Support Vector Machines

André Hopfgartner & Matthias Rupp

08.06.2021

Vorarlberg University of Applied Sciences

FH Vorarlberg
University of Applied Sciences

Agenda

1. Einführung
2. Hard-Margin Support Vector Machine
3. Soft-Margin Support Vector Machine
4. Vergleich Hard- & Soft-Margin Support Vector Machine
5. Nichtlineare Trennung

Einführung

Intuition

Ziel: lineare Trennung zweier Klassen

Intuition

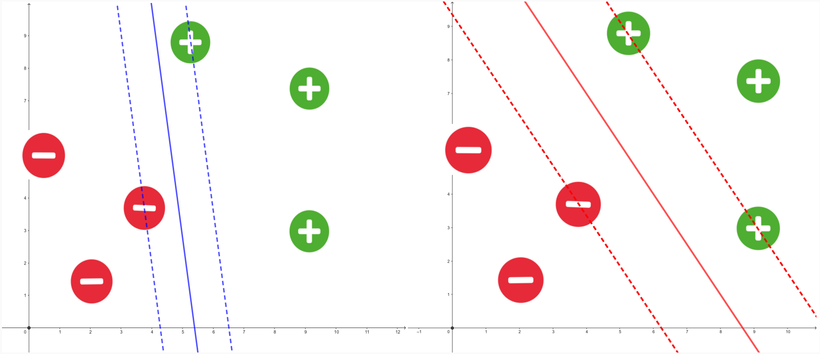
Ziel: lineare Trennung zweier Klassen
Wie?: Definition einer (Hyper-) Ebene

Intuition

Ziel: lineare Trennung zweier Klassen
Wie?: Definition einer (Hyper-) Ebene
Nebenbedingung: Möglichst großer freier Bereich

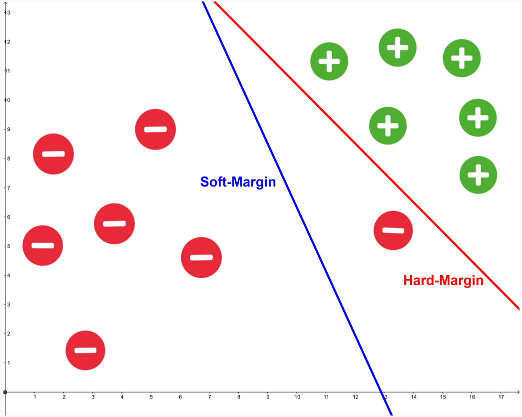
Intuition

Ziel: lineare Trennung zweier Klassen
Wie?: Definition einer (Hyper-) Ebene
Nebenbedingung: Möglichst großer freier Bereich



Arten von SVM:

- *Hard-Margin SVM*: Daten werden 100% korrekt getrennt
- *Soft-Margin SVM*: Einzelne Datenpunkte können falsch klassifiziert werden um insgesamt bessere Trennung zu erhalten



Hard-Margin Support Vector Machine

Gegeben sei ein Gewichtsvektor $w \in \mathbb{R}^K$, ein Bias $b \in \mathbb{R}$, ein beliebiger Punkt $x_n \in \mathbb{R}^K$ und ein zugehöriges Label $y_n \in \{-1, +1\}$. Eine Ebene im Raum kann allgemein definiert werden durch:

$$w^T x_n + b = 0 \tag{1}$$

Ziel der SVM: w und b bestimmen für optimale Trennung

Klassifikation

Annahme: w und b bereits bekannt

Wie klassifiziert man einen Punkt x_n?

Klassifikation

Annahme: w und b bereits bekannt

Wie klassifiziert man einen Punkt x_n?

Liegt x_n über oder unter Ebene = Vorzeichen:

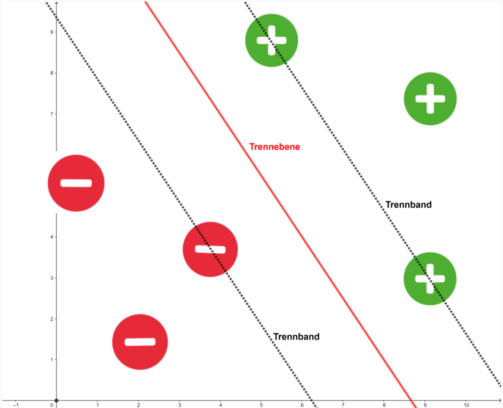
$$
\begin{aligned}
y = sign(w^T x_n + b) \qquad & \text{ist gleichbedeutend mit} \\
w^T x_n + b > 0 \qquad & \text{für } y_n = +1 \\
w^T x_n + b < 0 \qquad & \text{für } y_n = -1
\end{aligned}
$$

Bisher: Punkte können genau auf der Grenze liegen wenn $w^T x_n + b = 0$

Einführung eines Trennbandes

Striktere Regel: Um Ebene soll Band frei bleiben

$$\begin{aligned} w^T x_n + b &\geq +1 \qquad \text{für } y_n = +1 \\ w^T x_n + b &\leq -1 \qquad \text{für } y_n = -1 \end{aligned}$$



Einführung eines Trennbandes

Beidseitige Multiplikation mit y_n

$$
\begin{aligned}
y_n(w^T x_n + b) &\geq 1 \qquad \text{für } y_n = +1 \\
y_n(w^T x_n + b) &\geq 1 \qquad \text{für } y_n = -1
\end{aligned}
$$

