

Magdalena Topczewska<sup>1</sup>

## RANGOWA KLASYFIKACJA OBIEKTÓW ZA POMOCĄ KUL W RÓŻNYCH NORMACH

**Streszczenie:** Mając dany zbiór uczący, który zawiera obiekty należące do dwóch lub większej liczby klas, można zbudować najmniejsze kule otaczające obiekty z wybranej klasy rozwiązując zadanie programowania kwadratowego. Ze względu na to, że najmniejsze kule są konstruowane oddzielnie dla każdej klasy, problem może być w prosty sposób rozszerzony do przypadków wieloklasowych. W pracy przedstawiamy propozycje klasyfikatorów opartych na kulach w normie euklidesowej oraz w normie  $l_1$ . Przedstawione eksperymenty zostały przeprowadzone tak na syntetycznych jak i rzeczywistych zbiorach danych.

**Słowa kluczowe:** klasyfikacja, kule, metody nieliniowe

### 1. Wstęp

Analiza dyskryminacyjna jest jedną z metod eksploracyjnej analizy danych, która służy budowaniu obszarów decyzyjnych pozwalających przydzielać obiekty do odpowiednich klas. W przypadku, gdy nie są znane warunkowe gęstości prawdopodobieństwa w klasach możliwe jest ich szacowanie za pomocą metod nieparametrycznych lub założenie dotyczące ogólnej klasy modelu klasyfikatora. Wybór klasyfikatora wiąże się z decyzją o stopniu złożoności linii decyzyjnej.

Popularne i powszechnie stosowane są klasyfikatory w postaci liniowych funkcji dyskryminacyjnych, dla których znane są uogólnienia za pomocą różnych metod [2,8]. Można również założyć inne kształty funkcji decyzyjnych. W literaturze znane są przykłady zastosowania kul, sześcianów, wielościanów wypukłych lub innych kształtów do podziału przestrzeni.

Pierwszy raz dyskryminacja za pomocą kul zastosował do klasyfikacji obrazów w 1962 roku P.W. Cooper. W kolejnych latach była rozwijana przez wielu naukowców [7]. Znanym algorytmem opartym na kulach jest sieć *RCE* (ang. *Restricted Coulomb Energy*) zaproponowana przez D.L. Reilly, L.N. Coopera i C. Elbauma [4,5]. Każda klasa reprezentowana jest jako zbiór prototypowych obszarów, które

---

<sup>1</sup> Wydział Informatyki, Politechnika Białostocka, Białystok

mogą być przedstawione w postaci kul. Pojedyncze kule dobierane są w ten sposób, aby zawierały obiekty należące tylko do jednej klasy i nieobjęte przez inne kule. Po zakończeniu budowania obszarów decyzyjnych w postaci zbioru kul nowy obiekt jest klasyfikowany zgodnie z etykietą klasy, jaką miały obiekty zawarte w odpowiedniej kuli.

Innym znanym algorytmem opartym na klasyfikacji za pomocą kul jest *SCM* (ang. *Set Covering Machine*), zaproponowany przez M. Marchanda oraz J. Shawe-Taylor. Głównym celem jest znalezienie najmniejszej liczby kul, które idealnie podzielią przestrzeń na obszary odpowiadające odpowiednim klasom. Klasyfikator wynikowy jest koniunkcją lub dysjunkcją zbioru klasyfikatorów w postaci kul, a każda z nich dzieli przestrzeń wejściową na obszary przynależności do dwóch klas [3].

W cytowanych pracach klasyfikacja za pomocą kul opiera się na takim podziale przestrzeni, by poszczególne kule zawierały jedynie obiekty należące do danej klasy. Jeśli dane z obu klas bardzo mocno zachodzą na siebie, to liczba kul potrzebnych do idealnego podzielenia przestrzeni na obszary decyzyjne może być bardzo duża.

W niniejszej pracy prezentowane są nowe funkcje dyskryminacyjne, które pozwalają na to, by pewna liczba obiektów z jednej klasy znalazła się w obszarze decyzyjnym innej klasy. Funkcje te mają postać kul w różnych normach, więc kształty obszarów decyzyjnych można lepiej dobrać do analizowanych danych.

Najmniejsza kula otaczająca zbiór obiektów z  $k$ -tej klasy, scharakteryzowana przez środek  $s$  oraz promień  $R$ , może być wyznaczona poprzez rozwiązanie zadania programowania kwadratowego [6]. Modyfikując podejście można uzyskać funkcje celu służącą klasyfikacji obiektów.

