
Nichtlineare Trennung von Punktmenge

www.bankboard.de - kontakt@bankboard.de

1. März 2024

Bereits 1965 hat Olvi L. Mangasarian in seinem Artikel “Linear and Nonlinear Separation of Patterns by Linear Programming“ (vgl. Mangasarian, 1965) die Trennung beliebiger Punktmenge mittels Linear Programmierung betrachtet. 1990 wurden diese Methoden dann weiterentwickelt und als Neuronales Netz zur Erkennung von Krebszellen angewandt (vgl. O. Mangasarian, 1990). Insbesondere die massiv gesteigerten Rechnerleistungen machen diesen Algorithmus heute wieder interessant. Im Folgenden sollen diese Verfahren dargestellt werden. Anschließend werden zur Trennung beliebiger Punktmenge verschiedene Algorithmen entwickelt.

1 Konvexe Menge

Die folgenden Aussagen gelten für beliebige Vektorräume und überwiegend auch für affine Räume. Trotzdem soll zur besseren Lesbarkeit auf eine Verallgemeinerung verzichtet werden. Es wird daher stets der reelle Vektorraum \mathbb{R}^n betrachtet.

Bezeichnung 1.1. Für $a \in \mathbb{R}^n$ bezeichnet a_i die i -te Vektorkomponente. Bei Vektoren sind $<, \leq, >, \geq, =, \neq$ komponentenweise zu verstehen, d.h. für $a \in \mathbb{R}^n$ bedeutet $a > 0$, dass $a_i > 0$ für alle $i = 1, \dots, n$ gilt. 1_n bezeichnet den Vektor $(1, \dots, 1)^T$ aus \mathbb{R}^n .

Definition 1.2. Eine Menge $C \subset \mathbb{R}^n$ heißt **konvex**, wenn für beliebige $x^{(1)}, x^{(2)} \in C$ und $\lambda \in \mathbb{R}$ mit $0 \leq \lambda \leq 1$ gilt:

$$\lambda \cdot x^{(2)} + (1 - \lambda) \cdot x^{(1)} \in C.$$

Mit $x^{(1)}, x^{(2)} \in C$, C konvex, ist somit auch die Strecke

$$\{x \in \mathbb{R}^n \mid x^{(1)} + \lambda \cdot (x^{(2)} - x^{(1)}), 0 \leq \lambda \leq 1\}$$

zwischen $x^{(1)}$ und $x^{(2)}$ in C enthalten. Weitere leicht zu verifizierende Beispiele von konvexen Mengen sind:

Beispiel 1.3. (i) Die leere Menge \emptyset ist konvex.

(ii) Jede einelementige Menge $\{x\}$, $x \in \mathbb{R}^n$, ist konvex.

(iii) \mathbb{R}^n und jeder Untervektorraum von \mathbb{R}^n sind konvex.

(iv) Jeder affine Unterraum von \mathbb{R}^n ist konvex.

(v) Sei $a \in \mathbb{R}^n, b \in \mathbb{R}$. Der offene Halbraum $\{x \in \mathbb{R}^n \mid \langle a, x \rangle < b\}$ und der abgeschlossene Halbraum $\{x \in \mathbb{R}^n \mid \langle a, x \rangle \leq b\}$ sind konvex.

(vi) Sei $A \in \mathbb{R}^{n \times m}$ und $b \in \mathbb{R}^m$. Dann ist $\{x \in \mathbb{R}^n \mid Ax = b\}$ konvex.

Es soll beispielhaft (vi) gezeigt werden:

Sei $x^{(1)}, x^{(2)} \in \{x \in \mathbb{R}^n \mid Ax = b\}$ und $\lambda \in \mathbb{R}$ mit $0 \leq \lambda \leq 1$. Dann ist

$$A(\lambda \cdot x^{(2)} + (1 - \lambda) \cdot x^{(1)}) = \lambda b + (1 - \lambda)b = b.$$

Somit ist $\lambda \cdot x^{(2)} + (1 - \lambda) \cdot x^{(1)}$ in $\{x \in \mathbb{R}^n \mid Ax = b\}$. Oder auch: Falls keine Lösung des Gleichungssystem existiert, ist die Menge nach (i) konvex. Andernfalls ist $\{x \in \mathbb{R}^n \mid Ax = b\} = b + \text{Ker}(A)$. Letzteres ist ein affiner Raum, daher konvex nach (v).

