Zaawansowane uczenie maszynowe:
wykład 13

Paweł Cichosz

Nienadzorowana detekcja anomalii

Zbiór trenujący: $T \subseteq D \subset X$

Prawdziwy status: $c : X \to \{0, 1\}$, dla $x \in T$ **nieznane** $c(x)$

Założenie: przykłady trenujące są w zdecydowanej większości lub wyłącznie prawidłowe.

Podejście: detekcja przykładów *odstających* – nietypowych, niepodobnych do (większości) przykładów trenujących:

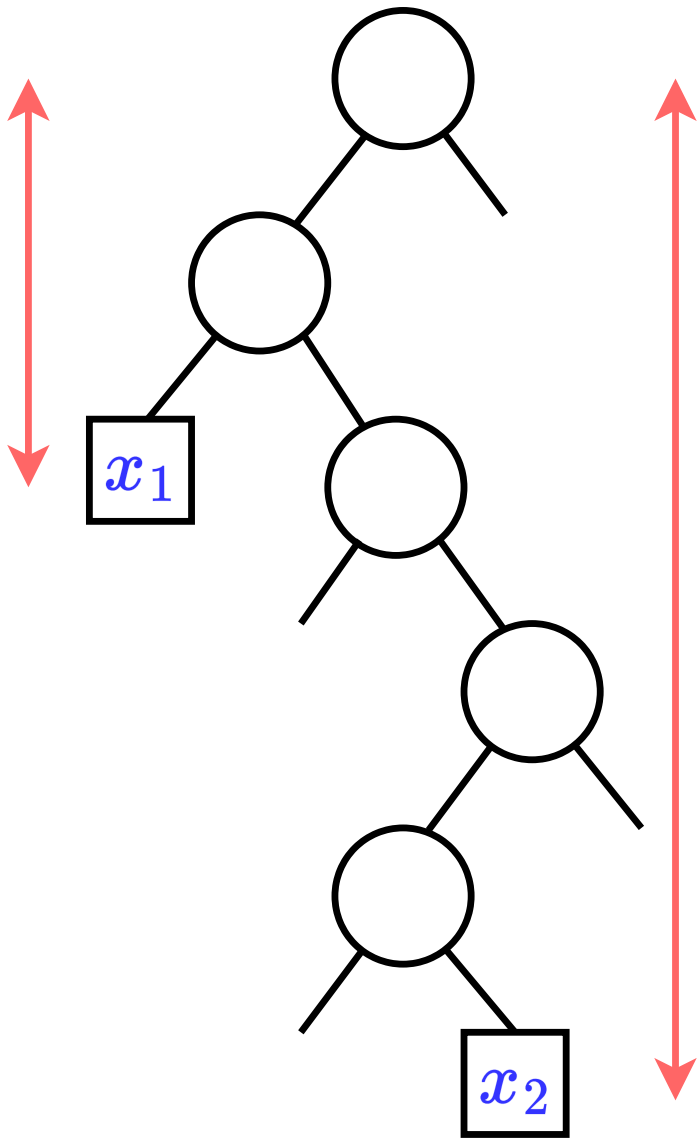
- klasyfikacja jednoklasowa,
- niepodobieństwo do sąsiadów,
- niepodobieństwo do grup.

Wskaźnik nieprawidłowości: $\psi(x)$ – liczba wyrażająca „przekonanie” modelu o nieprawidłowości przykładu x.

Klasyfikacja jednoklasowa

- Najczęściej modyfikacja standardowych algorytmów klasyfikacji binarnej.
- Sposób modyfikacji specyficzny dla poszczególnych algorytmów.
- Możliwe także podejście ogólne: sztuczne generowanie przykładów odstających i zastosowanie klasyfikacji binarnej.

Las izolacyjny



Model zespołowy: zespół m drzew izolacyjnych, np. $m = 100$.

Zbiory trenujące dla poszczególnych drzew: k-elementowe próby $T_1, T_2, \ldots$ z T (losowane bez zwracania, stosunkowo małe k – kilkaset-kilka tysięcy, zwykle 2^l dla $l \in \{8, 9, 10, \ldots\}$).

Wybór podziałów: jednostajnie losowy.

Kryterium stopu: brak możliwości podziału lub głębokość przekracza $\log_2 k$.