W pracy proponujemy klasyfikację punktów należących do kilku klas przy użyciu kryterium nieliniowego opartego na aproksymacji funkcji *signum*, obliczającej znak argumentu, funkcją gładką o własnościach opisanych w dalszej części pracy. Pojedynczy klasyfikator stanowi rozsądny kompromis między ilością prawidłowo i nieprawidłowo sklasyfikowanych przypadków, dopuszczając, aby w obrębie kuli znalazły się także obiekty z innych klas. Takie podejście pozwala zachować jedną z najważniejszych cech klasyfikatorów – *generalizację* przy zachowaniu wysokiej *jakości klasyfikacji*. Wpływa to również na efektywność algorytmu, gdyż zdecydowanie redukuje liczbę kul potrzebnych do podziału przestrzeni wejściowej.

Dodatkowo jako inny typ klasyfikatora proponowane jest zastosowanie kul w normie  $l_1$ , dla których optymalne położenie znajdowane jest także przy użyciu kryterium nieliniowego. Zarówno dla kul w normie  $l_1$  jak i  $l_2$ , pojedyncze klasyfikatory składane są w warstwy rangowe dzieląc przestrzeń na sekwencję aktywnych pól, pozwalającą na wyjściu uzyskać informację, do której klasy należy przyporządkować dany obiekt.

Struktura pracy jest następująca: w rozdziale drugim opisany jest sposób znajdowania najmniejszej kuli otaczającej obiekty, które należą do jednej klasy. Rozdział trzeci rozszerza ideę najmniejszej kuli otaczającej dane do postaci klasyfikatora, służącego do podziału przestrzeni na obszary decyzyjne dwóch klas. Zaprezentowany jest przykład klasyfikacji oraz wizualizacji przedstawionej funkcji celu. W rozdziale czwartym przedstawiona jest idea klasyfikacji rangowej, czyli strategii wykorzystywanej w sytuacjach, gdy dla prawidłowej klasyfikacji potrzebny jest wielokrotny podział przestrzeni za pomocą wybranego klasyfikatora. Rozdział piąty zawiera opis nowego klasyfikatora mającego postać kuli w normie  $l_1$ , a także przykład zastosowania tego typu klasyfikacji. Ostatni, szósty rozdział jest podsumowaniem oraz wytyczeniem dalszych kierunków badań.

## 2. Najmniejsze kule

Dany jest zbiór obiektów zgromadzonych w zbiorze uczącym składającym się z par  $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)\}$ , gdzie  $\mathbf{x}_i$  ( $i = 1, \dots, m$ ) jest wektorem zawierającym wartości atrybutów  $i$ -tego obiektu,  $y_i$  zawiera etykietę klasy, do której należy obiekt  $\mathbf{x}_i$ . Atrybuty mogą przyjmować wartości ze zbioru liczb rzeczywistych, mogą to być również cechy mające wartości binarne ( $\mathbf{x}_i \in R^n$ ,  $\mathbf{x}_i \in \{0, 1\}^n$ ). W przypadku dwóch klas dla etykiety preferowana w pracy jest reprezentacja bipolarna ( $y_i \in \{-1, +1\}$ ), w przypadku wielu klas mogą to być kolejne wartości ze zbioru liczb naturalnych.

Niech  $k$ -ta klasa będzie oznaczona jako  $C_k = \{\mathbf{x}_i : i \in I_k\}$ , gdzie  $I_k$  jest zbiorem indeksów obiektów należących do tej klasy. Najmniejsza kula otaczająca ten zbiór obiektów, scharakteryzowana przez środek  $\mathbf{s}$  oraz promień  $R$ , może być wyznaczona poprzez rozwiązanie zadania programowania kwadratowego

$$\min_{\mathbf{s}, R} R^2 \quad (1)$$

z ograniczeniami

$$\forall i \quad \|\mathbf{x}_i - \mathbf{s}\|^2 \leq R^2, \quad (2)$$

$$R \geq 0. \quad (3)$$

gdzie  $\|\cdot\|$  oznacza normę euklidesową. Ujmując inaczej  $C_k \subseteq B_{\mathbf{s}, R}$ .