Lemma 1.4.

(i) Seien $C_i, i \in I$, konvexe Mengen in \mathbb{R}^n . Dann ist auch $\bigcap_{i \in I} C_i$ konvex.

(ii) Mit $C_i, i = 1, \dots, k, C_i \subset \mathbb{R}^n$ ist auch

$$C_1 + \dots + C_k := \{x \in \mathbb{R}^n \mid x = x_1 + \dots + x_k; x_i \in C_i\}$$

konvex.

(iii) Ist $C \subset \mathbb{R}^n$ konvex und $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine lineare Abbildung, dann ist auch das Bild $L(C) \subset \mathbb{R}^m$ konvex.

Beweis.

zu (i): Falls $\bigcap_{i \in I} C_i$ leer oder einelementig ist, ist die Behauptung nach Beispiel 3.6 (i) bzw. (ii) erfüllt. Sei nun $x, y \in \bigcap_{i \in I} C_i$. Dann gilt für alle $i \in I$: $x, y \in C_i$ und $\lambda x + (1 - \lambda)y \in C_i$ mit $0 \leq \lambda \leq 1$ wegen C_i konvex.

zu (ii): komponentenweise Anwendung der Definition

zu (iii): Mit a, b in $L(C)$ gibt es $x, y \in C$ mit $a = L(x), b = L(y)$. Für $0 \leq \lambda \leq 1$ ist auch $\lambda x + (1 - \lambda)y \in C$ und $L(\lambda x + (1 - \lambda)y) = \lambda L(x) + (1 - \lambda)L(y) = \lambda a + (1 - \lambda)b$ in $L(C)$. \square

Definition 1.5. Sei $k \in \mathbb{N}$ und $x^{(i)} \in \mathbb{R}^n, i = 1, \dots, k$. Eine **konvexe Kombination** der Menge $\{x^{(1)}, \dots, x^{(k)}\}$ ist jeder Punkt der Form:

$$\sum_{i=1}^k \lambda_i x^{(i)}, \lambda_i \geq 0, \sum_{i=1}^k \lambda_i = 1.$$

In einer konvexen Menge C sind offensichtlich alle zweielementigen konvexen Kombinationen wieder in C . Dass dies für alle konvexen Kombinationen gilt, zeigt folgendes Lemma:

Lemma 1.6. Eine Menge $C \subset \mathbb{R}^n$ ist genau dann konvex, wenn es alle konvexen Kombinationen aus beliebigen Elementen aus C enthält.

Beweis. Der Beweis wird über Induktion geführt (vgl. Tuy, 2016, Proposition 1.5, p. 6). \square

Definition 1.7. Sei $A \subset \mathbb{R}^n$ eine beliebige Menge. Dann heißt

$$\text{conv}(A) := \bigcap \{C \mid C \subset \mathbb{R}^n \text{ konvex und } A \subset C\}$$

konvexe Hülle von A .

Es ist $A \subset \mathbb{R}^n$ und \mathbb{R}^n konvex. Somit ist die konvexe Hülle für alle A definiert. Zudem ist $\text{conv}(A)$ wegen Lemma 1.4 (i) konvex. Nach Definition ist $\text{conv}(A)$ zudem die kleinste konvexe Menge, die A enthält. Für eine konvexe Menge C gilt offensichtlich $\text{conv}(C) = C$.

Es lässt sich folgender wichtiger Satz zeigen:

Satz 1.8 (Caratheodory). Sei $A \subset \mathbb{R}^n$. Dann gilt:

$$\text{conv}(A) = \left\{ \sum_{i=1}^{n+1} \lambda_i x^{(i)} \mid x^{(i)} \in A, \lambda_i \geq 0, \sum_{i=1}^{n+1} \lambda_i = 1 \right\},$$

d.h. jedes Element ist die konvexe Kombination aus maximal $(n + 1)$ Elementen von A , wobei n die Dimension des Vektorraums \mathbb{R}^n ist.

