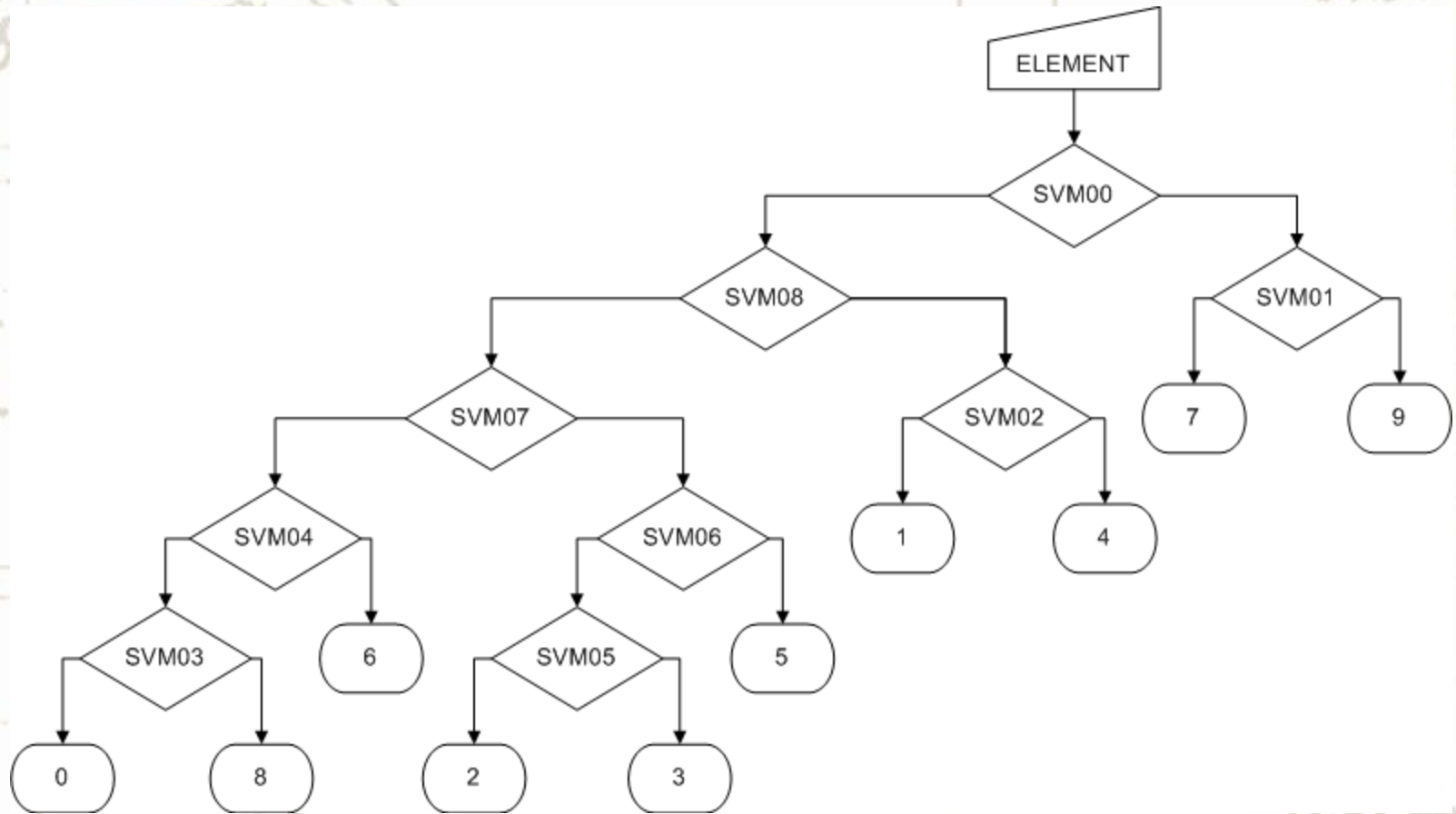
$$d(C_X, C_Y) = d\left(\frac{1}{n_{C_X}} \sum_{x \in C_X} x, \frac{1}{n_{C_Y}} \sum_{y \in C_Y} y\right)$$