Ze względu na to, że tak sformułowane zadanie jest bardzo wrażliwe na obserwacje odstające, można wprowadzić zmienne rozluźniające, pozwalając, by część obiektów znalazła się poza kulą. Uzyskuje się wówczas sformułowanie zawierające w funkcji celu dodatkowo sumę zmiennych rozluźniających. Poprzez dobór odpowiednich parametrów możemy wpływać na wielkość kuli, a zatem na liczbę obiektów,

które znajdują się poza nią [6]. Ponieważ problem jest wypukły, można zastosować metodę mnożników Lagrange'a i doprowadzić funkcję celu do postaci, w której obliczane są wartości iloczynów skalarnych między wektorami danych, a następnie, w celu znalezienia rozwiązania, zastosować standardowe algorytmy programowania kwadratowego [7].

Należy nadmienić, że rozwiązując (1) z ograniczeniami (2) oraz (3) poszukuje się najmniejszej kuli otaczającej obiekty z wybranej klasy biorąc pod uwagę wyłącznie obiekty z tej klasy. Przypadki należące do innych klas nie są w ogóle rozpatrywane. Jeśli dwie klasy będą się pokrywać może wystąpić sytuacja, że wewnątrz znalezionej kuli znajdują się wszystkie obiekty z obydwu klas. Należy więc tak sformułować zadanie, aby znalezione parametry kuli były tak dobrane, by odpowiednio dużo obiektów z wybranej klasy znalazło się wewnątrz kuli przy jednoczesnym zapewnieniu, że minimalna liczba obiektów z innych klas znajdzie się w jej obrębie. Dlatego w pracy proponowane są modyfikacje (1) pozwalające na dobrą klasyfikację nowych obiektów, dla których informacja, do której klasy powinny zostać przyporządkowane nie jest jeszcze znana.

### 3. Klasyfikacja oparta na kuli

#### 3.1 Funkcja celu

W tej pracy proponujemy klasyfikację punktów należących do dwóch klas przy użyciu kryterium nieliniowego opartego na aproksymacji funkcji *signum*, obliczającej znak argumentu, dogodną funkcją gładką  $\zeta$ . Modyfikując (1) przy ograniczeniach (2) oraz (3), badanie czy pojedynczy obiekt  $\mathbf{x}_i$  znajduje się wewnątrz kuli czy poza nią odbywa się poprzez obliczenie wartości funkcji

$$\phi(\mathbf{x}_i; \mathbf{s}, R) = \zeta(R^2 - (\mathbf{x}_i - \mathbf{s})^2). \quad (4)$$

Funkcja celu przyjmie następującą postać

$$\Phi(\mathbf{x}, \mathbf{y}; \mathbf{s}, R) = R^2 + \frac{1}{4} \sum_{i=1}^m (\phi(\mathbf{x}_i; \mathbf{s}, R) - y_i)^2. \quad (5)$$

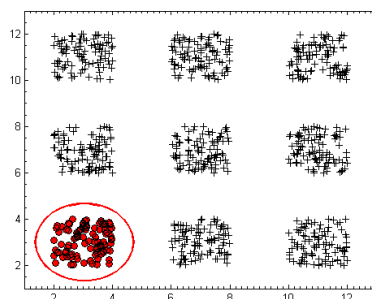
Pierwszy wyraz funkcji celu odpowiada poszukiwaniu najmniejszej kuli wokół zbioru obiektów, drugi natomiast jest regułą minimalizacji błędu kwadratowego przy klasyfikacji obiektów należących do dwóch klas. Jeśli  $i$ -ty obiekt posiadający etykietę 1 będzie usytuowany wewnątrz kuli, to wartość funkcji  $\phi(\mathbf{x}_i; \mathbf{s}, R)$  wyniesie 1, więc wartość różnicy  $(\phi(\mathbf{x}_i; \mathbf{s}, R) - y_i)$  (błędu klasyfikacji) wyniesie 0. Analogicznie,

jeśli obiekt  $\mathbf{x}_i$  o etykiecie  $-1$  będzie usytuowany poza kulą wartość błędu również wyniesie 0. W pozostałych przypadkach wartość różnicy będzie wynosić  $+2$  lub  $-2$ , co po podniesieniu do kwadratu da wartość 4. Liczbę błędnie sklasyfikowanych obiektów, zarówno tych z etykietą  $+1$  leżących poza kulą, jak i obiektów z etykietą  $-1$ , usytuowanych wewnątrz kuli, otrzymamy poprzez podzielenie wartości błędów opisanych powyżej przez wartość 4.

Ze względu na ciągłość i różniczkowalność (5) do znalezienia optymalnych wartości  $s$  oraz  $R$  można stosować metody gradientowe.