Beweis. vgl. Güler, 2010, Theorem 4.13, p. 91 \square

Folgende Bemerkung wird im Abschnitt 3 benötigt:

Bemerkung 1.9. Sei $A = \{a^{(1)}, \dots, a^{(k)}\} \subset \mathbb{R}^n$. Gilt nun $c^T \cdot a^{(i)} > \alpha$ für ein $c \in \mathbb{R}^n$ und ein $\alpha \in \mathbb{R}$ für alle $i = 1, \dots, k$, dann ist nach dem Satz von Caratheodory ein beliebiges $a \in \text{conv}(A)$ durch $a = \sum_{i=1}^{n+1} \lambda_i x^{(i)}$, $x^{(i)} \in A$, $\lambda_i \geq 0$, $\sum_{i=1}^{n+1} \lambda_i = 1$ darstellbar. Damit ist $c^T \cdot a = \sum_{i=1}^{n+1} \lambda_i c^T \cdot x^{(i)} < \sum_{i=1}^{n+1} \lambda_i \alpha = \alpha$ für alle $a \in \text{conv}(A)$.

2 Trennbarkeit

Der folgende Abschnitt zeigt zwei wichtige Trennbarkeitssätze für konvexe Mengen.

Definition 2.1. Die Mengen $A, B \subset \mathbb{R}^n$ heißen **trennbar**, falls es ein $c \in \mathbb{R}^n$, $c \neq 0$ und ein $\alpha \in \mathbb{R}$ mit

$$x \in A \Rightarrow c^T x \leq \alpha \text{ und } x \in B \Rightarrow c^T x \geq \alpha$$

gibt.

$A, B \subset \mathbb{R}^n$ heißen **strikt trennbar**, falls

$$x \in A \Rightarrow c^T x < \alpha \text{ und } x \in B \Rightarrow c^T x > \alpha$$

gilt.

Die Hyperebene $E = \{x \in \mathbb{R}^n \mid c^T x = \alpha\}$ heißt dabei **trennende** bzw. **strikt trennende Hyperebene**.

Satz 2.2 (Trennung konvexer Mengen). Sind $A, B \subset \mathbb{R}^n$ konvex und gilt $A \cap B = \emptyset$, dann sind A und B trennbar.

Beweis. vgl. Tuy, 2016, Theorem 1.3, p. 17 □

Satz 2.3 (Strikte Trennung konvexer Mengen). Seien $A, B \subset \mathbb{R}^n$ konvexe, abgeschlossen Mengen. Zudem sei eine der Mengen kompakt. Gilt nun $A \cap B = \emptyset$, dann sind A und B strikt trennbar.

Beweis. vgl. Tuy, 2016, Theorem 1.4, p. 18 □

Die Kompaktheit einer der Mengen ist für die strikte Trennung notwendig, wie folgendes Beispiel zeigt:

Beispiel 2.4. Sei $A = \{(x, y) \in \mathbb{R}^2 | x \leq 0\}$ und $B = \{(x, y) \in \mathbb{R}^2 | x \cdot y \geq 1\}$. Dann sind beide Mengen konvex, abgeschlossen und disjunkt, aber nicht kompakt. Die Hyperebene $E = \{(x, y) \in \mathbb{R}^2 | x = 0\}$ trennt die beiden Mengen. Diese Trennung ist aber nicht strikt, da beide Punktmenge sich für $x \rightarrow \infty$ beliebig nahe kommen.

3 Lineare Programmierung und Trennbarkeit

Bezeichnung 3.1. Im Folgenden werden nun endliche Punktfolgen $\mathcal{A} = \{a^{(1)}, \dots, a^{(m)}\}$ und $\mathcal{B} = \{b^{(1)}, \dots, b^{(k)}\}$ aus \mathbb{R}^n betrachtet. Mit $A = [a^{(1)}, \dots, a^{(m)}]^T \in \mathbb{R}^{m \times n}$ und $B = [b^{(1)}, \dots, b^{(k)}]^T \in \mathbb{R}^{k \times n}$ werden die Matrizen betrachtet, deren Zeilen die einzelnen Punkte aus \mathcal{A} bzw. \mathcal{B} darstellen.

\mathcal{A} und \mathcal{B} sollen trennbar heißen, wenn ihre konvexen Hüllen $\text{conv}(\mathcal{A})$ bzw. $\text{conv}(\mathcal{B})$ trennbar sind. Falls endliche Punktfolgen trennbar sind, sind diese stets strikt trennbar, da $\text{conv}(\mathcal{A})$ und $\text{conv}(\mathcal{B})$ kompakt und abgeschlossen sind.