Las izolacyjny

Predykcja:

$$\psi(x) = 2^{-\frac{L(x)}{L_{\mathrm{BST}}(k)}}$$

gdzie $L(x)$ – średnia długość ścieżki (liczba gałęzi) do liścia, do którego trafia w poszczególnych drzewach przykład x, powiększona o liczbę dodatkowych przykładów w tym liściu, $L_{\mathrm{BST}}(k)$ – oczekiwana długość ścieżki nieudanego wyszukiwania w drzewie BST o k węzłach:

$$L_{\mathrm{BST}}(k) = 2H(k-1) - 2\frac{k-1}{k}$$

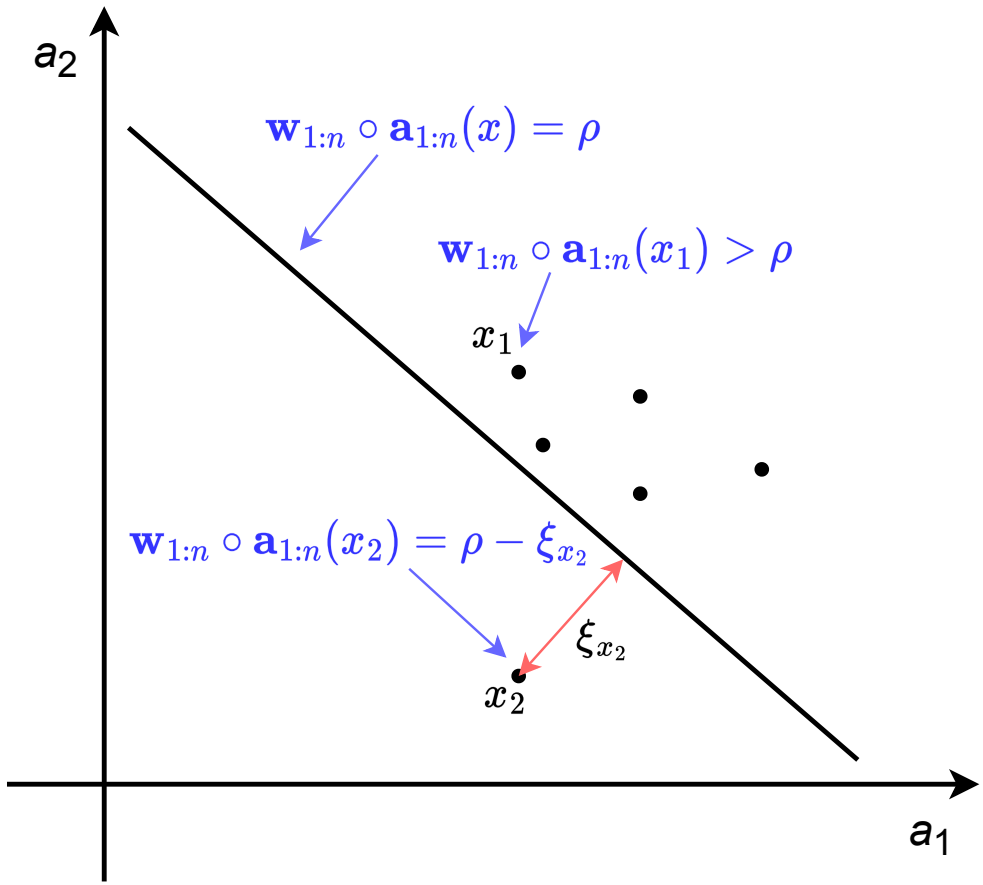
przy czym $H(i)$ – i-ta liczba harmoniczna, $H(i) \approx \ln i + \gamma$, γ – stała Eulera-Mascheroniego.

Domyślny próg decyzyjny: 0.5.

Jednoklasowy SVM (OC-SVM)

Granica decyzyjna:

hiperpłaszczyzna maksymalnie odległa od początku układu współrzędnych separująca od niego większość przykładów trenujących.



Jednoklasowy SVM (OC-SVM)

Zadanie optymalizacji:

minimalizacja:

$$\frac{1}{2}\|\mathbf{w}_{1:n}\|^2 + \frac{1}{\nu|T|}\sum_x \xi_x - \varrho$$

przy ograniczeniach:

$$(\forall x \in T) \quad \mathbf{w}_{1:n} \circ \mathbf{a}_{1:n}(x) \geq \varrho - \xi_x$$
$$(\forall x \in T) \quad \xi_x \geq 0$$

gdzie ν – maksymalny zakładany udział przykładów odstających w zbiorze trenującym.