### 3.2 Przykład z wizualizacją funkcji celu

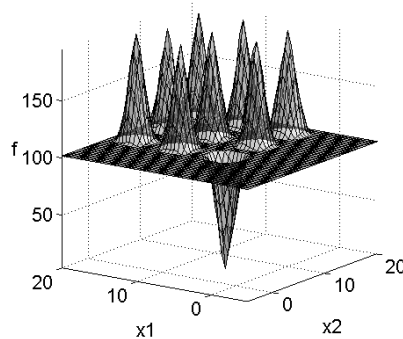
Jako przykład poszukiwania pojedynczej kuli oddzielającej obiekty z wybranej klasy od pozostałych, został wygenerowany zbiór zawierający punkty opisane w  $R^2$ , należące do dziewięciu klas. Każda klasa zawiera 100 obiektów. Obiekty jednej klasy można oddzielić od pozostałych stosując klasyfikator w postaci kuli (rys. 1). Zaważmy, że kule w  $R^2$  to koła.



Rys. 1. Zbior – *szach.txt*

Biorąc pod uwagę przykład z przestrzeni dwuwymiarowej, znalezienie optymalnego podziału w pojedynczym kroku algorytmu wiąże się ze znalezieniem trzech parametrów - dwóch współrzędnych opisujących środek kuli oraz jej promienia. Poniżej znajduje się wykres funkcji celu do klasyfikacji skrajnej klasy w przypadku danych *szach.txt* przy ustalonej (optymalnej) wartości promienia.

Jeśli wartość promienia kuli jest ustalona, poszukiwanie minimum funkcji celu (1) ogranicza się do poszukiwania położenia środka kuli. Na rysunku można za-



**Rys. 2.** Wykres funkcji celu dla zbioru *szach.txt* przy ustalonej (optymalnej) wartości promienia

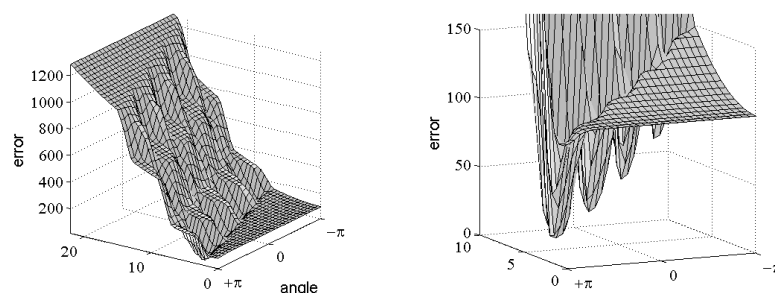
uważyć, że minimum funkcji znajduje się w środku wyróżnionej klasy. Przesuwanie środka kuli powoduje wzrost wartości funkcji celu, czyli gorszą klasyfikację obiektów. Usytuowanie środka kuli w miejscu, gdzie żaden obiekt nie jest wewnątrz kuli, powoduje, że wartość funkcji celu wynosi około 100 (wszystkie obiekty z wyróżnionej klasy są źle sklasyfikowane). Jeśli zaś kula zawiera inną niż wyróżnioną klasę, to wartość funkcji celu wzrośnie do 200 (zliczone są źle sklasyfikowane obiekty z wyróżnionej klasy oraz obiekty z innych klas, które nieprawidłowo znalazły się wewnątrz kuli).

W celu graficznego przedstawienia funkcji celu z wszystkimi parametrami należy posłużyć się opisem w przestrzeni parametrów. Jako pierwszy parametr wzięty jest kąt między wektorem  $s$  i osią  $OX$ , jako drugi – parametr  $R$ .

Zarówno na rysunku 2 jak i 3 funkcja jest gładka i tym samym dogodna do minimalizacji.

#### 4. Klasyfikacja rangowa

Przedstawiona w poprzednim paragrafie funkcja celu dotyczy klasyfikacji obiektów należących do dwóch klas – obiekty z jednej klasy powinny znaleźć się wewnątrz kuli, podczas gdy obiekty z drugiej klasy powinny pozostać na zewnątrz. W przypadku, gdy obiekty należące do wybranej klasy rozmieszone są w przestrzeni danych w kilku rozłącznych klastrach lub gdy w zbiorze danych znajduje się liczba klas większa niż 2, można zastosować wielokrotne podziały zgodnie z wybraną strategią. W pracy proponujemy zastosowanie strategii rangowej oraz wykorzystanie pól aktywnych [1].