Lemma 3.2. Es gelten die Bezeichnungen wie in 3.1 beschreiben. \mathcal{A} und \mathcal{B} sind genau dann trennbar, wenn es $c \in \mathbb{R}^n$, $c \neq 0$, und $\alpha \in \mathbb{R}$ gibt, mit

$$Ac - \alpha 1_m > 0 \text{ und } Bc - \alpha 1_k < 0.$$

Beweis. Seien zunächst \mathcal{A} und \mathcal{B} trennbar, dann gibt es per Definition ein $c \in \mathbb{R}^n$ und ein $\alpha \in \mathbb{R}$ mit $c^T x - \alpha > 0$ für $x \in \mathcal{A}$ und $c^T x - \alpha < 0$ für $x \in \mathcal{B}$. Wegen $c^T x = x^T c$ und der Definition von A und B gilt somit $Ac - \alpha 1_m > 0$ bzw. $Bc - \alpha 1_k < 0$. Gilt nun umgekehrt $Ac - \alpha 1_m > 0$ und $Bc - \alpha 1_k < 0$, dann ist $c^T x - \alpha > 0$ für $x \in \mathcal{A}$ und $c^T x - \alpha < 0$ für $x \in \mathcal{B}$. Nach Bemerkung 1.9 gilt dies dann für alle $a \in \text{conv}(\mathcal{A})$ bzw. $b \in \text{conv}(\mathcal{B})$ und \mathcal{A} und \mathcal{B} sind trennbar. □

Lemma 3.3. Es gelten die Bezeichnungen wie in 3.1 beschreiben. \mathcal{A} und \mathcal{B} sind genau dann trennbar, wenn es $c \in \mathbb{R}^n$ und $\alpha, \beta \in \mathbb{R}$ gibt, mit

$$Ac - \alpha 1_m \geq 0, -Bc + \beta 1_k \geq 0,$$

$$\alpha - \beta > 0 \text{ und } 1_n \geq c \geq -1_n.$$

Beweis. Angenommen es gibt ein $c \in \mathbb{R}^n$ und $\alpha, \beta \in \mathbb{R}$, die obige Ungleichungen erfüllen. Dann ist $\alpha > \beta$ und damit $\frac{\alpha+\beta}{2} < \alpha$ bzw. $\frac{\alpha+\beta}{2} > \beta$. Mit $\gamma := \frac{\alpha+\beta}{2}$ gilt daher

$$Ac - \gamma 1_m > 0 \text{ und } -Bc + \gamma 1_k > 0.$$

Nach Lemma 3.2 sind dann \mathcal{A} und \mathcal{B} trennbar. Dies zeigt die eine Richtung.

Sei nun \mathcal{A} und \mathcal{B} trennbar. Dann gibt es nach Lemma 3.2 ein $\gamma \in \mathbb{R}$ und ein $d \in \mathbb{R}^n$, $d \neq 0$, mit

$$Ad - \gamma 1_m > 0 \text{ und } -Bd + \gamma 1_k > 0 \quad (\star)$$

Definiere mit δ die größte, absolute Vektorkomponente von d , d.h. für $d = (d_1, \dots, d_n)$ ist $\delta = \max_{i=1, \dots, n} |d_i|$. Dann gilt für $c := \frac{d}{\delta}$:

$$-1_n \leq c \leq 1_n.$$

Bezeichne mit α die kleinste Vektorkomponente des Vektors Ac und mit β die größte Vektorkomponente des Vektors Bc , d.h. $Ac - \alpha 1_m \geq 0$ und $-Ac + \beta 1_m \geq 0$. Zudem gilt wegen Konstruktion von α bzw. β und (\star) :

$$Ac - \frac{\gamma}{\delta} \cdot 1_m \geq (\alpha - \frac{\gamma}{\delta}) \cdot 1_m > 0$$

$$Bc - \frac{\gamma}{\delta} \cdot 1_k \leq (\beta - \frac{\gamma}{\delta}) \cdot 1_k < 0.$$

Aus $\alpha - \frac{\gamma}{\delta} > 0$ und $\beta - \frac{\gamma}{\delta} < 0$ folgt:

$$\alpha - \frac{\gamma}{\delta} - (\beta - \frac{\gamma}{\delta}) > 0$$

und damit $\alpha - \beta > 0$. □

Lemma 3.3 lässt sich nun leicht in ein lineares Programm übersetzen:

