FH Vorarlberg
University of Applied Sciences

Support Vector Machines (SVM)

Computational Intelligence II

Informatik - Software and Information Engineering
Fachhochschule Vorarlberg

Erstellt von
André Hopfgartner & Matthias Rupp

Dornbirn, am 1. März 2021

Inhaltsverzeichnis

Abkürzungsverzeichnis

SVM Support Vector Machine

1 Einführung

1.1 Intuition

Ziel: möglichst breites Band zwischen den 2 verschiedenen Klassen aufziehen.

1.2 Mathematische Herleitung

Gegeben sei ein Gewichtsvektor $\vec{w} \in \mathbb{R}^M$ und ein beliebiger Vektor $\vec{u} \in \mathbb{R}^M$. Der Gewichtsvektor $\vec{w}$ spannt mit dem Bias $b \in \mathbb{R}$ eine Ebene im Raum auf. Um entscheiden zu können, ob ein Punkt $\vec{u}$ über oder unter der Trennebene liegt, kann folgende Entscheidungsregel eingeführt werden:

$$\vec{w} \cdot \vec{u} + b \geq 0 \tag{1.1}$$

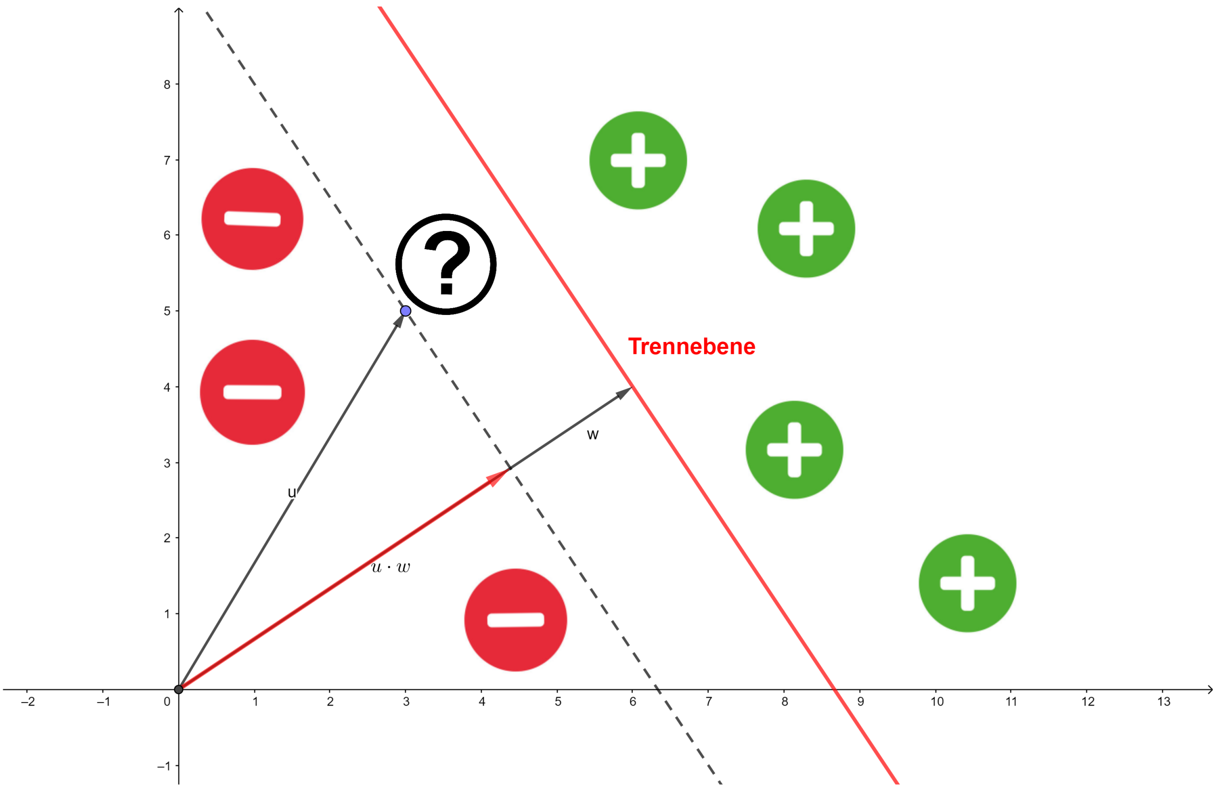

Abbildung 1.1: Plot of Golden section with quadratic function

Als nächstes führen wir ein Band um die Trennebene ein. Gegeben sei eine Trainingsmenge von Tupeln $(\vec{x_i}, d_i)$ mit $\vec{x_i} \in \mathbb{R}^M$ und $d_i \in \{-1, +1\}$. Bei $\vec{x_i}$ handelt es sich um einen Eingabevektor, d_i ist das jeweils zugehörige Label. Wir führen folgende Regeln ein:

$$\begin{aligned} \vec{w} \cdot \vec{x_i} + b \geq +1 \textbf{ für } d_i = +1 \\ \vec{w} \cdot \vec{x_i} + b \leq -1 \textbf{ für } d_i = -1 \end{aligned} \tag{1.2}$$

Gleichung 1.2 kann weiter verallgemeinert werden, indem auf beiden Seiten mit d_i multipliziert wird. Weil für jedes beliebige d_i gilt $d_i d_i = 1$ ergibt sich:

$$\begin{aligned} d_i(\vec{w} \cdot \vec{x_i} + b) \geq 1 \textbf{ für } d_i = +1 \\ d_i(\vec{w} \cdot \vec{x_i} + b) \geq 1 \textbf{ für } d_i = -1 \end{aligned} \tag{1.3}$$

