



Wrocław University of Technology

Metody tworzenia efektywnych komitetów klasyfikatorów jednoklasowych

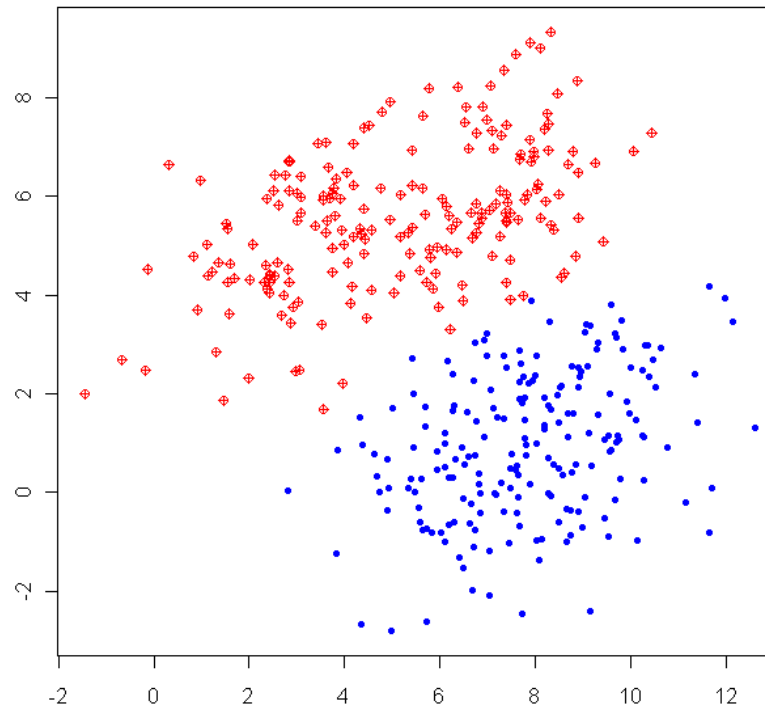
Bartosz Krawczyk

Katedra Systemów i Sieci Komputerowych
Politechnika Wrocławska

e-mail: bartosz.krawczyk@pwr.wroc.pl

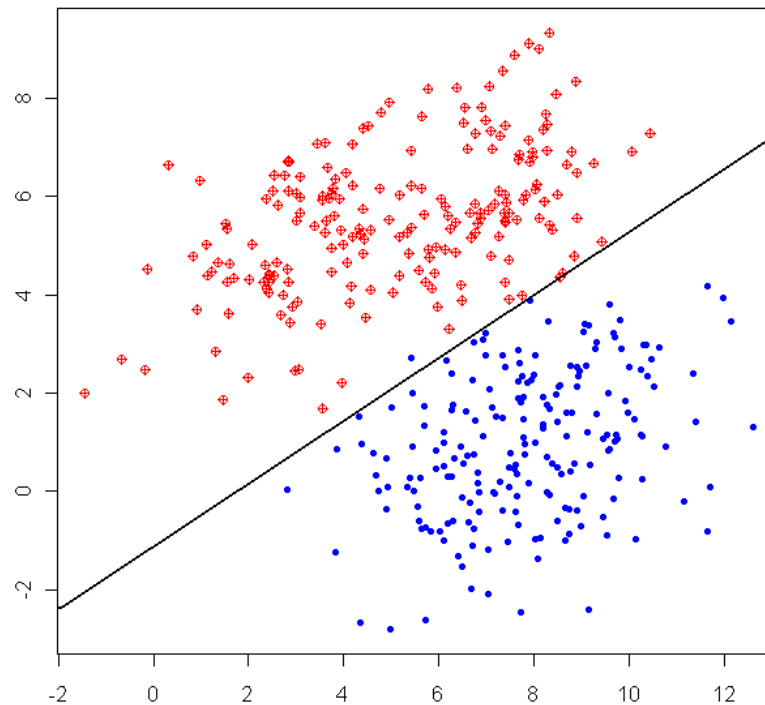
Czym jest klasyfikacja jednoklasowa?