Satz 3.4. *Es gelten die Bezeichnungen wie in 3.1 beschreiben. \mathcal{A} und \mathcal{B} sind genau dann trennbar, wenn folgendes Lineares Programm eine positive Lösung hat:*

$$\begin{aligned} \text{Maximiere} \quad & \alpha - \beta \\ \text{u.d.B.} \quad & Ac - \alpha 1_m \geq 0 \\ & -Bc + \beta 1_k \geq 0 \\ & 1_n \geq c \geq -1_n. \end{aligned}$$

Beweis. Wenn es eine positive Lösung gibt, ist $\alpha - \beta > 0$. Die Trennbarkeit folgt dann direkt aus Lemma 3.3. Wenn auf der anderen Seite die Punkt Mengen trennbar sind, gibt es laut Lemma 3.3 α, β mit $\alpha - \beta > 0$ und das Programm hat eine positive Lösung. □

Durch das Lineare Programm werden bei trennbaren Punkt Mengen \mathcal{A} und \mathcal{B} zwei parallele Hyperbenen $E_1 = \{x \in \mathbb{R}^n | c^T x = \alpha\}$ und $E_2 = \{x \in \mathbb{R}^n | c^T x = \beta\}$ erzeugt. E_1 bzw. E_2 haben mindestens einen Punkt mit \mathcal{A} bzw. \mathcal{B} gemeinsam. Zudem ist der Abstand der beiden Hyperbenen wegen der Zielfunktion $\alpha - \beta$ maximal. Die Hyperebene $E = \{x \in \mathbb{R}^n | c^T x = \frac{\alpha+\beta}{2}\}$ ist

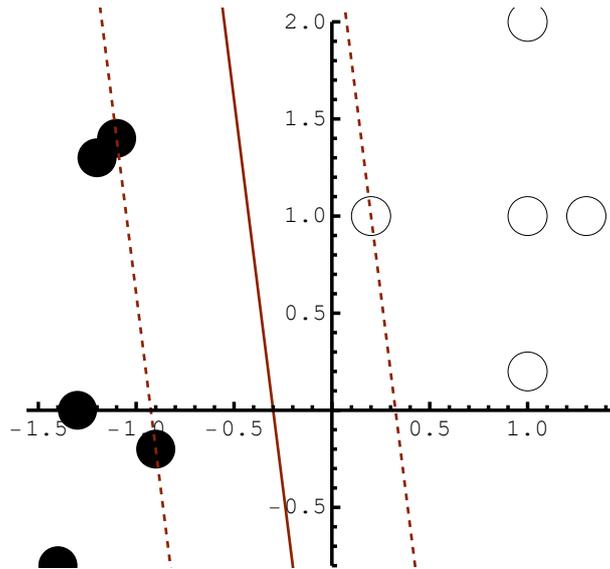


Abbildung 1: Darstellung der Trennung aus Beispiel 3.6

eine "optimale" Hyperebene zum Trennen der Punktmengen.

Satz 3.5. Das Lineare Programm aus Satz 3.4 lässt sich in die Form

$$\begin{aligned} & \text{Maximiere } d^T x \\ & \text{u.d.B. } Dx \leq b \\ & \quad x \geq 0 \end{aligned}$$

überführen.

Beweis. Zur Vereinfachung der Notation wird mit $0_n = (0, \dots, 0)^T$ der Nullvektor in \mathbb{R}^n bezeichnet und mit I_n die Identitätsmatrix aus $\mathbb{R}^{n \times n}$.

Zunächst muss $x \geq 0$ erfüllt sein. Hierzu wird aus $c = c^+ - c^-$, $\alpha = \alpha^+ - \alpha^-$ und $\beta = \beta^+ - \beta^-$ mit $c^+, c^-, \alpha^+, \alpha^-, \beta^+, \beta^- \geq 0$. Der gesuchte Vektor x ist dann $x = ((c^+)^T, \alpha^+, \beta^+, (c^-)^T, \alpha^-, \beta^-)^T \in \mathbb{R}^{2n+4}$ und $d = ((0_n)^T, 1, -1, (0_n)^T, -1, 1)^T \in \mathbb{R}^{2n+4}$.