Jednoklasowy SVM (OC-SVM)

Predykcja:

- jako nieprawidłowe klasyfikowane przykłady leżące po ujemnej stronie granicy decyzyjnej, dla których:

$$\mathbf{w}_{1:n} \circ \mathbf{a}_{1:n}(x) < \varrho$$

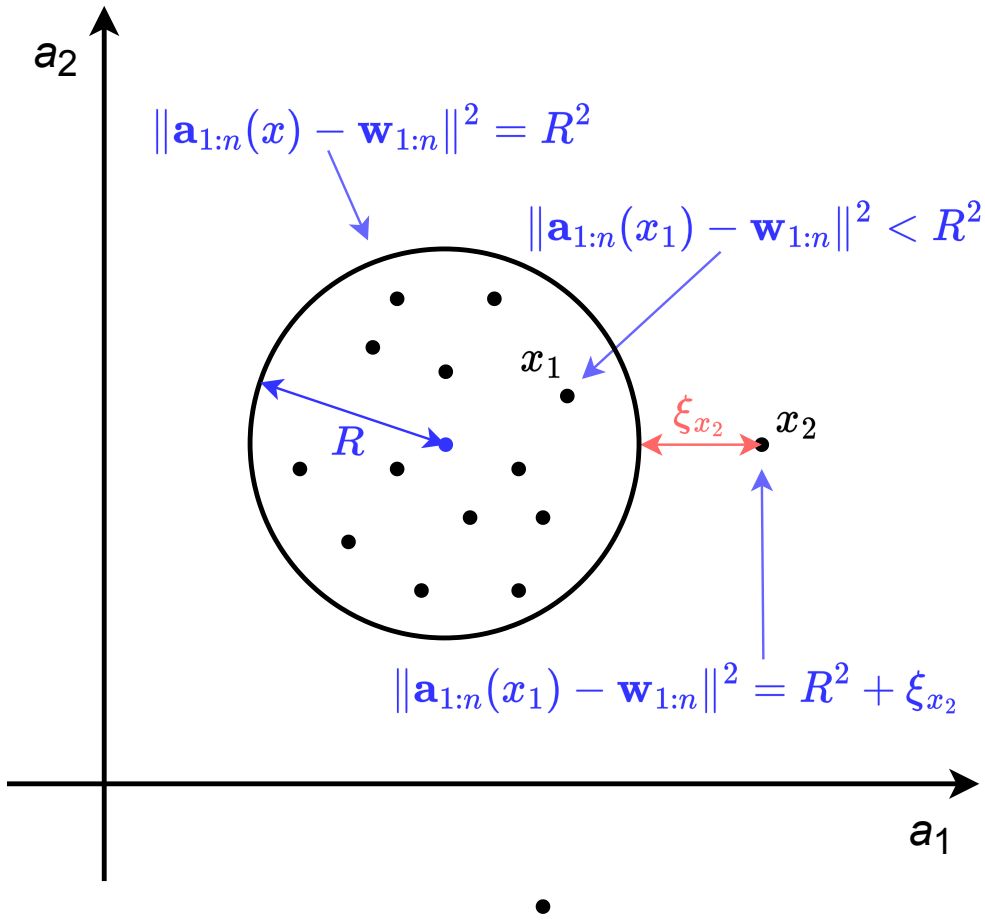
- $\psi(x)$ proporcjonalne do odległości od granicy decyzyjnej:

$$\psi(x) = -(\mathbf{w}_{1:n} \circ \mathbf{a}_{1:n}(x) - \varrho)$$

Postać dualna: uzyskiwana metodą mnożników Lagrange'a analogicznie jak dla SVM (szczegóły poza zakresem wykładu).

Support Vector Data Description (SVDD)

Granica decyzyjna: sfera o minimalnym promieniu obejmująca większość przykładów trenujących.