- Celem **klasyfikacji wieloklasowej** jest rozróżnienie między obiektami pochodzącymi z dwóch lub więcej klas, zachowując **jak najmniejszy błąd**.



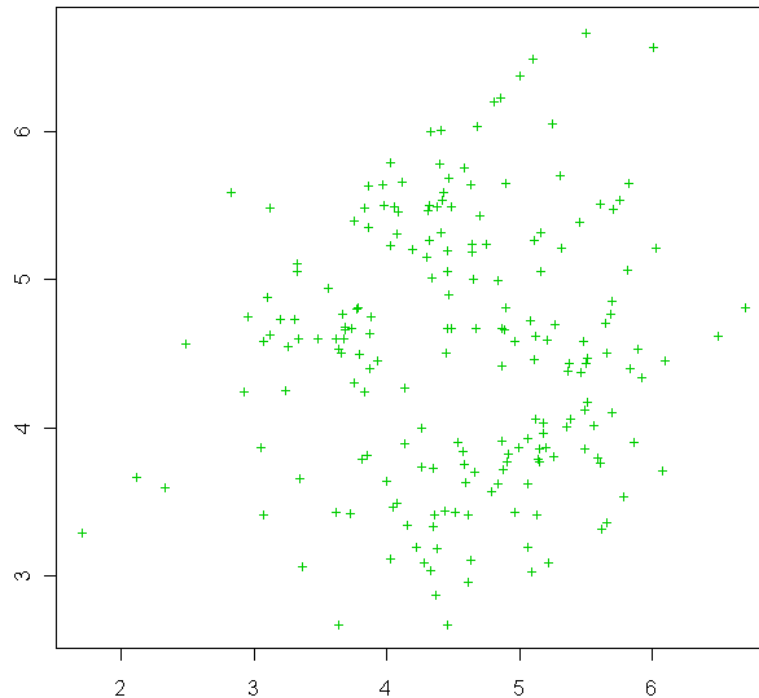
Czym jest klasyfikacja jednoklasowa?

- Celem **klasyfikacji wieloklasowej** jest rozróżnienie między obiektami pochodzącymi z dwóch lub więcej klas, zachowując **jak najmniejszy błąd**.



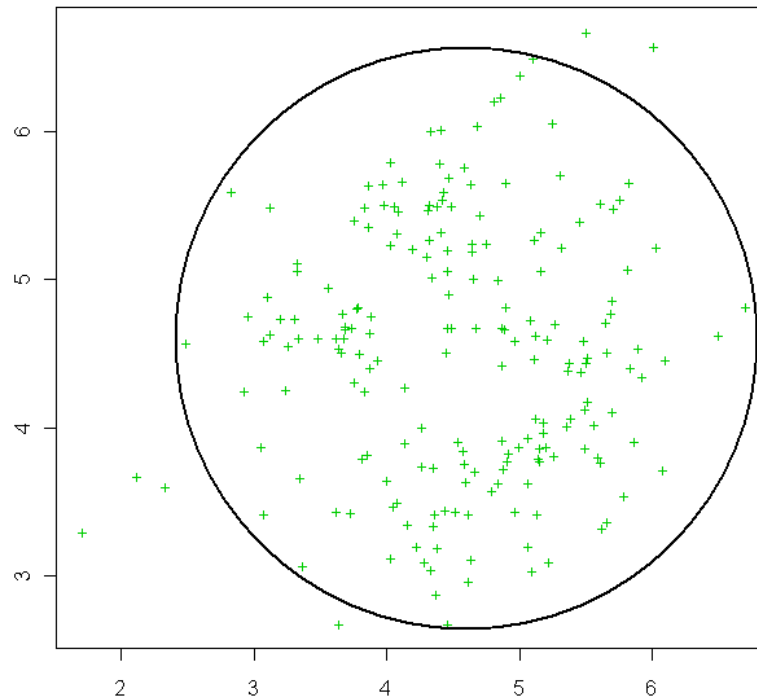
Czym jest klasyfikacja jednoklasowa?