Drzewo jeden kontra wszyscy

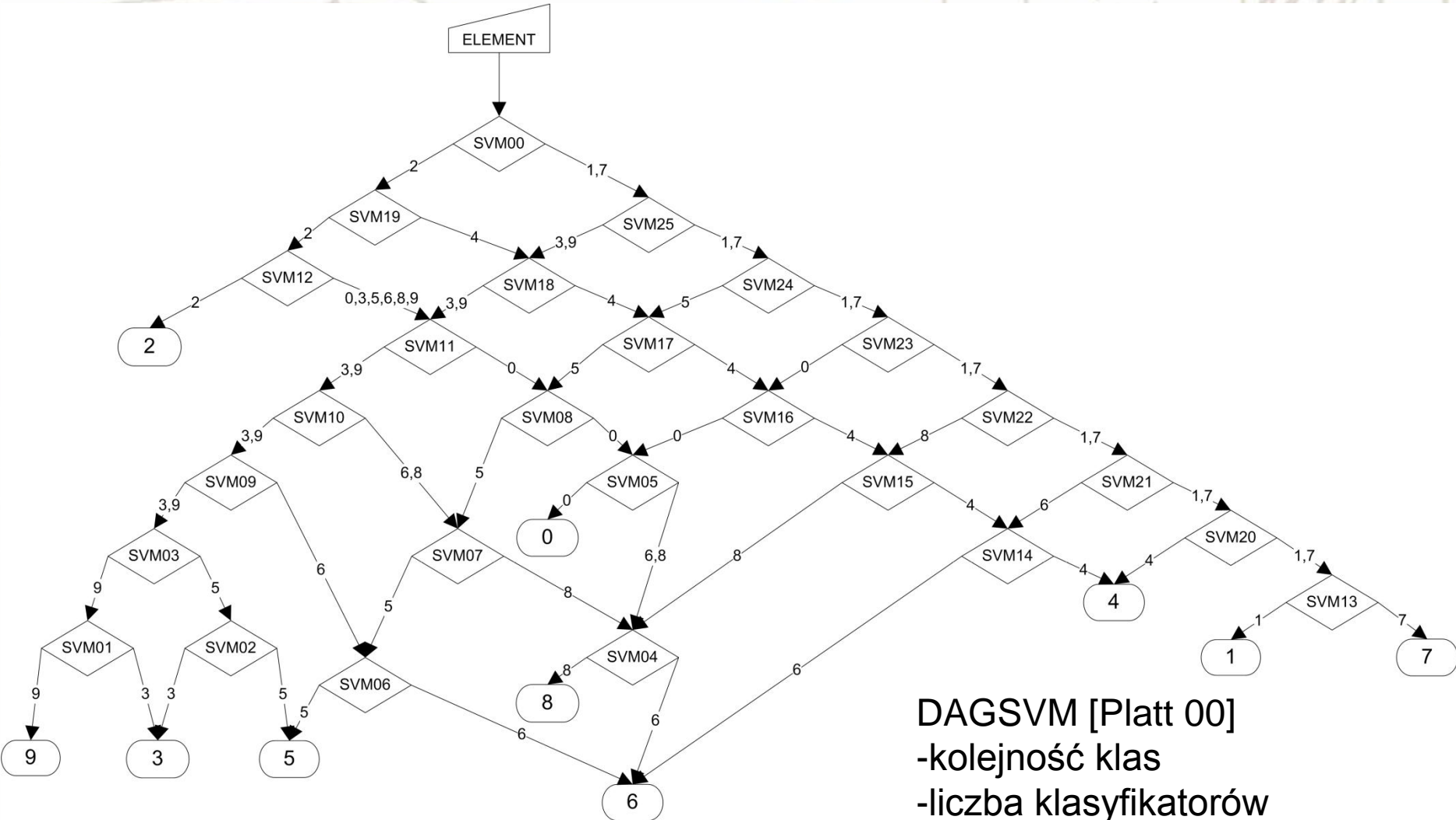


Implementacja drzewiasta [Han 07]
- Nieustalona kolejność klas

Drzewo grupujące



Graf grupujący



Zbiór testowy

Klasyfikacja

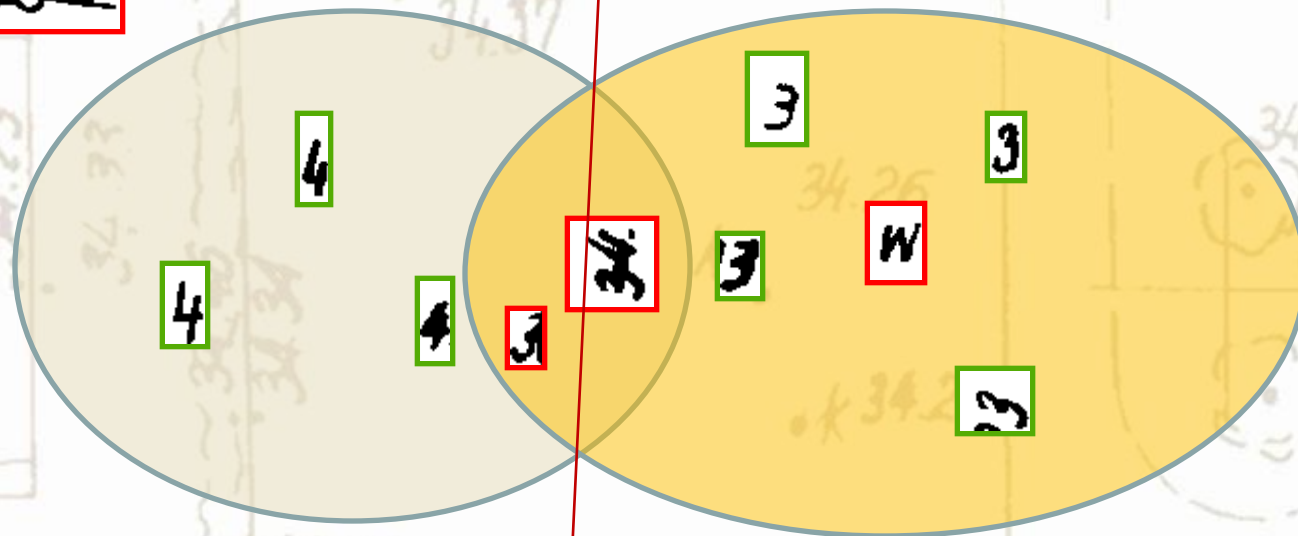
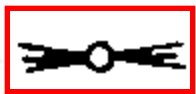
Klasa	Ucząca		Testowa	
	Liczność	Procent [%]	Liczność	Procent [%]
0	236	4.46	87	4.86
1	688	13.00	234	13.07
2	850	16.07	312	17.43
3	1685	31.85	535	29.89
4	452	8.54	172	9.61
5	314	5.93	102	5.70
6	327	6.18	93	5.20
7	273	5.16	86	4.80
8	211	3.99	82	4.58
9	255	4.82	87	4.86
Razem	5291	74.72	1790	25.28