Die Bedingung $1_n \geq c \geq -1_n$ wird zu $c^+ - c^- \leq 1_n$ und $-c^+ + c^- \leq 1_n$. Die Matrix $D \in \mathbb{R}^{(2n+m+k) \times (2n+4)}$ lässt sich folgendermaßen schreiben:

$$D = \begin{pmatrix} -A & 1_m & 0_m & A & -1_m & 0_m \\ B & 0_k & -1_k & -B & 0_k & 1_k \\ I_n & 0_n & 0_n & -I_n & 0_n & 0_n \\ -I_n & 0_n & 0_n & I_n & 0_n & 0_n \end{pmatrix}$$

und b ist $((0_m)^T, (0_k)^T, (1_n)^T, (1_n)^T)^T$ mit $b \in \mathbb{R}^{2n+m+k}$. □

Beispiel 3.6. Es sollen die Mengen $\mathcal{A} = \{(1, 1)^T, (0.2, 1)^T, (1, 0.2)^T, (1.3, 1)^T, (1, 2)^T\}$ und $\mathcal{B} = \{(-1.2, 1.3)^T, (-1.4, -0.8)^T, (-1.3, 0)^T, (-0.9, -0.2)^T\}$,

$(-1.1, 1.4)^T$ getrennt werden. Es ist $m = 5$, $k = 5$ und $n = 2$.

Dann ist $d = (0, 0, 1, -1, 0, 0, -1, 1)^T$ und $b = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1)^T$. Die Matrix D ist

$$D = \begin{pmatrix} -1 & -1 & 1 & 0 & 1 & 1 & -1 & 0 \\ -0.2 & -1 & 1 & 0 & 0.2 & 1 & -1 & 0 \\ -1 & -0.2 & 1 & 0 & 1 & 0.2 & -1 & 0 \\ -1.3 & -1 & 1 & 0 & 1.3 & 1 & -1 & 0 \\ -1 & -2 & 1 & 0 & 1 & 2 & -1 & 0 \\ -1.2 & 1.3 & 0 & -1 & 1.2 & -1.3 & 0 & 1 \\ -1.4 & -0.8 & 0 & -1 & 1.4 & 0.8 & 0 & 1 \\ -1.3 & 0 & 0 & -1 & 1.3 & 0 & 0 & 1 \\ -0.9 & -0.2 & 0 & -1 & 0.9 & 0.2 & 0 & 1 \\ -1.1 & 1.4 & 0 & -1 & 1.1 & -1.4 & 0 & 1 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Die Berechnung des Linearen Programm mit Standardalgorithmen liefert für

$$\begin{aligned} & \text{Maximiere } d^T x \\ & \text{u.d.B. } Dx \leq b \\ & \quad x \geq 0 \end{aligned}$$

den Lösungsvektor $x = (1, 0.125, 0.325, 0, 0, 0, 0, 0.925)^T$ und damit

$$(c, \alpha, \beta) = ((1, 0.125), 0.325, 0)^T - ((0, 0), 0, 0.925)^T = ((1, 0.125), 0.325, -0.925)^T.$$

Die "optimal" trennende Hyperbene von \mathcal{A} und \mathcal{B} ist dann

$$E = \left\{ x \in \mathbb{R}^2 \mid (1, 0.125)x - \frac{0.325 - 0.925}{2} \right\}.$$

Der Algorithmus 1 in Abbildung 2 fasst die Ergebnisse zusammen.

4 Nicht-linear trennbare Punktmengeten

Falls die Punktmengeten \mathcal{A} und \mathcal{B} nicht trennbar sind, liefert das lineare Programm die Lösung $x = (0, \dots, 0)^T \in \mathbb{R}^{2n+4}$ und damit $c = (0, \dots, 0)^T \in \mathbb{R}^n$.

Es ist allerdings möglich in diesem Fall eine Lösung mit $c \neq 0$ zu erzwingen. Hierzu wird das Lineare Programm aus Satz 3.5 so verändert, dass eine Komponente von c ungleich 0 ist. Beispielsweise kann $c_i = 1$ durch Veränderung von b folgendermaßen "erzungen" werden. Sei dazu $1_n^{(i)} = (1, \dots, 1, -1, 1, \dots, 1) \in \mathbb{R}^n$ der Vektor 1_n mit -1 an i -ter Komponente. Dann ist mit

A1: Lineare Trennung

INPUT: \mathcal{A} und \mathcal{B} zwei Punktmengen aus \mathbb{R}^n

STEP 1:

Setze

$$D = \begin{pmatrix} -A & 1_m & 0_m & A & -1_m & 0_m \\ B & 0_k & -1_k & -B & 0_k & 1_k \\ I_n & 0_n & 0_n & -I_n & 0_n & 0_n \\ -I_n & 0_n & 0_n & I_n & 0_n & 0_n \end{pmatrix}$$

$$b = ((0_m)^T, (0_k)^T, (1_n)^T, (1_n)^T)^T \in \mathbb{R}^{2n+m+k}.$$

$$d = ((0_n)^T, 1, -1, (0_n)^T, -1, 1)^T \in \mathbb{R}^{2n+4}$$

STEP 2:

Maximiere $d^T x$

u.d.B. $Dx \leq b$

$x \geq 0$

OUTPUT:

Aus der Lösung x berechne:

$$c = (x_1, \dots, x_n)^T - (x_{n+3}, \dots, x_{2n+2})^T$$

$$\alpha = x_{n+1} - x_{2n+3}$$

$$\beta = x_{n+2} - x_{2n+4}$$

INTERPRETATION:

$c \neq 0$: $E = \{x \in \mathbb{R}^n \mid c^T x = \frac{\alpha+\beta}{2}\}$ trennt \mathcal{A} und \mathcal{B} .

$c = 0$: \mathcal{A} und \mathcal{B} sind nicht linear trennbar.

Abbildung 2: Algorithmus zur linearen Trennung zweier beliebigen Punktmengen aus \mathbb{R}^n

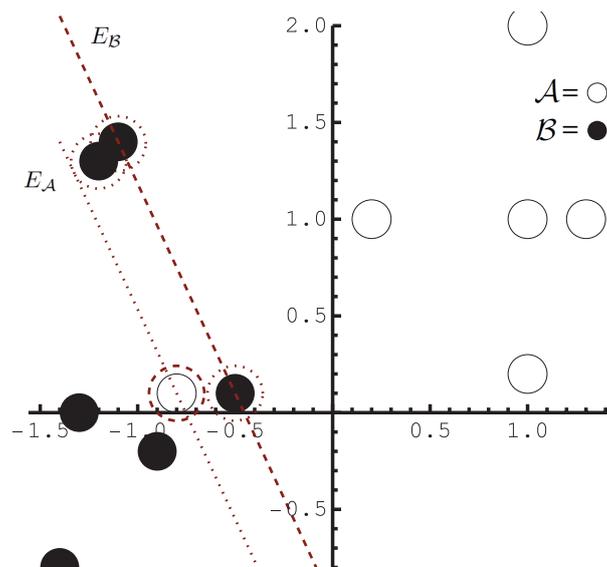


Abbildung 3: Darstellung der Trennung aus Beispiel 4.1. Die markierten Punkte werden von den Ebenen falsch klassifiziert.

A2: Trennung nichtleerer Punktmenge

INPUT: \mathcal{A} und \mathcal{B} zwei Punktmenge aus \mathbb{R}^n und $i = 1$.

STEP 1:

Löse das Lineare Programm aus Satz 3.5.

Falls $c \neq 0$ dann

Setze $c(i) = c$, $a(i) = \alpha$ und $b(i) = \beta$ und STOPP

Sonst

Setze $\#F = \infty$.

Für $j = 1$ bis n und für $t \in \{-1, 1\}$:

Setze $c_j = t$ und löse modifiziertes LP $(1_n^{(j)})$ in b

Ermittle alle Punkte mit $x \in \mathcal{A}$ und $c^T x - \beta < 0$
und alle Punkte mit $x \in \mathcal{B}$ und $c^T x - \alpha > 0$.