- W przypadku **klasyfikacji jednoklasowej** na etapie uczenia mamy do dyspozycji wyłącznie obiekty **pochozące z jednej klasy**. Nazywa się to nauką przy **braku kontrprzykładów**.



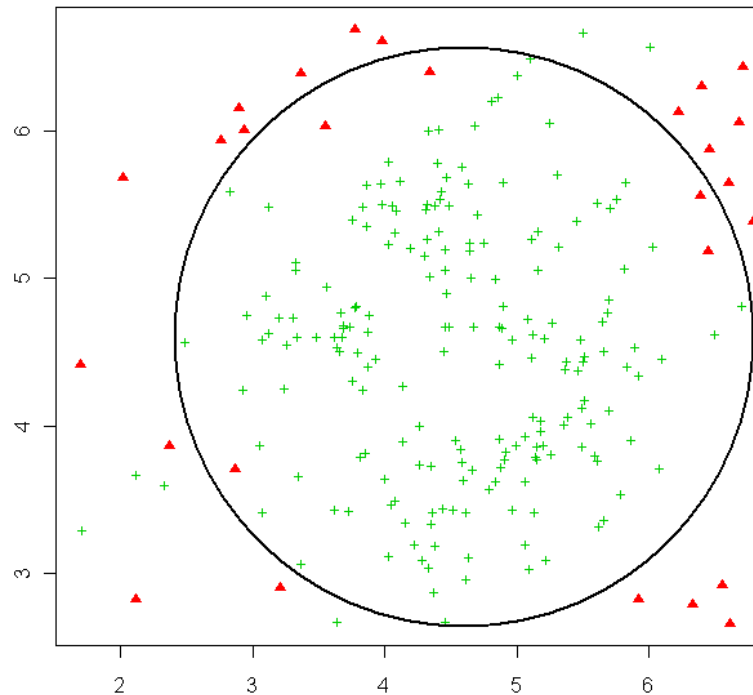
Czym jest klasyfikacja jednoklasowa?

- Jednym z najbardziej popularnych podejść jest estymacja **zamkniętej granicy decyzyjnej**, zawierającej w sobie wszystkie istotne obiekty z klasy celu.



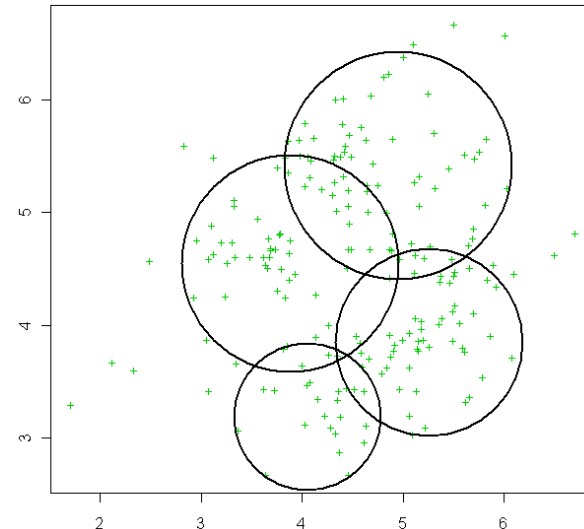
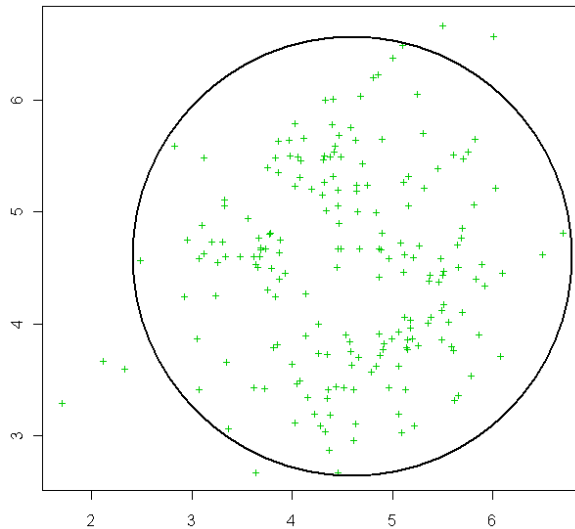
Czym jest klasyfikacja jednoklasowa?