Porównanie metod klasyfikacji

	Metoda				
	MLP	J-K-J	J-K-W	Drzewo	Graf
Wynik klasyfikacji	94.25	95.75	93.35	93.02	95.25
Najgorszy wynik	87.21	82.76	78.43	63.22	83.91
Najlepszy wynik	97.67	97.94	97.67	97.01	97.76
Odchylenie standartowe	3.97	4.95	5.91	9.92	4.08
Liczba klasyfikatorów SVM	-	40	9	9	26

- Graf grupujący może zastąpić metodę jeden-kontra-jeden
- Znacząca redukcja użytych klasyfikatorów
- Wskazane zostaną metody implementacji mechanizmu odrzucania

Lokalizacja błędów w przestrzeni danych



34.42

34.36 · 34.50

34.58

ch. bet.

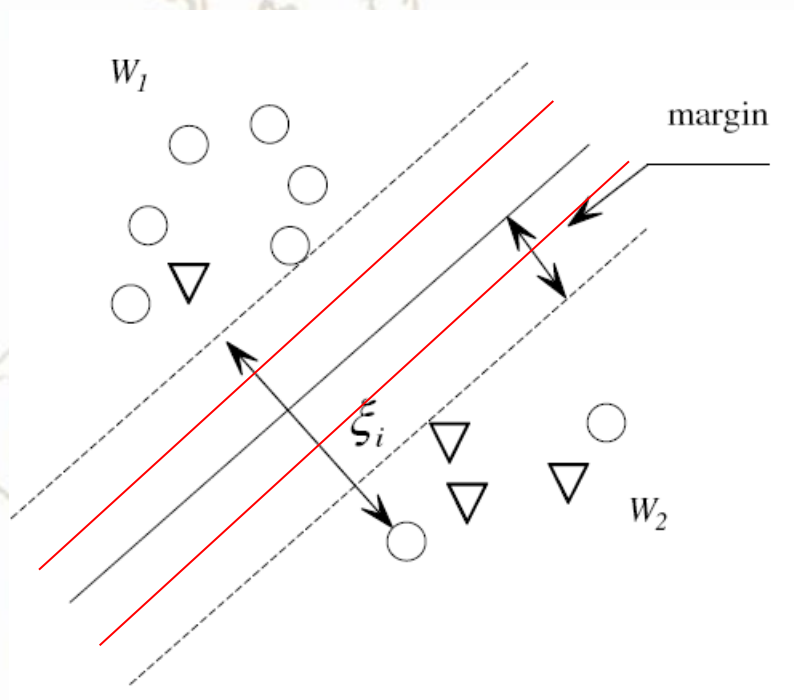
Metody realizacji odrzucania

- Odrzucanie z zastosowaniem progu
 - Przypadki o niskim prawdopodobieństwie prawidłowego rozpoznania
- Odrzucanie jako zadanie klasyfikacyjne
 - Przypadki rozpoznane jako błędne przez alternatywne zadanie klasyfikacyjne
- Odrzucanie z zastosowaniem zbioru ograniczającego
 - Przypadki nienależące do zbioru wyznaczonego przez elementy poprawne
- Metody hybrydowe

Odrzucanie z zastosowaniem progu

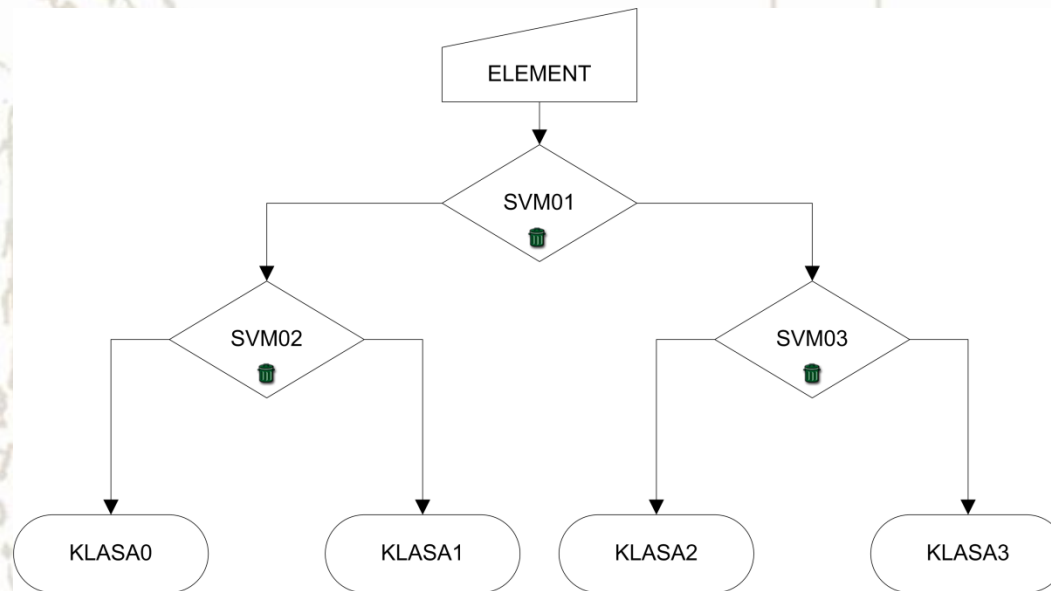
Reguła Chowa [Chow 70]

$$\max_i p_i F(x|C_i) < (1 - t) \sum_{j=1, j \neq i}^n p_j F(x|C_j)$$



$$\rho(x) = \begin{cases} i & p(\omega_i|x) \geq (1 - t) \\ j & p(\omega_j|x) \geq (1 - t) \\ 0 & \text{w p.p.} \end{cases}$$

