# Kapitel 5 Klassifikation

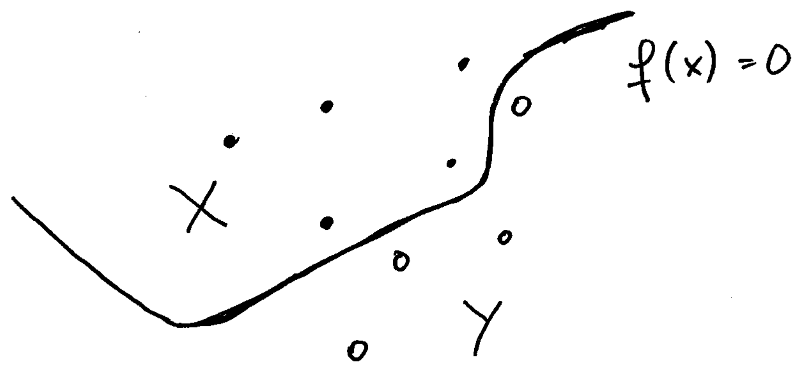
<u>gegeben</u>: Punktmengen

$$X = \{x_1, \dots, x_M\}, \quad Y = \{y_1, \dots, y_N\} \subseteq \mathbb{R}^n$$

<u>gesucht</u>: Funktion $f: \mathbb{R}^n \to \mathbb{R}$ mit

$$f(x_i) > 0, \quad i = 1, \dots, N \quad \text{und} \quad f(y_i) < 0, \quad i = 1, \dots, M.$$

Die <u>0-Level-Menge</u> $\{x \in \mathbb{R}^n : f(x) = 0\}$ von $f$ <u>trennt</u> / <u>klassifiziert</u> die Mengen $X$ und $Y$.




Bei der Suche nach $f$ beschränkt man sich auf eine geeignete (hängt von der konkreten Anwendung ab) Funktionenklasse.

# § 1 Lineare Klassifikation

Verwende für $f$ nur affin-lineare Fkt., d.h. man sucht $f(x) = a^T x - \beta$ mit $a \in \mathbb{R}^n$, $\beta \in \mathbb{R}$.

Dann kann das Problem der linearen Klassifikation mit Methoden der linearen Programmierung behandelt werden.

<u>Satz</u> Die folgenden Aussagen sind äquivalent:

(1) Man kann $X$ und $Y$ mit einer affinen Fkt. trennen.

(2) $\exists a \in \mathbb{R}^n, \beta \in \mathbb{R}$ : $\quad a^T x_i - \beta > 0, \quad i = 1, \dots, M,$
$\quad a^T y_i - \beta < 0, \quad i = 1, \dots, N.$

(3) $\exists a \in \mathbb{R}^n, \beta \in \mathbb{R}$ : $\quad a^T x_i - \beta \geq 1, \quad i = 1, \dots, M,$
$\quad a^T y_i - \beta \leq -1, \quad i = 1, \dots, N.$

(4) $\nexists \lambda \in \mathbb{R}^N_{\geq 0}, \mu \in \mathbb{R}^M_{\geq 0}$ : $\quad e^T \lambda = 1, \quad e^T \mu = 1,$

$$\sum_{i=1}^{M} \lambda_i x_i = \sum_{i=1}^{N} \mu_i y_i ,$$

wobei $e^T = [1 \dots 1]$.

(5) $\operatorname{conv} X \cap \operatorname{conv} Y = \emptyset$

<u>Bew.</u>: <u>klar</u>: $(1) \Longleftrightarrow (2) \Longleftrightarrow (3)$ und $(4) \Longleftrightarrow (5)$

$(3) \Longleftrightarrow (4)$: Farkas Lemma für lineare Programme:

Allgemein: Seien $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Dann gilt

$$\exists x \in \mathbb{R}^n : Ax \le b \iff \nexists y \ge 0 : A^T y = 0, \; b^T y < 0.$$

Wende dies auf

$$x = (a, \beta) \in \mathbb{R}^{n+1}$$

$$A = \begin{bmatrix} -x_i & 1 \\ y_i & -1 \end{bmatrix} \in \mathbb{R}^{(M+N)\times(n+1)}, \quad b = \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix} \in \mathbb{R}^{M+N}$$

an.

Dann $y \in \mathbb{R}^{M+N}$, $y = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_M \\ \mu_1 \\ \vdots \\ \mu_N \end{bmatrix} \ge 0$

$$A^T y = \begin{bmatrix} -x_i & y_i \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_M \\ \mu_1 \\ \vdots \\ \mu_N \end{bmatrix} = \begin{bmatrix} \sum \lambda_i(-x_i) + \sum \mu_i y_i \\ \sum \lambda_i - \sum \mu_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$b^T y = [-1 \dots -1] \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_M \\ \mu_1 \\ \vdots \\ \mu_N \end{bmatrix} = -\sum \lambda_i - \sum \mu_i < 0.$$

Also $\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{N} \mu_i$ und $\sum_i \lambda_i + \sum_i \mu_i > 0$.