Falls die Anzahl l dieser Punkte $l < \#F$ setze
 $\#F = l$ und speichere die Punkte in $\tilde{\mathcal{A}}$ bzw. $\tilde{\mathcal{B}}$
und setze $c(i) = c$, $a(i) = \alpha$ und $b(i) = \beta$

Next

STEP 2:

Setze $i = i + 1$, $\mathcal{A} = \tilde{\mathcal{A}}$, $\mathcal{B} = \tilde{\mathcal{B}}$

Gehe zu STEP 1

OUTPUT: $i_{max} = i$ und $c(i)$, $a(i)$, $b(i)$

Abbildung 4: Algorithmus zur (auch nicht) linearen Trennung zweier beliebigen Punktmenge aus \mathbb{R}^n

$b = ((0_m)^T, (0_k)^T, (1_n)^T, (1_n^{(i)})^T)^T$ die Nebenbedingung $-c_i^+ + c_i^- \leq -1$. Zugleich gilt $c_i^+ - c_i^- \leq 1$ unverändert, d.h. in Summe $c_i = c_i^+ - c_i^- = 1$. Analog ist für $b = ((0_m)^T, (0_k)^T, (1_n^{(i)})^T, (1_n)^T)^T$ $c_i = -1$.

Beispiel 4.1. Der Punkt $(-0.8, 0.1)^T$ wird in \mathcal{A} und $(-0.5, 0.1)^T$ in \mathcal{B} zusätzlich aufgenommen. Das Lineare Programm aus Satz 3.5 liefert einen Nullvektor als Lösung, d.h. die Punkte sich nicht linear trennbar. Das modifizierte Lineare Programm mit $c_1 = 1$ liefert folgende Lösungen:

$$\alpha = -0.753846 \text{ und } \beta = -0.453846$$

$$E_{\mathcal{A}} = \{x \in \mathbb{R}^2 \mid (1., 0.461538)x = -0.753846\}$$

$$E_{\mathcal{B}} = \{x \in \mathbb{R}^2 \mid (1., 0.461538)x = -0.453846\}.$$

Ermittle nun alle Punkte $(x_1, x_2)^T$ aus \mathcal{B} mit $(x_1, x_2)^T(1., 0.461538) + 0.753846 > 0$ und alle Punkte $(x_1, x_2)^T$ aus \mathcal{A} mit $(x_1, x_2)^T(1., 0.461538) + 0.453846 < 0$ (vgl. Abbildung 3) und streiche die jeweils identifizierten Punkte aus \mathcal{B} und aus \mathcal{A} . Dann sind die Punktmenge trennbar. Für die gestrichenen Punkte kann ein neuer Trennungsversuch unternommen werden. So können sukzessive alle Punkte getrennt werden.

Durch die Variation $c_i = \pm 1$ für alle $i = 1, \dots, n$ lassen sich die Ebene finden, für die die wenigsten Punkte für eine Trennung eliminiert werden müssen. Diese eliminierten Punkte können nun neu getrennt werden. Durch den Algorithmus 2 in Abbildung 4 können so beliebige

A3: Testen von unbekanntem Punkten

INPUT: P aus \mathbb{R}^n , i_{\max} und $c(i)$, $a(i)$, $b(i)$ aus Algorithmus 4.

STEP:

Für $i = 1$ bis i_{\max}

Falls $c(i)^T P - \beta \geq 0$ dann $K = \text{“A“}$, $k = i$, $i = i_{\max}$

Falls $c(i)^T P - \alpha \leq 0$ dann $K = \text{“B“}$, $k = i$, $i = i_{\max}$

Next

OUTPUT: Zugehörigkeit K des Punktes P und
Trennungstiefe k .

Abbildung 5: Algorithmus zum Testen unbekannter Muster P aus \mathbb{R}^n

Punktmengen getrennt werden.

Ein unbekannter Punkt kann einer Menge eindeutig zugeordnet werden, wenn die Lösungen des Algorithmus 2 aus Abbildung 4 nach und nach angewandt werden bis eine Trennung gefunden ist (vgl. Algorithmus 3 in Abbildung 5). Dabei wird die Trennungskategorisierung zur Punktmenge \mathcal{A} bzw. \mathcal{B} mit höherer Trennungstiefe unschärfer und unsicherer.

Bibliography

- Güler, Osman (2010). *Convex Analysis and Global Optimization*. New York: Springer.
- Mangasarian, O. (1965). „Linear and Nonlinear Separation of Patterns by Linear Programming“. In: *Operation Research, Volume 13, Issue 3*, S. 444–452.
- O. Mangasarian R. Setiono, W. Wolberg (1990). „Pattern Recognition via Linear Programming: Theory and Application to Medical Diagnosis“. In: *T. Coleman, Y. Li (Eds.): Large-Scale Numerical Optimization, SIAM*, S. 22–30.
- Tuy, Hoang (2016). *Convex Analysis and Global Optimization*. Switzerland: Springer.