**Rys. 3.** Wykres funkcji celu w przestrzeni parametrów (z prawej strony – powiększenie ukazujące minimum globalne funkcji)

Niech będzie dany zbiór elementarnych klasyfikatorów  $Q_k$ , ( $i = 1, \dots, k$ ), każdy zdefiniowany na podstawie obiektów  $\mathbf{x}_i$ , należących do zbioru uczącego, oraz przyjmujący jedną z wartości bipolarnych  $q_k$  ( $q_k \in \{-1, +1\}$ ):

$$q_k = q_k(\mathbf{p}, \mathbf{x}_i) \quad (6)$$

gdzie  $\mathbf{p}$  jest wektorem parametrów, w tej pracy  $\mathbf{p} = [s_k, R_k]$ .

Idea klasyfikacji rangowej polega na wielokrotnym stosowaniu klasyfikacji danych w celu uzyskania podziału całej przestrzeni na obszary odpowiadające odpowiednim klasom.

**Definition 1.** Klasyfikator  $Q_k$  jest nadrzędny w stosunku do klasyfikatora  $Q_r$  wtedy i tylko wtedy, gdy  $k < r$ .

**Definition 2.** Pole aktywne  $P_k$  jest to zbiór takich obiektów  $\mathbf{x}_i$ , które uaktywniają klasyfikator  $Q_k$ , jednocześnie nie uaktywniając żadnego nadrzędnego klasyfikatora.

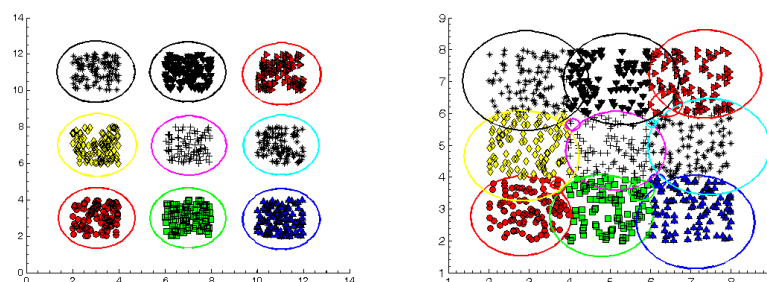
W pierwszym kroku poszukiwana jest kula najlepiej separująca obiekty z wyróżnionej klasy. Wnętrze znalezionej kuli stanowi *pole aktywne*. Zapamiętywana jest dla niego etykieta klasy oraz ranga. Klasyfikator znaleziony w taki sposób stanowi pierwszą warstwę ogólnego rangowego modelu klasyfikacji. Dla warstwy pierwszej ranga jest równa 1. W kolejnym kroku, wśród pozostałych obiektów poszukiwana jest następna kula, która możliwie dobrze oddziela obiekty z wyróżnionej klasy od obiektów z innych klas. Wyróżnioną klasą na tym etapie może być klasa z etapu pierwszego, dla której nie wszystkie obiekty zostały prawidłowo sklasyfikowane lub dowolna inna klasa spośród pozostałych, wybrana według dowolnego kryterium. Algorytm jest powtarzany, a dla znalezionych w kolejnych iteracjach kul zapamiętywana

jest etykieta klasy oraz ranga. Algorytm jest zatrzymany, gdy wszystkie obiekty zostaną podzielone na sekwencję aktywnych pól, pozwalając na wyjściu uzyskać informację, do której klasy należy przyporządkować dany obiekt.

Dzieląc przestrzeń na pola aktywne można posłużyć się ich dwoma rodzajami [1]. Pierwszy rodzaj to pola aktywne, które są deterministycznie dopuszczalne, czyli takie, które zawierają wyłącznie obiekty należące do jednej klasy. Są to pola jednorodne pod względem przynależności obiektów do klas. Drugi typ stanowią pola aktywne, które są statystycznie dopuszczalne. Są to pola, które z pewnym założonym współczynnikiem  $\alpha \in (0; 0.5)$  dopuszczają występowanie wewnątrz nich obiektów należących do innych klas. W pracy rozpatrujemy obydwie rodzaje pól aktywnych.

#### 4.1 Przykład

Rozwijając przykład podany w poprzednim paragrafie prezentujemy zbiory danych zawierające obiekty, które należą do dziewięciu klas ułożonych w postaci pól szachownicy. Pierwsza część wykresu przedstawia zbiór, w którym obiekty z różnych klas są separowalne, natomiast druga część dotyczy zbioru, gdy klasy przylegają do siebie.