Nun können wir $\lambda_i, \mu_i$ so skalieren, dass
$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{N} \mu_i = 1$ gilt. $\boxtimes$

__Bem.:__ Menge der affinen Fkt., die $X$ und $Y$ trennen bilden nach (3) des Satzes ein Polyeder.

__Ziel:__ Finde möglichst gute Trennung, d.h. eine Trennung, die weit von $X$ und $Y$ entfernt ist.

Das folgende konische Programm liefert eine solche Trennung:

$$t^* = \max t$$

$$(*) \quad \begin{aligned} &a^T x_i - \beta \geq t, \quad i = 1, \dots, M \\ &a^T y_i - \beta \leq -t, \quad i = 1, \dots, N \\ &\|a\| \leq 1 \qquad (\Longleftrightarrow (a, 1) \in \mathcal{L}^{n+1}) \\ &a \in \mathbb{R}^n, \beta \in \mathbb{R}, t \in \mathbb{R} \end{aligned}$$

,

wobei $\mathcal{L}^{n+1} = \{ (x, t) \in \mathbb{R}^{n+1} : \|x\| \leq t \}$ der __Lorentz-Kegel__ ist.

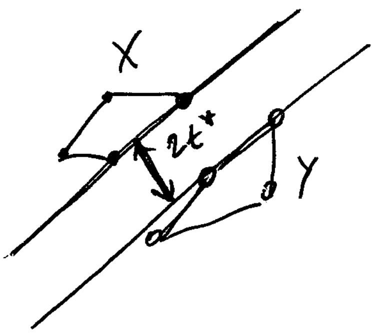
**Satz** (a) $t^* > 0 \Leftrightarrow$ X und Y können durch eine affine Fkt. getrennt werden.

(b) Falls $t^* > 0$, dann ~~gibt es eine~~ gilt für alle optimale Lösungen $a^* \in \mathbb{R}^n$: ~~mit~~ $\|a^*\| = 1$.

(c) Falls $t^* > 0$, dann gilt

$$t^* = \frac{1}{2} \min \{ \|x-y\| : x \in \operatorname{conv} X, y \in \operatorname{conv} Y \}.$$

M.a.W.: Aus (b) und (c) folgt, ~~dass $2t^*$~~ dass $2t^*$ die maximale Breite eines Streifens angibt, der conv X und conv Y trennt.




**Bew.:** (a): ✓

(b) Ang. $a \in \mathbb{R}^n, \beta \in \mathbb{R}, t \in \mathbb{R}$ sind zulässig für (*) und es ist $\|a\| < 1$, $t > 0$.

Dann gilt $\quad \frac{a^T}{\|a\|} x_i - \frac{\beta}{\|a\|} \geq \frac{t}{\|a\|} > t$

und $\quad \frac{a^T}{\|a\|} y_i - \frac{\beta}{\|a\|} \le -\frac{t}{\|a\|} < -t$ ,

D.h. wenn man a durch $\frac{a}{\|a\|}$ ersetzt, kann man $t$ durch $\frac{t}{\|a\|} > t$ ersetzen.

(c) Schreibe (*) als konisches Programm in primaler Standardform:

$$
\begin{aligned}
t^* = \max\ & t \\
& t \in \mathbb{R} \\
& (a, \alpha) \in \mathcal{L}^{n+1} \\
& \beta \in \mathbb{R} \\
& r \in \mathbb{R}^N_{\geq 0} \\
& s \in \mathbb{R}^M_{\geq 0}
\end{aligned}
$$

$$
\begin{aligned}
\gamma: \quad & \alpha = 1 \\
\lambda_i: \quad & -a^T x_i + \beta + t + r_i = 0 \quad , \quad i = 1, \ldots, M \\
\mu_i: \quad & a^T y_i - \beta + t + s_i = 0 \quad , \quad i = 1, \ldots, N
\end{aligned}
$$

Es gilt starke Dualität hier. Also:

$$
\begin{aligned}
t^* = \min\ & \gamma \\
t:\ & \sum_{i=1}^{M} \lambda_i + \sum_{i=1}^{N} \mu_i = 1 \\
(a,\alpha):\ & \left( \sum_i -\lambda_i x_i + \sum_i \mu_i y_i \,,\, \gamma \right) \in \mathcal{L}^{n+1} \\
\beta:\ & \sum_i \lambda_i - \sum_i \mu_i = 0 \\
r:\ & \lambda_i \geq 0 \\
s:\ & \mu_i \geq 0
\end{aligned}
$$

$$
\begin{aligned}
= \min\ & \left\| \sum \lambda_i x_i - \sum \mu_i y_i \right\| \\
& \lambda, \mu \geq 0 \\
& e^T \lambda = e^T \mu = \tfrac{1}{2}.
\end{aligned}
$$

$$
= \frac{1}{2} \min \left\{ \|x - y\| : x \in \operatorname{conv} X,\ y \in \operatorname{conv} Y \right\}.
$$

$\boxtimes$