Zastosowanie dla drzewa



Odrzucanie we wszystkich węzłach:

$$R(x) = \text{sgn}\left(\prod_{K \in P} \rho_K(x)\right) \phi_\alpha(x, \bigcup_{j=1}^n L_j)$$

Przy ostatnim klasyfikatorze:

$$R(x) = \text{sgn}(\rho_C(x)) \phi_\alpha(x, \bigcup_{j=1}^n L_j)$$

Odrzucanie jako zadanie klasyfikacyjne

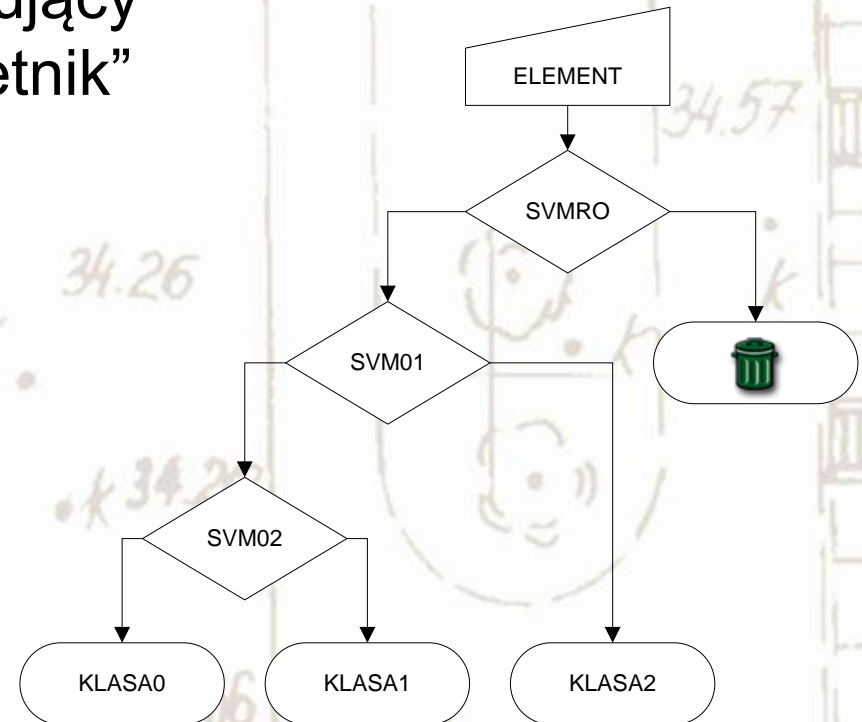
- Elementy odrzucane tworzą dodatkową klasę
 - Globalną
 - Lokalne
 - Każda rozpoznawana klasa ma przypisaną klasę elementów nieprzystających

Odrzucanie z zastosowaniem klasy globalnej

Nowy korzeń drzewa wskazujący na stary korzeń i klasę „śmiećnik”

$$\phi'_\alpha(x, \bigcup_{j=0}^n L_j) = i$$

$$R(x) = \phi'_\alpha(x, \bigcup_{j=0}^n L_j)$$



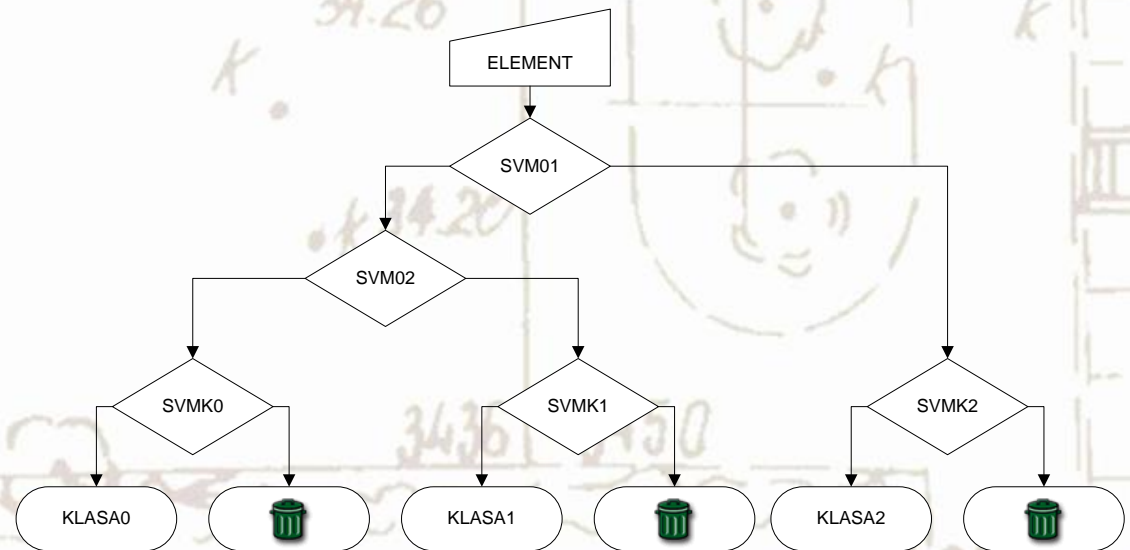
Odrzucanie z zastosowaniem klas lokalnych