Rys. 4. Zbiory *szach.txt* z zaznaczonymi granicami decyzyjnymi

Stosując klasyfikację rangową uzyskujemy podział przestrzeni na obszary przynależności do poszczególnych klas. W pierwszym przypadku każda klasa może być odseparowana poprzez jedno cięcie, czyli znalezienie parametrów dla pojedynczej kuli, wewnątrz której znajdą się wszystkie obiekty z odpowiedniej klasy. W drugim – do odseparowania pojedynczej klasy potrzeba jest kilku cięć. W pojedynczym



kroku algorytmu poszukiwana jest najbardziej liczna klasa, której obiekty mogą być sklasyfikowane prawidłowo.

W przypadku danych separowalnych dziewięć iteracji potrzebowano do znalezienia dziewięciu kul całkowicie separujących obiekty, które należały do dziewięciu klas. Każda kula jest jednorodnym polem aktywnym i zawiera obserwacje należące wyłącznie do jednej klasy. W przypadku danych, w których klasy przylegały do siebie algorytm znalazł 22 kule potrzebne do całkowitej separacji obiektów.

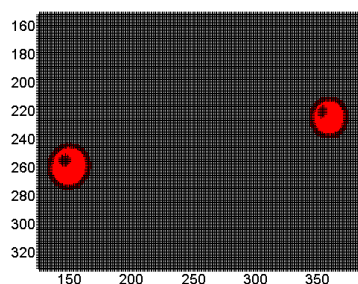
## 4.2 Przykład

Jako przykład drugi został wybrany zbiór zawierający kolorowe zdjęcie, na którym wystąpił efekt "czerwonych oczu". Efekt ten może wystąpić, gdy zdjęcie robione jest przy użyciu lampy błyskowej. Światło lampy odbija się od siatkówki dając zabarwienie czerwone na skutek przefiltrowania światła przez naczynia krwionośne.



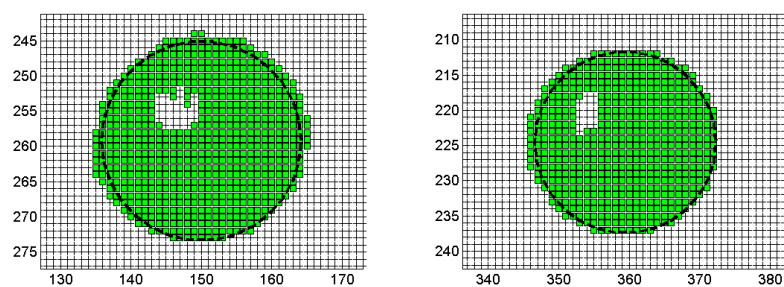
Rys. 5. Obraz oryginalny – Gabi

W tym przypadku jednokrotne zastosowanie klasyfikacji za pomocą kul nie przyniosłoby oczekiwanych rezultatów, gdyż obydwójce oczu powinno się sklasyfikować jako obiekty leżące wewnątrz znalezionych kul. Niezbędna w takim przypadku jest klasyfikacja rangowa. Ze względu na fakt, że oryginalny obraz jest w przestrzeni barw RGB24, poprzez badanie poziomów poszczególnych kanałów została stworzona bipolarna zmienna klasyfikująca, która przyjmuje wartość  $+1$  w przypadku gdy kolor piksela jest czerwony oraz  $-1$  dla pozostałych kolorów. Należy tu zaznaczyć, że nie wszystkie obiekty należące do klastra opisującego pojedyncze oko są w odcieniach koloru czerwonego. Odbłysek światła sprawia, że część



Rys. 6. Wyróżniona klasa – powiększenie

pikseli wewnątrz oka ma kolor biały lub zbliżony do białego (rys. 6). Z tego powodu parametry algorytmu powinny być dobrane tak, aby pozwalały umiejscowić wewnątrz kuli obiekty nienależące do wyróżnionej klasy.



Rys. 7. Znalezione kule otaczające obiekty opisujące lewe i prawe oko

W dwóch krokach algorytmu znaleziono optymalne parametry kul otaczających obiekty z danej klasy (rys. 7). Punkty znajdujące się wewnątrz znalezionych kul mogą być następnie poddane obróbce niwelującej efekt czerwonych oczu.