- Zakłada się, że w trakcie eksploatacji klasyfikatora jednoklasowego mogą pojawić się **nowe, nieznane obserwacje**. Nazywa się je **obserwacjami odstającymi**.



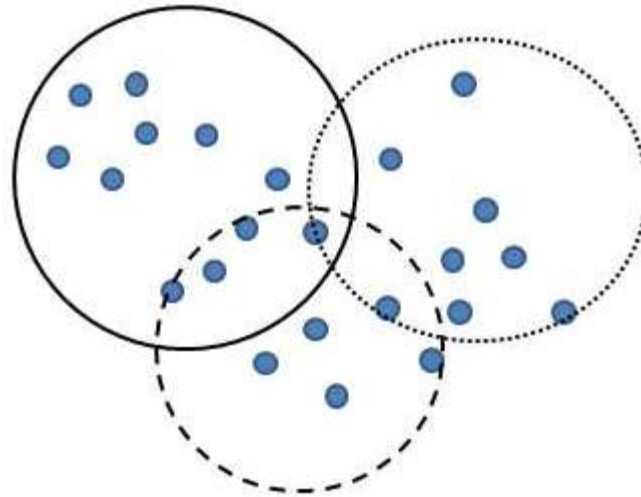
Komitety klasyfikatorów jednoklasowych

- W klasyfikacji jednoklasowej niezbędne są metody poprawiające jakość i stabilność modelu - **komitety klasyfikatorów** są więc obiecującym kierunkiem badań.
- Do tej pory większość prac w tej dziedzinie skupiała się na aplikacjach - istnieje **potrzeba rozbudowania teoretycznych podstaw** tej dziedziny.



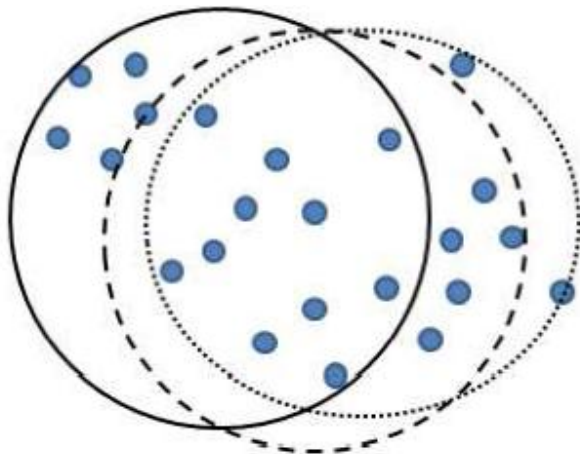
Klasyfikatory w puli

- Aby komitet klasyfikatorów osiągał dobrą skuteczność rozpoznawania, musi być on zbudowany z puli klasyfikatorów które jednocześnie mają **niski błąd klasyfikacji** oraz są **zróźnicowane względem siebie**.

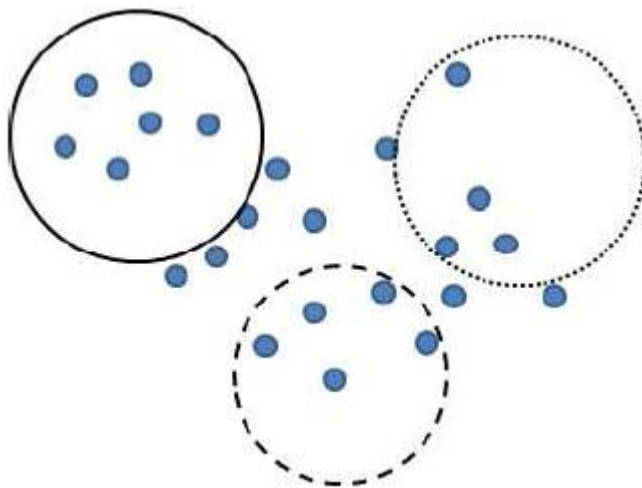


Klasyfikatory w puli

- Jeżeli klasyfikatory niezależnie uzyskują **niski błąd klasyfikacji**, ale **nie są zróżnicowane** to nie wnoszą nic do komitetu.