Support Vector Data Description (SVDD)

Zadanie optymalizacji:

minimalizacja:

$$R^2 + \mathcal{C} \sum_{x \in T} \xi_x$$

przy ograniczeniach:

$$(\forall x \in T) \quad \|\mathbf{a}_{1:n}(x) - \mathbf{w}_{1:n}\|^2 \leq R^2 + \xi_x$$
$$(\forall x \in T) \quad \xi_x \geq 0$$

Support Vector Data Description (SVDD)

Predykcja:

- jako nieprawidłowe klasyfikowane przykłady leżące na zewnątrz sfery reprezentującej granicę decyzyjną, dla których:

$$\|\mathbf{a}_{1:n}(x) - \mathbf{w}_{1:n}\|^2 > R^2$$

- $\psi(x)$ wyznaczane na podstawie odległości od granicy decyzyjnej:

$$\psi(x) = \delta_{\mathbf{w}_{1:n},R}(x) = \|\mathbf{a}_{1:n}(x) - \mathbf{w}_{1:n}\| - R$$

Postać dualna: uzyskiwana metodą mnożników Lagrange'a analogicznie jak dla SVM (szczegóły poza zakresem wykładu).

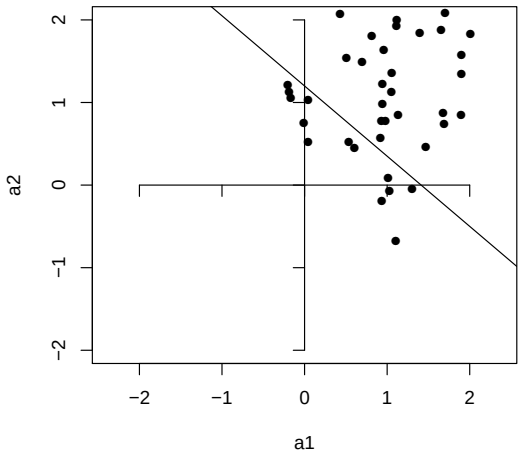
Funkcje jądrowe

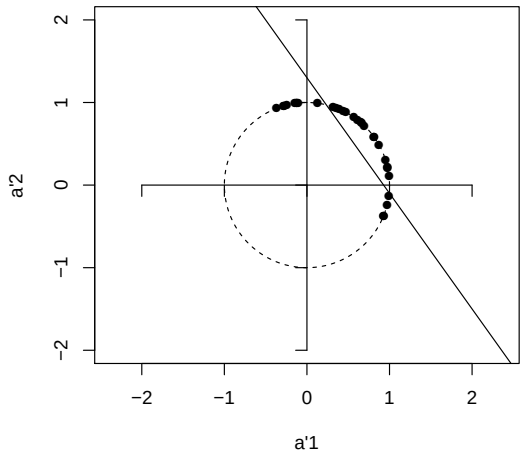
- Algorytmy OC-SVM i SVDD zazwyczaj stosowane z jądrem radialnym (niezmienniczym względem translacji):

$$K(\mathbf{a}_{1:n}(x_1), \mathbf{a}_{1:n}(x_2)) = K(\mathbf{a}_{1:n}(x_1) + \mathbf{b}, \mathbf{a}_{1:n}(x_2) + \mathbf{b})$$
$$K(\mathbf{a}_{1:n}(x), \mathbf{a}_{1:n}(x) + \mathbf{b}) = K(\mathbf{0}, \mathbf{b}) = \kappa(\mathbf{b})$$
$$K(\mathbf{a}_{1:n}(x), \mathbf{a}_{1:n}(x)) = \text{const}$$

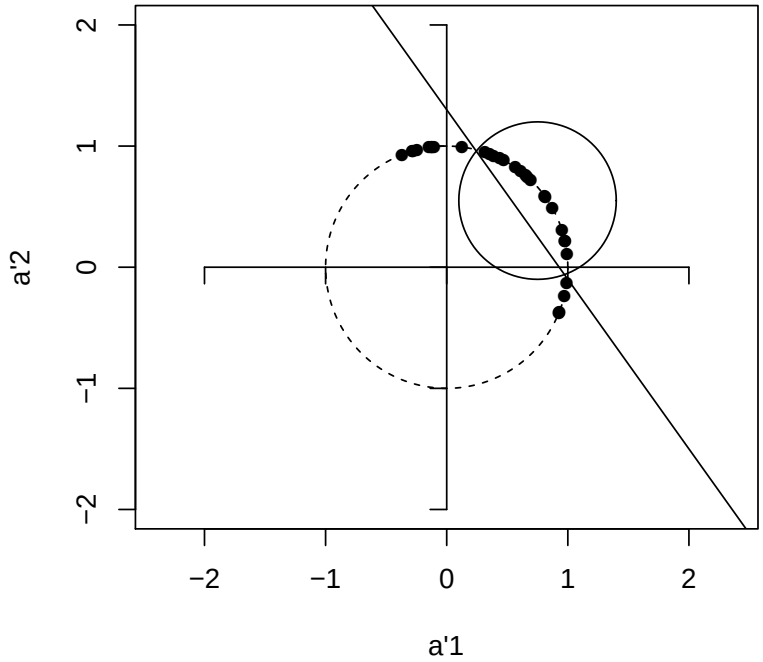
- W nowej reprezentacji wyznaczanej przez funkcję jądrową przykłady leżą na pewnej sferze.
- Wówczas OC-SVM i SVDD stają się w zasadzie równoważne.