Klasa „śmietnik” dla każdego liścia, dodatkowy podział na elementy zaakceptowane i odrzucone

Sposób realizacji: dodatkowy SVM dla każdej z klas

$$\phi'_\alpha(x, \bigcup_{j=-n, j \neq 0}^n L_j) = i$$

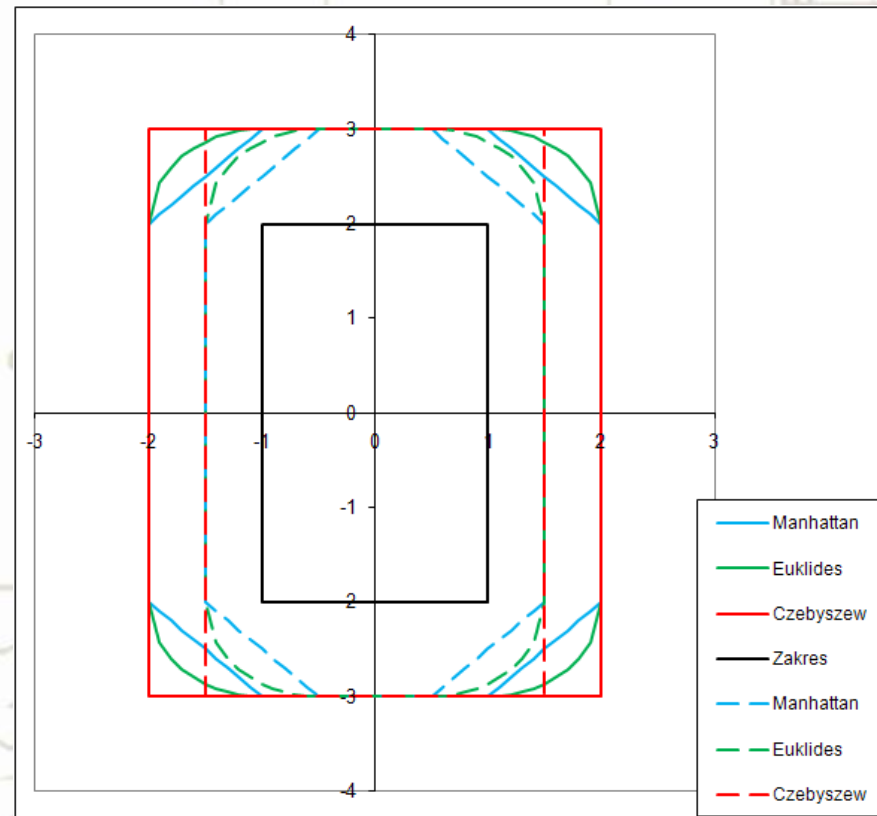
$$R(x) = \begin{cases} \phi'_\alpha(x) & \phi'_\alpha(x) > 0 \\ 0 & \phi'_\alpha(x) < 0 \end{cases}$$



Odrzucanie z zastosowaniem zbioru ograniczającego

- Akceptacja na podstawie przynależności do zbioru
- Zbiór wyznaczany przez elementy uczące klasyfikatora

$$R(x) = \begin{cases} \phi_\alpha(x) & \phi_\alpha(x) = i \wedge x \in \Gamma_i \\ 0 & \phi_\alpha(x) = i \wedge x \notin \Gamma_i \end{cases}$$

