

Support vector machines mit Kernel Trick

Jonas Klug, Hendrik Siek und Tim Schlottmann

TU Hamburg

29. Januar 2018

1 Abstract

2 Motivation

3 Support vector machines – Grundlagen

Support Vector Maschinen sind ein binärer Klassifizierer. Datenmengen werden also in zwei Klassen eingeteilt. Zur Unterteilung der Klassen wird eine Hyperebene benutzt. Wie die Unterteilung stattfindet soll in diesem Abschnitt erklärt werden.

3.1 Definitionen

Folgende Definitionen bilden die Grundlage für SVMs:

- Anzahl an Datenpunkten: $m \in \mathbb{R}$
- Input: $\mathbf{x} \in \mathbb{R}^N$
- Output: $y \in \{-1, +1\}$
- Trainingsset: $S \in (\mathbb{R}^N \times \{-1, +1\})^m$

Ziel ist es, eine Hypothese zu finden, die einen Input \mathbf{x} auf eine der beiden Klassen y abbildet:

$$h: \mathbb{R}^N \rightarrow \{+1, -1\}$$

$$\mathbf{x} \mapsto y$$

3.2 Hyperebene

Betrachten wir ein Trainingsset S mit eingezeichneter Hyperebene H wie in Abbildung 1. Es fällt auf, dass ein Spalt (engl. margin) entsteht, dessen Grenzen sich durch zwei weitere Hilfsebenen H_+ und H_- darstellen lassen. H_+ und H_- sind hierbei parallel zu H . In Abbildung 1 ist H_+ die Begrenzung zur $+1$ -Klasse und H_- die Begrenzung zur -1 -Klasse. Die Vektoren, durch die die Hilfsebenen verlaufen, heißen Supportvektoren und sind in der Abbildung schwarz umrandet.

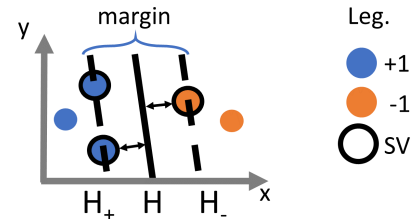


Abbildung 1: Hyperebene H zur Trennung der beiden Klassen $+1$ und -1

Nun gilt es die Hypothese zu finden. Hierfür betrachten wir zuerst die Hyperbenengleichung $H = \mathbf{w}^T \mathbf{x} + b = 0$. Hierbei werden die Parameter \mathbf{w} und b so gewählt, dass für die Supportvektoren gilt: $y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$. So lässt sich die Hypothese wie folgt definieren:

$$h(\mathbf{x}_i) = \begin{cases} +1 & \text{wenn } \mathbf{w}^T \mathbf{x}_i + b \geq 0 \\ -1 & \text{wenn } \mathbf{w}^T \mathbf{x}_i + b \leq 0 \end{cases}$$

3.3 Minimierungsproblem

Die algorithmische Bestimmung der Parameter \mathbf{w} und b erfolgt über ein Minimierungsproblem. Die Spalte zwischen den Hilfsebenen H_+ und H_- soll maximiert werden. Über die Projektionseigenschaft des Skalarproduktes lässt sich die Breite der Spalte über den Ausdruck $\frac{2}{\|\mathbf{w}\|}$ bestimmen. Die Breite der Spalte soll nun maximiert werden, was zu folgendem Minimierungsproblem führt:

$$\max_{\mathbf{w}, b} \frac{2}{\|\mathbf{w}\|} \Leftrightarrow \min_{\mathbf{w}, b} \|\mathbf{w}\| \Leftrightarrow \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2$$

$$\text{u.d.N. } y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 \geq 0$$

Durch die Verwendung von Lagrange Multiplikatoren lässt sich das Minimierungsproblem in eine Form bringen, die nur noch von der Lagrangevariable α sowie den Supportvektoren bzw. den neu zu klassifizierenden Vektoren abhängt:

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$$

$$\sum_i \alpha_i y_i = 0$$

$$h(\mathbf{x}) = \begin{cases} +1 & \sum_i \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \geq 0 \\ -1 & \sum_i \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \leq 0 \end{cases}$$

4 Kernel Trick – Jonas

5 Praktische Umsetzung – Hendrik