Funkcje jądrowe





- Granica decyzyjna OC-SVM odcina czaszę sfery, na której leżą przykłady.
- Sfera wyznaczana przez SVDD może obejmować tę odciętą czaszę.



Niepodobieństwo do sąsiadów

Podstawowa koncepcja: jako nieprawidłowe identyfikowane przykłady najmniej podobne do przykładów trenujących.

- Zwykle stosowane do detekcji przykładów w zbiorze danych odstających od większości tego zbioru, ale możliwe również wykorzystanie do tworzenia modelu stosowanego do nowych danych.

Identyfikacja najbliższych sąsiadów: $\mathrm{NN}_{T,k}(x)$ – zbiór k przykładów trenujących najbardziej podobnych do przykładu x (może być ich więcej jeśli są przykłady jednakowo podobne do x jak k-ty sąsiad).

Globalny wskaźnik nieprawidłowości

- Niepodobieństwo do k-tego (najmniej podobnego) sąsiada przykładu x:

$$\psi(x) = \delta_{T,k}(x) = \max_{x' \in \mathrm{NN}_{T,k}(x)} \delta(x, x')$$

- Średnie niepodobieństwo do k sąsiadów:

$$\psi(x) = \delta_{T,1:k}(x) = \frac{1}{|\mathrm{NN}_{T,k}(x)|} \sum_{x' \in \mathrm{NN}_{T,k}(x)} \delta(x, x')$$

- Odwrotność globalnej gęstości:

$$1/|\{x' \in T \mid \delta(x, x') < \Delta\}|$$

- Wartości globalnego wskaźnika nieprawidłowości nie uwzględniają zróżnicowania gęstości przykładów prawidłowych.

Lokalny wskaźnik nieprawidłowości

- Osiągalność x_1 z x_2 (skorygowana miara niepodobieństwa):

$$\delta_{T,k}(x_1, x_2) = \max\{\delta(x_1, x_2), \delta_{T,k}(x_2)\}$$

- Lokalna gęstość otoczenia x:

$$\varrho_{T,k}(x) = \frac{1}{\frac{1}{|\mathrm{NN}_{T,k}(x)|} \sum_{x' \in \mathrm{NN}_{T,k}(x)} \delta_{T,k}(x, x')}$$

- *Local outlier factor* (LOF):

$$\psi(x) = \mathrm{LOF}_{T,k}(x) = \frac{1}{|\mathrm{NN}_{T,k}(x)|} \sum_{x' \in \mathrm{NN}_{T,k}(x)} \varrho_{T,k}(x') / \varrho_{T,k}(x)$$

- Wartości lokalnego wskaźnika nieprawidłowości uwzględniają zróżnicowanie gęstości przykładów prawidłowych.

Niepodobieństwo do grup

Podstawowa koncepcja: za nieprawidłowe mogą być uważane przykłady niepodobne do grup wyznaczonych na zbiorze trenującym (lub podobne do małych/izolowanych grup).

Wskaźnik nieprawidłowości: różne możliwe warianty normalizacji i/lub ważenia niepodobieństwa do środka najbliższej grupy $d_x = \arg\min_d \delta(x, d)$.