Einführung eines Trennbandes

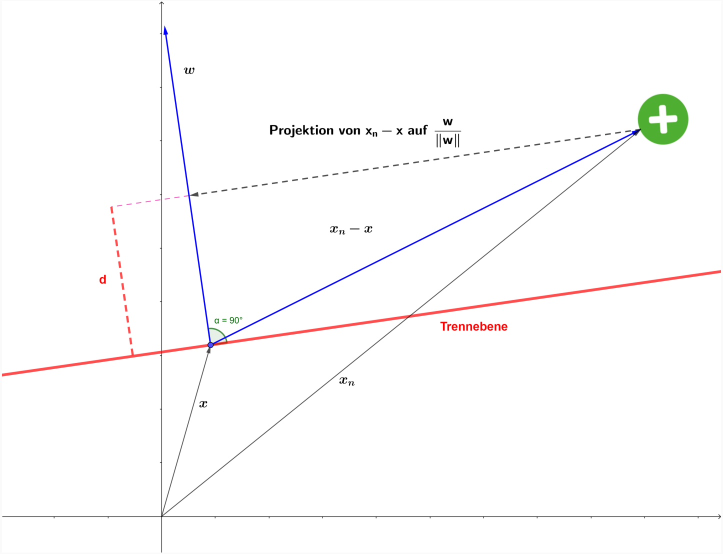
Beidseitige Multiplikation mit y_n

$$\begin{aligned} y_n(w^T x_n + b) &\geq 1 \qquad \text{für } y_n = +1 \\ y_n(w^T x_n + b) &\geq 1 \qquad \text{für } y_n = -1 \end{aligned}$$

Für den Fall, dass $x_n = \hat{x}$ genau an der Grenze des Trennbands liegt, gilt somit:

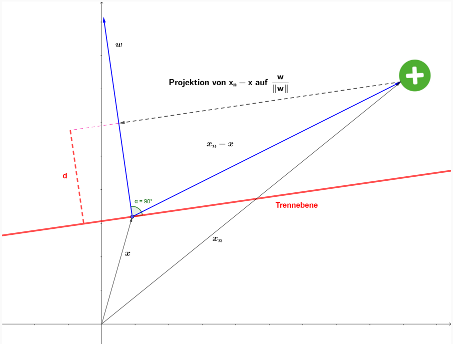
$$y_n(w^T \hat{x} + b) = 1 \tag{5}$$

Gesucht: Normalabstand d eines Punktes $x_n \in \mathbb{R}^K$ zur Ebene



Normalabstand eines Punktes zur Ebene

$$
\begin{aligned}
d = |\frac{w^T}{\|w\|}(x_n - x)| &= \\
= \frac{1}{\|w\|}|(w^T x_n - w^T x)| &= \\
= \frac{1}{\|w\|}|(w^T x_n + b - (w^T x + b))|
\end{aligned}
$$



Normalabstand eines Punktes zur Ebene

$$d = \frac{1}{\|w\|}|(w^T x_n + b - (w^T x + b))|$$

Weil der Punkt x auf der Ebene liegt gilt $w^T x + b = 0$ und somit für den Normalabstand eines beliebigen Punktes x_n:

$$d = \frac{1}{\|w\|}|(w^T x_n + b)|$$

Breite des Trennbands

$$d = \frac{1}{\|w\|} |(w^T x_n + b)|$$

Annahme: $x_n = \hat{x}$ ist der am nächsten zur Ebene liegende Punkt auf der Grenze des Trennbands
Weil $y_n(w^T \hat{x} + b) = 1 = |w^T \hat{x} + b|$ gilt ergibt sich der minimale Normalabstand D:

$$D = \frac{1}{\|w\|}$$

Weil D der minimale Normalabstand zur Ebene ist, ist $2D$ die Breite des freien Trennbands.

A test with images

- Some
- text
- on left side of slide here..
- Abb. 1 zeigt blabla.

A test with images

- Some
- text
- on left side of slide here..
- Abb. 1 zeigt blabla.

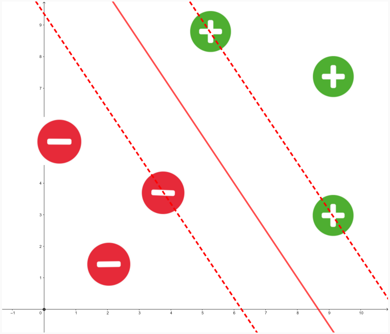


Abbildung 1: Abhängig von der Lage der Trennebene entstehen schmale (blau) oder breite (rot) Trennbänder. Ziel ist die Maximierung der Breite des Trennbands durch die Ermittlung der optimalen Lage der Trennebene.

citation tests

$$
\begin{aligned}
y = sign(w^T x + b) & \qquad \text{gleichbedeutend mit} && (6a) \\
w^T x + b > 0 & \qquad \text{für } y = +1 && (6b) \\
w^T x + b < 0 & \qquad \text{für } y = -1 && (6c)
\end{aligned}
$$

In Gleichung (6) wird ..

Footcite example[1]

Burges (1998)

[1] Platt 1998.

Soft-Margin Support Vector Machine

Vergleich Hard- & Soft-Margin Support Vector Machine

Nichtlineare Trennung

Fragen?

Literatur

Burges, Christopher J.C. (1. Juni 1998). „A Tutorial on Support Vector Machines for Pattern Recognition“. In: *Data Mining and Knowledge Discovery* 2.2, S. 121–167. ISSN: 1573-756X. DOI: 10.1023/A:1009715923555. URL: https://doi.org/10.1023/A:1009715923555 (besucht am 06. 03. 2021).

Platt, John (Apr. 1998). *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*. MSR-TR-98-14, S. 21. URL: https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines/.