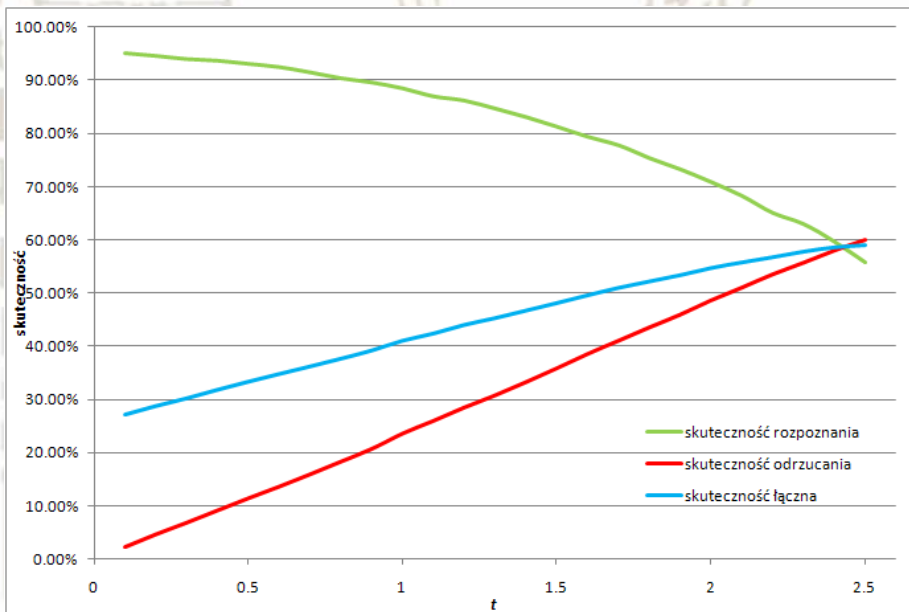


Klasyfikacja z odrzucaniem

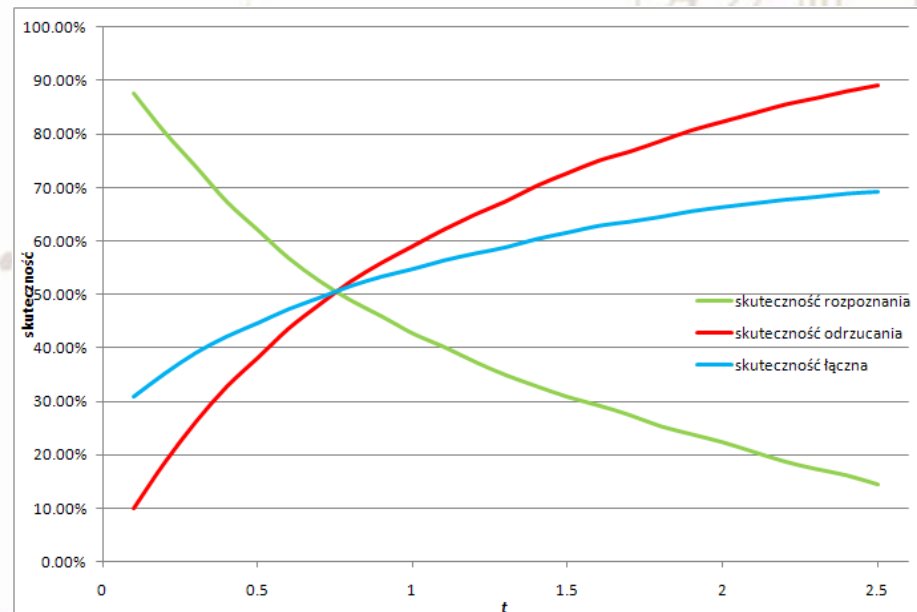
- Podział na grupę testową i uczącą w stosunku 1/3
- Grupa testowa - 6680 przykładów
 - 1790 z zadania klasyfikacji
 - 4890 spoza klas rozpoznawanych
- Skuteczność osiągnięta przy odrzuceniu wszystkich przykładów **73.20%**
- Maksymalna skuteczność, uwzględniająca skuteczność klasyfikatora, **98.73%**

Zastosowanie progu

Odrzucanie ograniczone do liści



Odrzucanie we wszystkich węzłach



Zbiega do skuteczności osiągniętej przy odrzuceniu wszystkich przykładów **73.20%**

Odrzucanie jako zadanie klasyfikacyjne

Klasa globalna

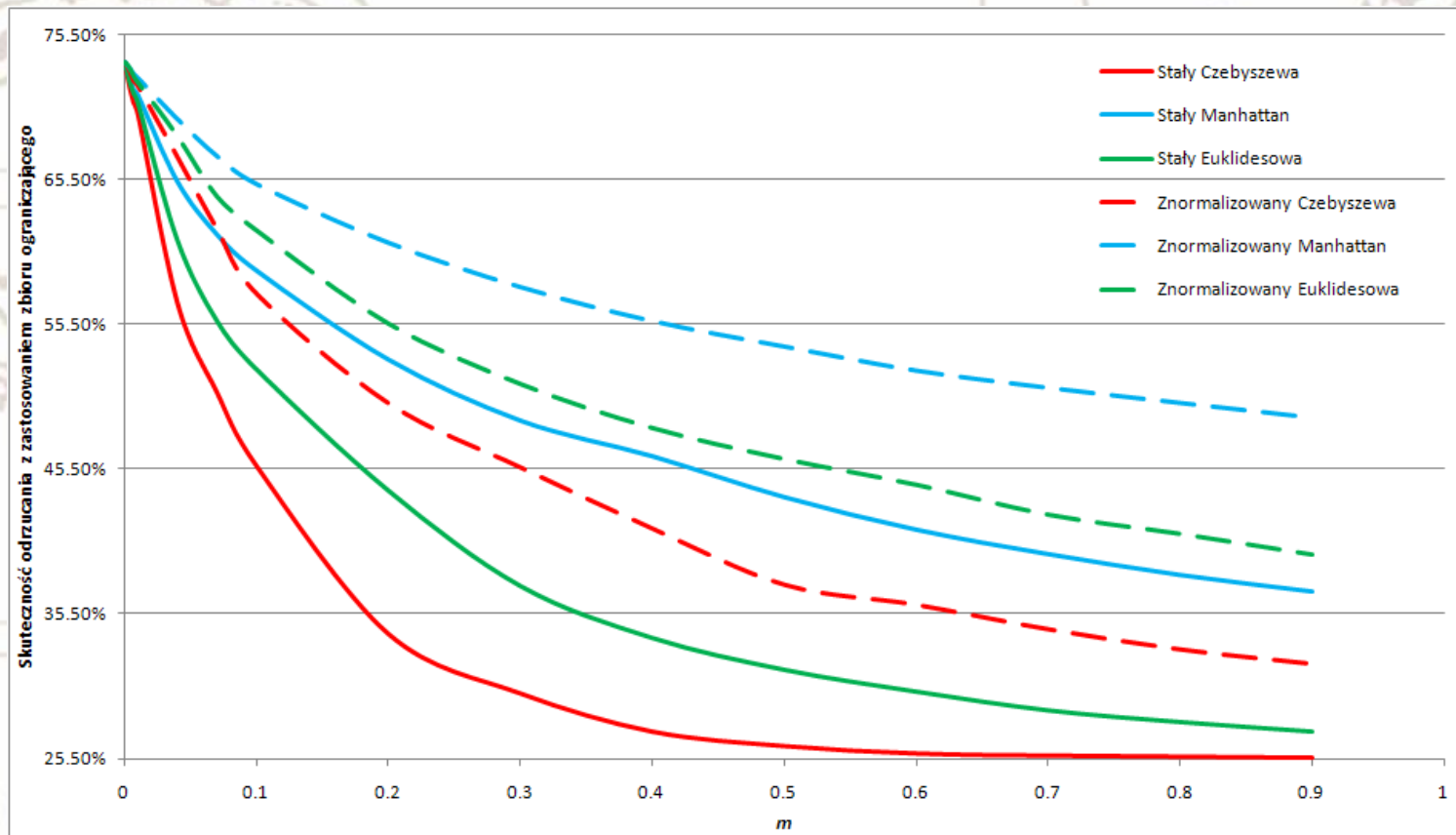
- Rozpoznanie 77.09%
- Odrzucanie 92.70%
- Łącznie **88.52%**