- Jeśli klasyfikatory **są zróżnicowane** względem siebie, ale **nie przejawiają dobrej jakości rozpoznawania** to pogorszymy jakość komitetu.

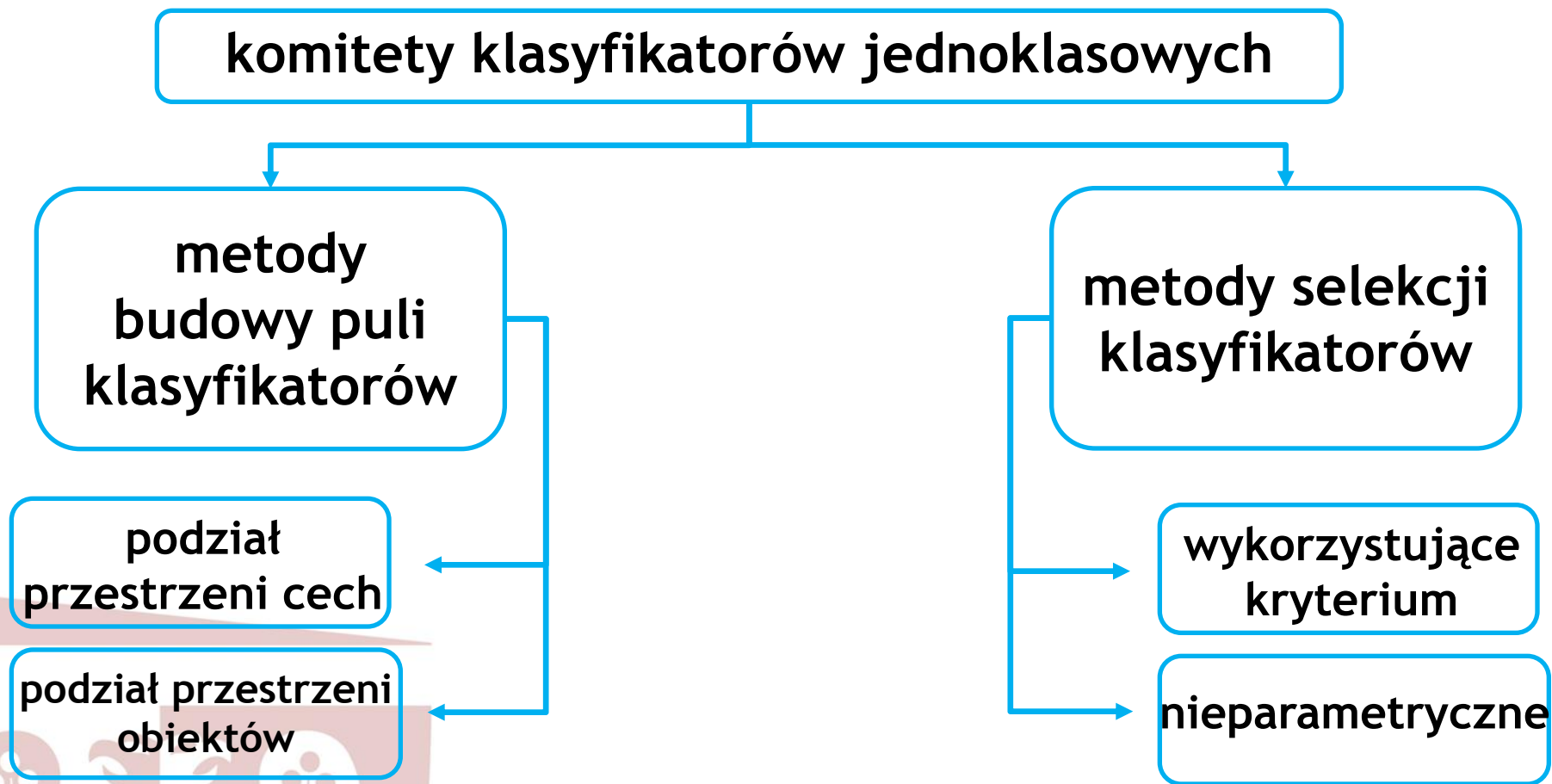


Trudności w budowie komitetów jednoklasowych

- W literaturze istnieje bogaty wachlarz sposobów w jaki budować i dobierać komitety klasyfikatorów - **dedykowany problemom wieloklasowym**.
- Niestety te metody **nie mogą zostać zastosowane** w problemie klasyfikacji jednoklasowej.
- Głównym problemem jest **brak kontrprzykładów** na etapie nauki:
 - **metody budowy komitetów**, takie jak boosting opierają się na minimalizacji błędu rozpoznawania
 - **miary różnorodności** bazują na wskazaniach dwóch i więcej klasyfikatorów dla danego obiektu i jego klasy
 - **metody selekcji / przycinania** komitetów potrzebują miary kryterialnej, którą najczęściej jest błąd komitetu



Dwa główne zagadnienia w budowie komitetów



Budowa komitetu przez podział przestrzeni cech

- Budowa komitetów klasyfikatorów jednoklasowych za pomocą podziału przestrzeni cech polega na **uczeniu klasyfikatorów na zredukowanej przestrzeni cech**, co zapewni wstępną różnorodność modeli.
- Jest to metoda bardzo obiecująca dla klasyfikacji jednoklasowej, gdyż **redukuje złożoność obliczeniową każdego klasyfikatora**.