## 5. Klasyfikacja za pomocą kul w normie $l_1$

### 5.1 Funkcja celu

Drugim rodzajem klasyfikatora opisanym w pracy jest kula w normie  $l_1$ , opisana równaniem

$$\|\mathbf{x} - \mathbf{s}\|_{l_1} \leq R, \quad (7)$$

gdzie  $\mathbf{s}$  jest środkiem kuli,  $R$  jej promieniem a  $\|\cdot\|_{l_1}$  oznacza normę taksówkową.

Analogicznie jak w przypadku normy euklidesowej, najmniejsza kula w normie  $l_1$  otaczająca zbiór obiektów z  $k$ -tej klasy, scharakteryzowana przez środek  $\mathbf{s}$  oraz promień  $R$ , może być wyznaczona poprzez rozwiązanie zadania

$$\min_{\mathbf{s}, R} R^2 \quad (8)$$

z ograniczeniami

$$\forall i \quad \|\mathbf{x}_i - \mathbf{s}\|_{l_1} \leq R \quad (9)$$

$$R \geq 0. \quad (10)$$

W celu znalezienia takiej linii decyzyjnej, by obiekty z etykietą 1 znalazły się wewnątrz kuli, a obiekty z etykietą  $-1$  poza nią, poszukujemy funkcji, która będzie zwracać wartość 1 wewnątrz kuli i wartość  $-1$  poza nią. Posłużymy się do tego funkcją charakterystyczną, która dla obiektów w kuli zwróci wartość jeden, a poza nią wartość zero. Następnie zmodyfikujemy funkcję, by na wyjściu uzyskiwać wartości  $\pm 1$ .

W przypadku jednowymiarowym możemy badać usytuowanie obiektu  $x_i$  poprzez obliczenie następującej funkcji

$$\phi_{1D}(x_i; s, R) = \frac{1}{2}(\text{sgn}(x_i - s + R) + \text{sgn}(s - x_i + R)). \quad (11)$$

Jeśli obiekt  $x$  jest usytuowany poza kula, wartość jednego wyrazu w sumie będzie dodatnia, drugiego ujemna, więc suma wyniesie zero. Jeżeli natomiast  $x$  jest wewnątrz kuli, to wartość sumy wyniesie 2. Mnożąc wynik przez  $\frac{1}{2}$  uzyskujemy funkcję, która zwraca wartość 1, gdy obiekt jest usytuowany wewnątrz kuli oraz wartość 0, gdy jest poza nią.

Jeżeli obiekty należą do przestrzeni wielowymiarowej, można wszystkie funkcje (11) dla każdego wymiaru przemnożyć przez siebie. Dodatkowo jako aproksymacja funkcji  $\text{sgn}$  wprowadzona została odpowiednia funkcja gładka  $\zeta$

$$\phi_{smooth}(\mathbf{x}_i; \mathbf{s}, R) = \prod_{j=1}^n \frac{1}{2}(\zeta(x_{ij} - s_j + R) + \zeta(s_j - x_{ij} + R)). \quad (12)$$

Jeśli obiekt jest umieszczony wewnątrz kuli, to dla każdego wymiaru wartość gładkiej aproksymacji funkcji (11) będzie równa około 1. Przemnożenie przez siebie wszystkich funkcji jednowymiarowych dla każdego wymiaru będzie również wynosić 1. Jeśli natomiast obiekt jest usytuowany poza kulą, wtedy dla wybranych wymiarów funkcja (11) wyniesie 0. Mnożąc wszystkie funkcje jednowymiarowe przez siebie uzyskamy ostatecznie wartość 0.

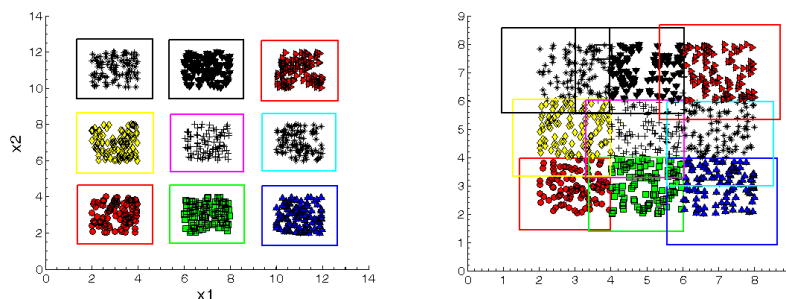
Ze względu na to, że zmienna klasyfikująca w danych ma reprezentację bipolarną, przekształcamy funkcję (12), by na wyjściu otrzymywać wartości 1 dla obiektów znajdujących się wewnątrz kuli oraz  $-1$  dla obiektów rozmieszczonych poza nią

$$\Phi(\mathbf{x}, \mathbf{y}; \mathbf{s}, R) = R_k^2 + \frac{1}{4} \sum_{i=1}^m ((2\phi_{smooth}(\mathbf{x}_i; \mathbf{s}, R) - 1) - y_i)^2. \quad (13)$$

Do funkcji celu dodajemy również człon odpowiedzialny za poszukiwanie jak najmniejszej kuli.