Klasy lokalne

- Rozpoznanie 85.02%
- Odrzucanie 96.03%
- Łącznie **93.08%**

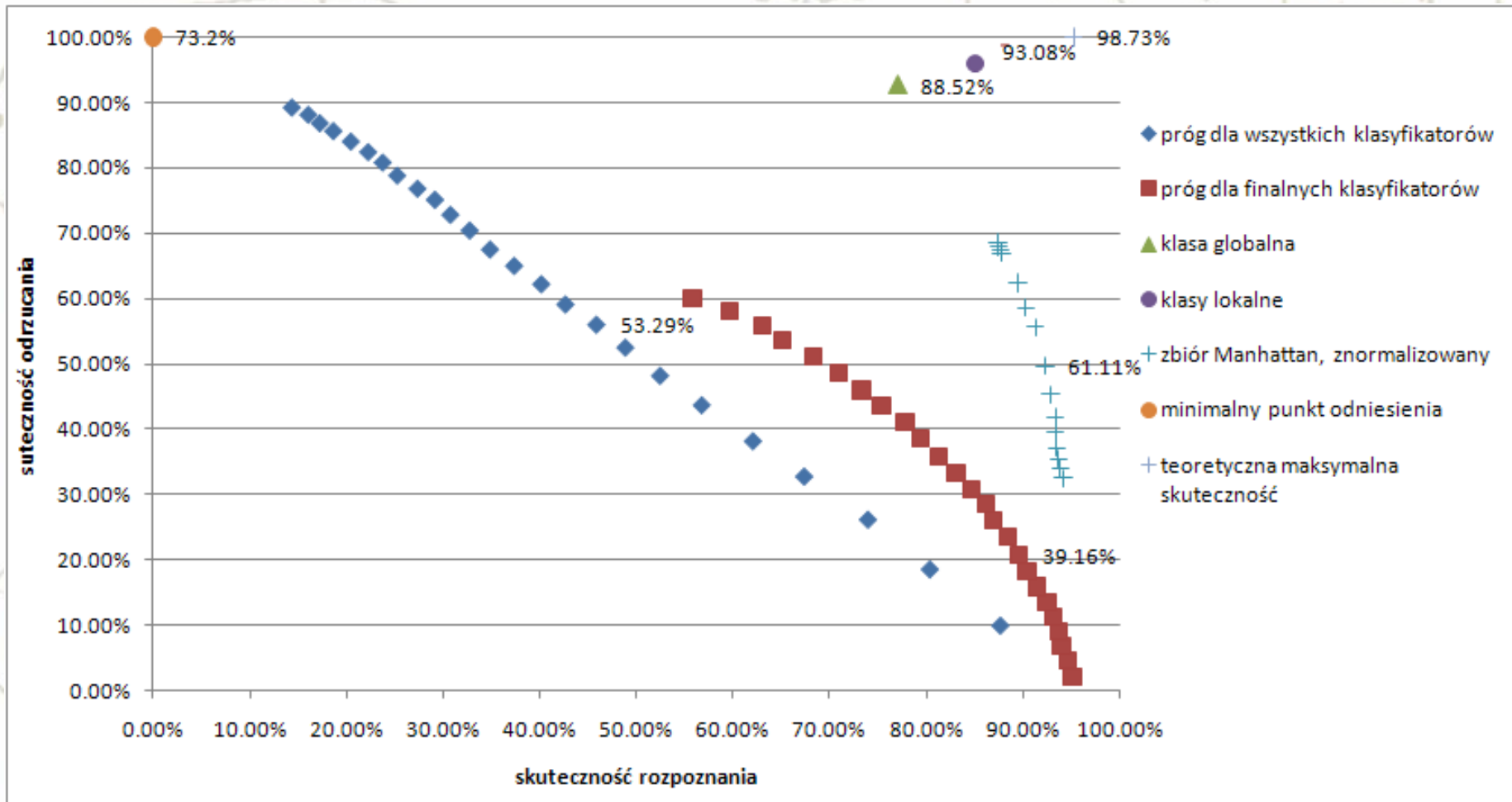
Wada - spadek skuteczności rozpoznania z 95.25%