- Niestety istnieje ryzyko utworzenia **wielu zbliżonych do siebie podprzestrzeni**, lub podprzestrzeni **wykorzystujących cechy o niskiej wartości informacyjnej**.
- Z tego powodu zaproponowano podejście **Overproduce and Select**, które opiera się na zbudowaniu dużej liczby podprzestrzeni a następnie przycinaniu komitetu.

Krawczyk B., Woźniak M., Diversity measures for one-class classifier ensembles. *Neurocomputing*, 2013

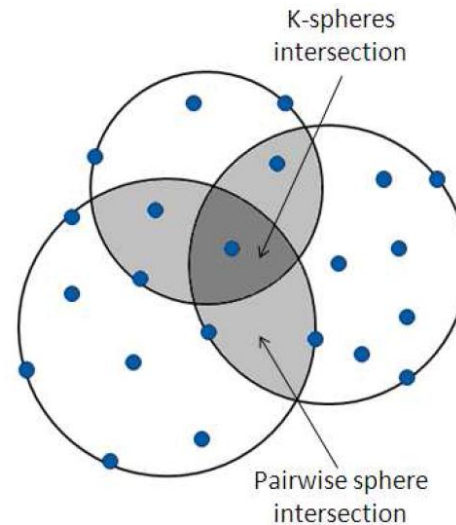


Metody selekcji klasyfikatorów jednoklasowych

- W wielu przypadkach (np. zastosowania metody Random Subspace) **nie należy wykorzystać wszystkich klasyfikatorów** dostępnych w puli.
- Istnieją **dwie główne motywacje** dokonania selekcji klasyfikatorów / przycinania komitetu:
 - **eliminacja słabych klasyfikatorów**, które mogą wpłynąć negatywnie na wynik decyzji komitetu
 - **redukcja ilości klasyfikatorów w puli**, co prowadzi do redukcji złożoności komitetu
- Zaproponowano dwa typy podejść: **wymagające miary kryterialnej** oraz **nieparametryczne**.
- Typ pierwszy jest modyfikacją istniejących metod selekcji, z reguły opierający się na **metodach optymalizacji**. Wymagana jednak jest **dedykowana miara** dla klasyfikacji jednoklasowej.
- Typ drugi jest alternatywnym podejściem, **nie wymagającym** dedykowanej miary.