## 5.2 Przykład

Poniżej przedstawiono klasyfikację danych opisanych w *przykładzie 1* przy użyciu kul w normie  $l_1$ .



Rys. 8. Zbiory *szach.txt* z zaznaczonymi granicami decyzyjnymi

W przypadku danych separowalnych w dziewięciu kolejnych krokach uzyskano podział przestrzeni na pola aktywne, z których każde było deterministycznie dopuszczalne, czyli zawierało obiekty należące do jednej klasy. Także dane, w których klasy przylegały do siebie, zostały całkowicie odseparowane w dziewięciu cięciach.

## 6. Podsumowanie

W pracy zaprezentowano klasyfikatory oparte na kulach w normie euklidesowej oraz w normie  $l_1$ . Przedstawiono dwie funkcje celu, do minimalizacji których można zastosować metody gradientowe. Niewątpliwą trudność stanowić może zależność metod nieliniowych od wartości wektora początkowego. Można je wyznaczyć stosując metodę  $k$ -średnich lub  $k$ -medoidów, obliczając oraz zapamiętując wartości średnie lub mediany dla znalezionych klastrów, czyli skupisk obiektów.

Obecnie badany jest przypadek modyfikacji przedstawionych funkcji celu, aby umożliwić niesymetryczne karanie obiektów należących do różnych klas. Eksperymenty będą przeprowadzone na dużych rzeczywistych zbiorach danych.

## Literatura

- [1] Bobrowski, L.: Ranked Patterns and Structural Linearization of Data Sets, In Poster Proceedings: Petra Perner (Ed.), Industrial Conference on Data Mining, ICDM 2006, IBAI CD-Report, ISSN 1617-2671, July 2006, pp. 30-36.
- [2] Duda, O.R., Hart, P.E., Stork, D.G.: Pattern Classification, Wydanie drugie, zmienione, John Wiley & Sons, 2001.
- [3] M. Marchand, M., Shawe-Taylor, J.: The set covering machine, Journal of Machine Research, 3, pp. 723-746, 2002.
- [4] Reilly, D.L., Cooper, L.N., Elbaum, C.: A neural model for category learning, Biological Cybernetics, 45, pp. 35-41, 1982.
- [5] Scofield, C.L., Reilly, D.L., Elbaum, C., Cooper, L.N.: Pattern class degeneracy in an unrestricted storage density memory. In Anderson, D.Z., ed.: Neural Information Processing Systems. American Institute of Physics, Denver, CO, pp. 674-682, 1987.
- [6] Tax, D.M.J., Duin, R.P.W.: Support vector domain description, Pattern Recognition Letters, vol. 20, no. 11-13, pp. 1191-1199, 1999.
- [7] Wang, J., Neskovic, P., Cooper, L. N.: Pattern classification via single spheres. Lecture Notes in Computer Science: Discovery Science (DS), A. Hoffmann, H. Motoda, and T. Scheffer (Eds.), Springer-Verlag, Vol. 3735. pp. 241-252, 2005.
- [8] Johnson, R. A., Wichern, D. W.: Applied Multivariate Statistical Analysis, Prentice-Hall, Inc., Englewood Cliffs, New York, 1991.

## **RANKED CLASSIFICATION OF DATA USING BOUNDING SPHERES IN DIFFERENT NORMS**

**Abstract:** If a training set containing objects from two or more classes is given, minimum bounding spheres enclosing objects belonging to a marked class can be built by solving a quadratic programming task. Because the minimum spheres are constructed separately for each class the problem can be easily extended to the multi-class cases.

In the paper classifiers both in  $l_1$  and  $l_2$  norms are proposed. Experiments were performed on artificial and on real data sets.

**Keywords:** classification, spheres, nonlinear methods

Praca naukowa finansowana ze środków na naukę w latach 2007-2008 jako projekt badawczy.