Zastosowanie zbioru ograniczającego

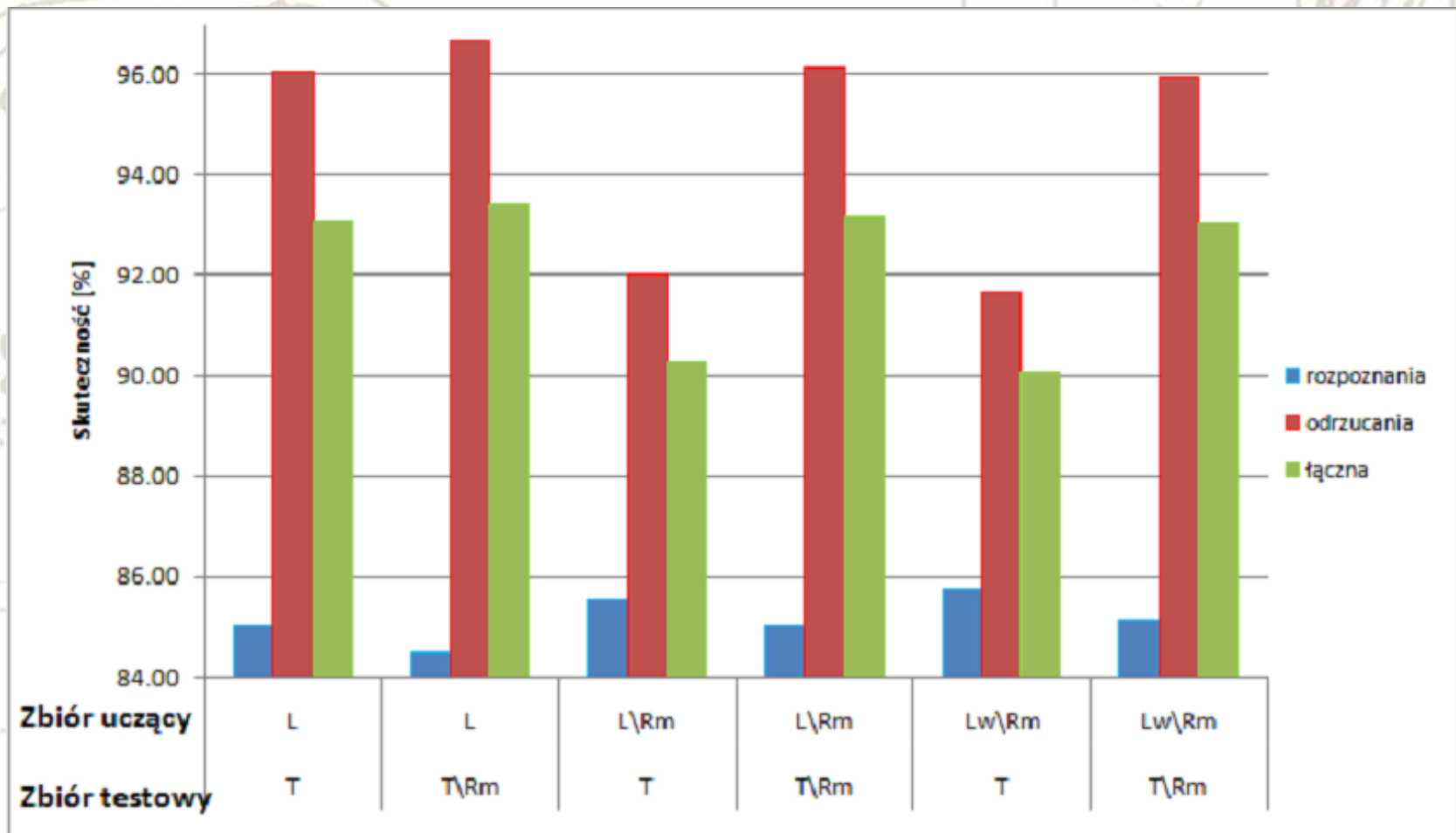


Skuteczność **73.20%** przekroczone tylko dla małych wartości m

Zestawienie metod



Metody hybrydowe



Bibliografia

- [Rhee 95] P. K. Rhee, J. S. Park & Y. G. Kang. *Rotation-invariant recognition of character strings for Korean cadastral map digitizing*. Document Analysis and Recognition, Proceedings of the Third International Conference on, vol. 2, 1995.
- [Vapnik 98] V. Vapnik. *Statistical learning theory*. Wiley, New York, 1998.
- [Hsu 02] C. W. Hsu & C. J. Lin. *A comparison of methods for multiclass support vector machines*. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 13, no. 2, 2002.
- [Han 07] Shunji Han, Wen You & Hui Li. *Application of Binary Tree Multi-class Classification Algorithm Based on SVM in Shift Decision for Engineering Vehicle*. *Control and Automation, 2007*
- [Platt 00] J. Platt, N. Cristianini & J. ShaweTaylor. *Large Margin DAGS for Multiclass Classification*. In S. A. Solla, T.K. Leen & K. R Mueller, editeurs, *Advances in Neural Information Processing Systems 12*, 2000
- [Chow 70] C. Chow. *On optimum recognition error and reject tradeoff*. *Information Theory, IEEE Transactions on*, vol. 16, no. 1, 1970.