- Iloraz niepodobieństwa do najbliższego środka grupy i średniego niepodobieństwa elementów tej grupy do jej środka:

$$\psi(x) = \frac{\delta(x, d_x)}{\frac{1}{|T_{d_x}|} \sum_{x' \in T_{d_x}} \delta(x', d_x)}$$

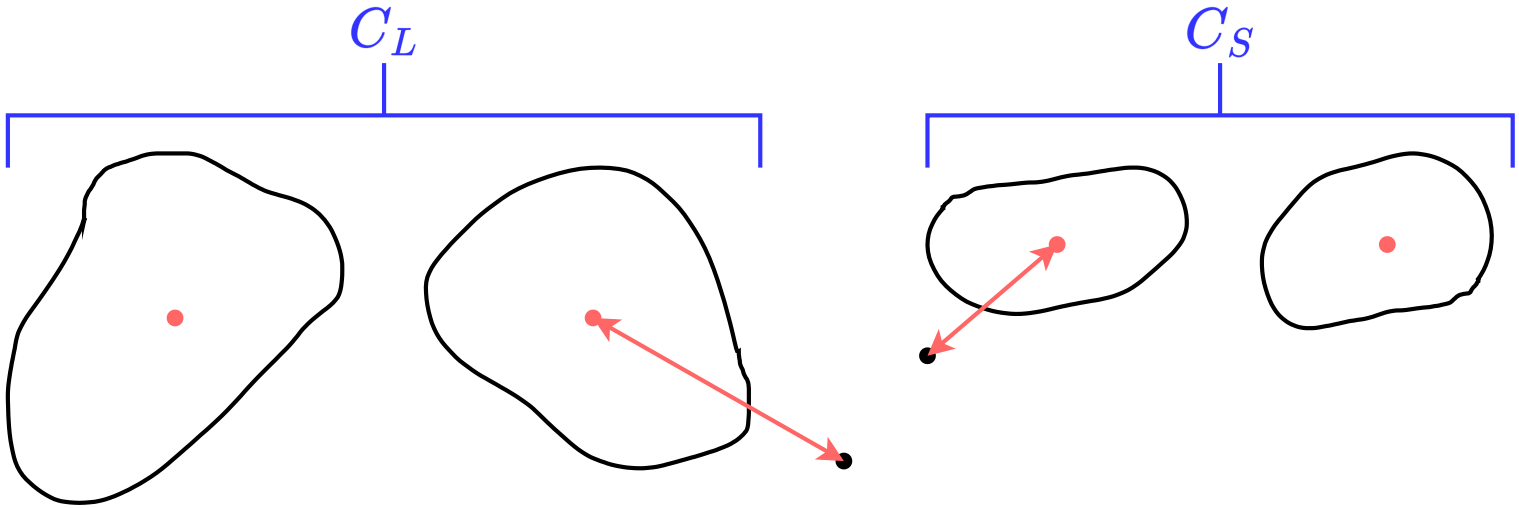
Wskaźniki nieprawidłowości oparte na grupowaniu

- *Cluster-based local outlier factor* (CBLOF – wariant standardowy, uCBLOF – uproszczony wariant bez ważenia rozmiarem grup):

$$\psi(x) = \mathrm{CBLOF}_\delta(x) = \begin{cases} |T_{d_x}|\delta(x, d_x) & \text{jeśli } d_x \in C_L \\ |T_{d_x}|\delta(x, d'_x) & \text{jeśli } d_x \in C_S \end{cases}$$

$$\psi(x) = \mathrm{uCBLOF}_\delta(x) = \begin{cases} \delta(x, d_x) & \text{jeśli } d_x \in C_L \\ \delta(x, d'_x) & \text{jeśli } d_x \in C_S \end{cases}$$

gdzie C_L, C_S – podzbiory dużych i małych grup, $d_x = \arg\min_d \delta(x, d)$, $d'_x = \arg\min_{d \in C_L} \delta(x, d)$



Wskaźniki nieprawidłowości oparte na grupowaniu

- *Local density cluster-based outlier factor* (LDCOF):

$$\psi(x) = \mathrm{LDCOF}_\delta(x) = \begin{cases} \dfrac{\delta(x,d_x)}{\sum_{x' \in T_{d_x}} \delta(x',d_x)/|T_{d_x}|} & \text{jeśli } d_x \in C_L \\ \dfrac{\delta(x,d'_x)}{\sum_{x' \in T_{d'_x}} \delta(x',d'_x)/|T_{d'_x}|} & \text{jeśli } d_x \in C_S \end{cases}$$

gdzie C_L, C_S – podzbiory dużych i małych grup,
$d_x = \arg\min_d \delta(x,d)$, $d'_x = \arg\min_{d \in C_L} \delta(x,d)$.