Miary w klasyfikacji jednoklasowej

- Zaproponowano szereg **nowych miar różnorodności**, wyspecjalizowanych pod kątem klasyfikacji jednoklasowej.
- Jedną z najskuteczniejszych miar jest **miara nakładania się sfer**.



- Jako miarę jakości zamiast tradycyjnego błędu klasyfikacji proponuje się zastosowanie **miary spójności przestrzeni**.



Metody selekcji z kryterium

- Aby uniknąć problemu doboru **tylko dokładnych albo tylko zróżnicowanych klasyfikatorów** jednoklasowych, zaproponowano wykorzystanie metod **optymalizacji wielokryterialnej**.
- Funkcją celu jest liniową kombinacją jednej z miar różnorodności (np. miary nakładania się sfer) oraz miary spójności przestrzeni.
- Wcześniejsze badania przeprowadzono z użyciem **algorytmów genetycznych**.
- Dobre wyniki zachęciły do dalszych badań nad metodą przeszukiwania. Obecnie proponuje się stosowanie **algorytmu memetycznego**, łączącego algorytm genetyczny z lokalnym przeszukiwaniem za pomocą **metody tabu search**.

Szczurek A., Krawczyk B. & Maciejewska M., VOCs classification based on committee of classifiers coupled with single sensor signals. Chemometrics and Intelligent Laboratory Systems vol. 125: 1-10.



Metody selekcji z kryterium

- Metody oparte na podejściu optymalizacyjnym mogą być **złożone obliczeniowo** dla rozbudowanych komitetów.
- Alternatywą może być wykorzystanie **algorytmów klasteryzacji**.
- Proponuje się, aby działały one na macierzy złożonej z **wartości parowych miar różnorodności**.

$$\begin{bmatrix} DIV(\Psi_1, \Psi_1) & \cdots & DIV(\Psi_1, \Psi_j) \\ \vdots & \ddots & \vdots \\ DIV(\Psi_i, \Psi_1) & \cdots & DIV(\Psi_i, \Psi_j) \end{bmatrix}$$

- W ten sposób miara różnorodności jest **tożsama z odległością** między obiektami w klasycznej analizie skupień.
- Następnie dla każdego z klastrów **wybiera się reprezentanta** za pomocą jednej z kilku badanych metod (najbliższy centroidowi, najdalszy od innych centroidów, uczenie nowego klasyfikatora itd).

Metody nieparametryczne

- Alternatywnym podejściem do wprowadzania nowych miar jest szukanie **metod nieparametrycznych** do przycinania komitetów klasyfikatorów jednoklasowych.
- Nasuwającym się tu rozwiązaniem jest zastosowanie **metod klasteryzacji**.
- Analizy skupień dokonuje się w nowej przestrzeni, gdzie **obiektom jest klasyfikator**, zaś cechami **wartości funkcji wsparć** dla poszczególnych obiektów.

$$\begin{bmatrix} F_1(x_1, \omega_T) & \cdots & F_1(x_i, \omega_T) \\ \vdots & \ddots & \vdots \\ F_L(x_1, \omega_T) & \cdots & F_L(x_i, \omega_T) \end{bmatrix}$$

