Zestaw 6 - Zadanie 1

Metody probabilistyczne w uczeniu maszynowym

Łukasz Trzos

Treść zadania

Wykaż, że skończone zbiory $A, B \subset \mathbb{R}^k$ są liniowo separowalne wtedy i tylko wtedy, gdy $\mathrm{Conv}A \cap \mathrm{Conv}B = \emptyset$.

Rozwiązanie

Udowodnimy implikacje w obydwie strony

Definicje

Dla dowolnego zbioru $A \subset \mathbb{R}^k$ definiujemy jego otoczkę wypukłą Conv A jako najmniejszy wypukły zbiór zawierający A. Jeśli $A = \{a_1, ..., a_n\}$, to

$$\mathrm{Conv}A = \{\sum_{i=1}^{n} \lambda_i a_i : \sum_{i=1}^{n} \lambda_i = 1 \wedge \forall_{i \in [n]} \lambda_i \geqslant 0\}.$$

Mówimy, że dwa zbiory $A, B \subseteq \mathbb{R}^k$ są liniowo separowalne jeśli istnieją wektor $w \in \mathbb{R}^k$ oraz $w_0 \in \mathbb{R}$ takie, że dla wszystkich $a \in A$ i $b \in B$ zachodzą nierówności

$$w^T a + w_0 > 0,$$

$$w^T b + w_0 < 0.$$

Dowód $\implies$

Niech $A, B \subset \mathbb{R}^k$ będą liniowo separowalnymi zbiorami oraz niech $w \in \mathbb{R}^k$ oraz $w_0 \in \mathbb{R}$ świadczą o ich separowalności. Weźmy dowolny element $x \in \mathrm{Conv}A$. Zachodzi:

$$w^T x = w^T (\sum_{i=1}^{n} \lambda_i a_i) = \sum_{i=1}^{n} \lambda_i w^T a_i > \sum_{i=1}^{n} \lambda_i (-w_0) = -w_0$$

gdzie w 1 równości korzystamy z definicji otoczki wypukłej w celu zmiany postaci x, natomiast w nierówności wykorzystujemy definicję wektora świadczącego o separowalności zbiorów.
Analogicznie dla dowolnego $y \in \mathrm{Conv}B$:

$$w^T y = w^T (\sum_{j=1}^{m} \mu_j b_j) = \sum_{j=1}^{m} \mu_j w^T b_j < \sum_{j=1}^{m} \mu_j (-w_0) = -w_0$$

Skoro $\forall_{x \in \mathrm{Conv}A} : w^T x > -w_0$, oraz $\forall_{y \in \mathrm{Conv}B} : w^T y < -w_0$, to $\mathrm{Conv}A \cap \mathrm{Conv}B = \emptyset$ $\square$

Dowód $\Longleftarrow$

Niech $A, B \subset \mathbb{R}^k$ takie, że $\text{Conv}A \cap \text{Conv}B = \emptyset$. Wiemy, że istnieje unikalna para punktów $a_0 \in \text{Conv}A$ oraz $b_0 \in \text{Conv}B$ taka, że:

$$\forall_{a \in ConvA, b \in ConvB} : (a \neq a_0 \vee b \neq b_0) \implies ||a - b|| < ||a_0 - b_0||$$

czyli innymi słowy, a_0 i b_0 minimalizują dystans między zbiorami. Wynika to z wypukłości otoczek wypukłych.

Niech $w = b_0 - a_0$. Oczywiście, $w \neq 0$, bo otoczki wypukłe sa rozłączne, zatem $b_0 \neq a_0$. Definiujemy dwie hiperpłaszczyzny prostopadłe do wektora w przechodzące kolejno przez a_0 i b_0:

$$v^T x = w^T a_0 = c_A$$

$$v^T x = w^T b_0 = c_B$$

Dystans między tymi hiperpłaszczyznami jest niezerowy, ponieważ wynosi on $w^T b_0 - w^T a_0 = w^T(b_0 - a_0) = (b_0 - a_0)^T(b_0 - a_0) = ||b_0 - a_0||^2 > 0$. Chcemy pokazać, że $\forall_{a \in ConvA} : w^T a \leqslant c_A$ oraz $\forall_{b \in ConvA} : w^T b \geqslant c_B$. Wtedy możemy wziąć dowolne $w_0 \in (c_A, c_B)$, (wiemy że istnieje ze względu na dystans między hiperpłaszczyznami), które wraz z wektorem w będzie świadczyć o liniowej separowalności.

Pokażemy, że $\forall_{a \in ConvA} : w^T a \leqslant c_A$ nie wprost. Niech $a \in ConvA$ takie, że $w^T a \geqslant c_A$. Wiemy, że do $ConvA$ musi należeć cały odcinek $[a_0, a]$, ponieważ możemy brać: $\lambda a_0 + (1 - \lambda)a$ dla dowolnej $\lambda \in [0, 1]$. Jeśli $[a_0, a]$ jest równoległy do $[a_0, b_0]$ to oczywiście a leży bliżej b_0, sprzeczność. Jeśli nie, to rozważamy punkt a' będący rzutem b_0 na odcinek $[a_0, a]$. Dystans od b_0 do a' to z definicji najkrótszy dystans z b_0 do tego odcinka, w szczególności krótszy niż do a_0, a wiemy, że $a' \in ConvA$, sprzeczność. Dowód drugiego faktu wygląda dokładnie identycznie.

Znaleźliśmy więc wektor świadczący o liniowej separowalności zbiorów A i B. $\square$