Für Punkte $\vec{x_i}$, die genau auf den Grenzen des Bandes liegen, muss also gelten:

$$d_i(\vec{w} \cdot \vec{x_i} + b) - 1 = 0 \tag{1.4}$$

Die Breite ξ des Bandes lässt sich berechnen, indem zwei Punkte $\vec{x_+}$ und $\vec{x_-}$ gewählt werden, die jeweils auf den Grenzen des Bandes liegen. Der Differenzvektor $\vec{x_+}$ - $\vec{x_-}$ kann anschließend auf den Einheitsvektor $\vec{w}^0$ projiziert werden mittels des Skalarprodukts. Als Ergebnis erhält man den Normalabstand zwischen den Grenzen des Bandes, was genau der Breite des Bandes entspricht:

$$\begin{aligned} \xi &= (\vec{x_+} - \vec{x_-}) \cdot \frac{\vec{w}}{\|\vec{w}\|} \\ &= \frac{1}{\|\vec{w}\|} \cdot (\vec{x_+} \cdot \vec{w} - \vec{x_-} \cdot \vec{w}) \end{aligned} \tag{1.5}$$

Setzt man in Gleichung 1.4 die Tupel $(\vec{x_+}, +1)$ und $(\vec{x_-}, -1)$ ein erhält man:

$$\begin{aligned} \vec{w} \cdot \vec{x_+} = 1 - b \textbf{ für } d_i = +1 \\ \vec{w} \cdot \vec{x_-} = -b - 1 \textbf{ für } d_i = -1 \end{aligned} \tag{1.6}$$

Damit ergibt sich für Gleichung 1.5:

$$\begin{aligned} \xi &= \frac{1}{\|\vec{w}\|} \cdot ((1 - b) - (-b - 1)) \\ &= \frac{2}{\|\vec{w}\|} \end{aligned} \tag{1.7}$$

Weil wir bei Support Vector Machine (SVM) ein möglichst breites Band zwischen den Trainingsdaten haben möchten führen wir eine Maximierung der Breite durch, die sich auch als Minimierung darstellen lässt:

$$\hat{\xi} = \max \frac{2}{\|\vec{w}\|} = \min \|\vec{w}\| \quad = \frac{1}{2} \min \|\vec{w}\|^2 \tag{1.8}$$

Die in Gleichung 1.8 beschriebene Umformungen wurden gemacht, weil dadurch die später zu optimierende Gleichung besser darstellbar ist.

Um Gleichung 1.8 optimieren zu können wird die Lagrangegleichung mit der in Gleichung 1.4 beschriebenen Nebenbedingung aufgestellt:

$$L = \frac{1}{2}\|\vec{w}\|^2 - \sum(\alpha_i(d_i(\vec{w} \cdot \vec{x_i} + b) - 1)) \tag{1.9}$$

Um den Extremwert einer Lagrangegleichung ermitteln zu können müssen alle partiellen Ableitungen 0 gesetzt werden:

$$\begin{aligned} \frac{\partial L}{\partial \vec{w}} = \vec{w} - \sum(\alpha_i d_i \vec{x_i}) &\stackrel{!}{=} 0 \\ \vec{w} &= \sum(\alpha_i d_i \vec{x_i}) \end{aligned} \tag{1.10}$$

$$\begin{aligned} \frac{\partial L}{\partial b} = -\sum(\alpha_i d_i) &\stackrel{!}{=} 0 \\ \sum(\alpha_i d_i) &= 0 \end{aligned} \tag{1.11}$$

Die Ergebnisse von Gleichung 1.10 und Gleichung 1.11 können in Gleichung 1.9 wieder eingesetzt werden:

$$\begin{aligned} L &= \frac{1}{2}(\sum(\alpha_i d_i \vec{x_i})) \cdot (\sum(\alpha_j d_j \vec{x_j})) - \sum(\alpha_i d_i x_i \cdot (\sum(\alpha_j d_j \vec{x_j})) - \sum(\alpha_i d_i b) + \sum(\alpha_i) = \\ &= \frac{1}{2}(\sum(\alpha_i d_i \vec{x_i})) \cdot (\sum(\alpha_j d_j \vec{x_j})) - \sum(\alpha_i d_i x_i \cdot (\sum(\alpha_j d_j \vec{x_j})) - b\sum(\alpha_i d_i) + \sum(\alpha_i) \end{aligned} \tag{1.12}$$

Aus Gleichung 1.11 gilt $\sum(\alpha_i d_i) = 0$:

$$L = \frac{1}{2}(\sum(\alpha_i d_i \vec{x_i})) \cdot (\sum(\alpha_j d_j \vec{x_j})) - \sum(\alpha_i d_i x_i \cdot (\sum(\alpha_j d_j \vec{x_j})) + \sum(\alpha_i) \tag{1.13}$$

Bei den Summen ergibt sich eine Symmetrie der Form $\frac{1}{2}a - a$ wobei a als Abkürzung für die Summen verstanden werden kann:

$$L = \frac{1}{2}(\sum\sum(\alpha_i \alpha_j d_i d_j \vec{x_i} \cdot \vec{x_j})) + \sum(\alpha_i) \tag{1.14}$$

Das bemerkenswerte an Gleichung 1.14 ist, dass die Optimierung nur von dem Skalarprodukt von je 2 Samples abhängt.

Weiters lässt sich die Entscheidungsregel Gleichung 1.15 mit Gleichung 1.10 umformen:

$$(\sum(\alpha_i d_i \vec{x_i})) \cdot \vec{u} + b \geq 0 \tag{1.15}$$

Somit hängt auch die Entscheidungsregel nur von den Skalarprodukten ab.