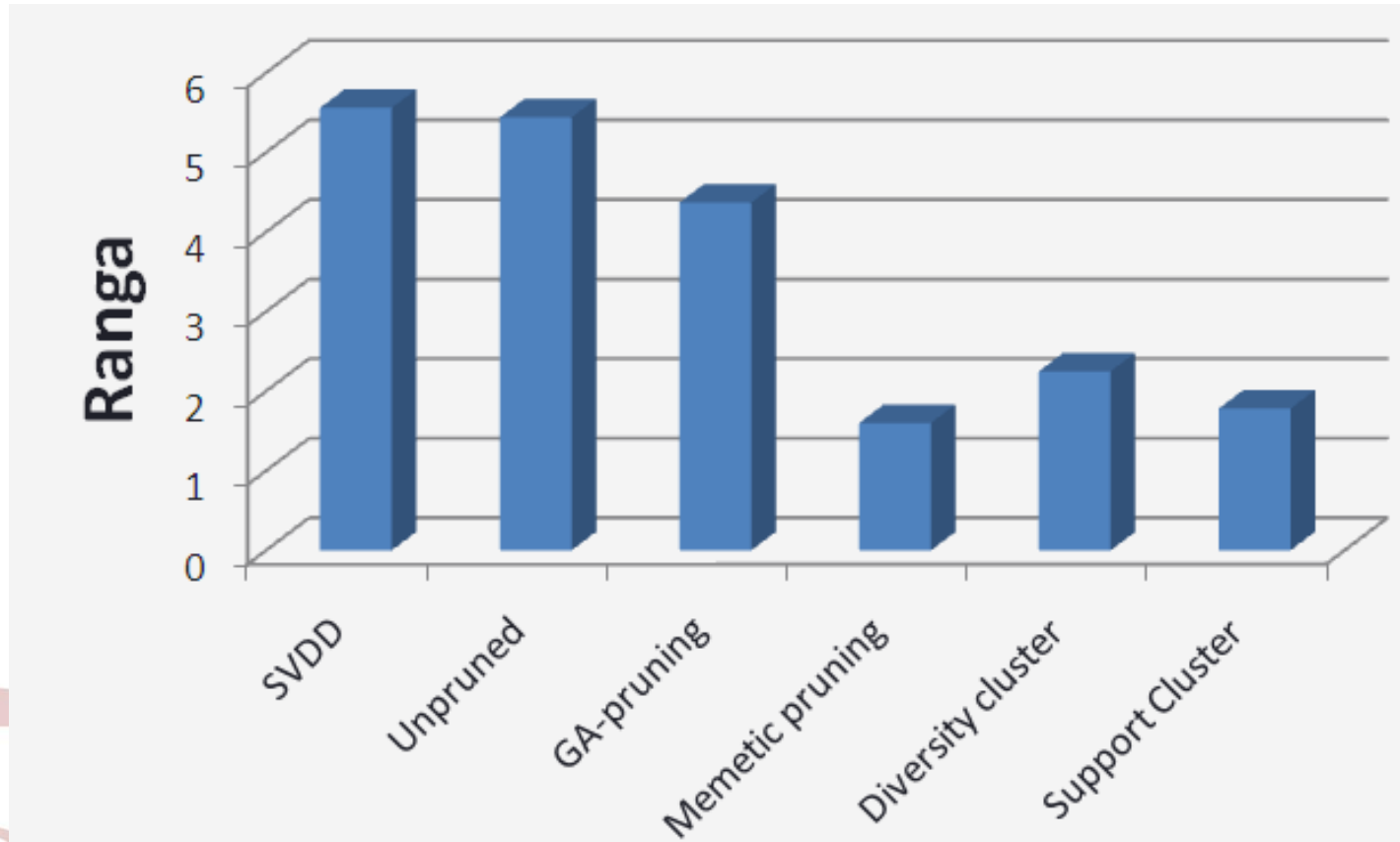
- Wyszukuje się najbardziej podobne klastry klasyfikatorów, tworzące pewne **profile decyzyjne**.

Badania eksperymentalne

- Wykonane na 20 bazach z UCI machine learning repository.
- Klasyfikator bazowy Support Vector Domain Description z jądrem RBF.
- Pula 30 klasyfikatorów, budowana na Random Subspace złożonej z 40% oryginalnych przestrzeni cech.
- Wykorzystano test rankingowy Friedmana.



Badania eksperymentalne



Dziękuję za uwagę

bartosz.krawczyk@pwr.wroc.pl

<http://www.kssk.pwr.wroc.pl/krawczyk>

