

Bei der Separierung von zwei Datenklassen scheint eine lineare Trennlinie zwischen die Klassen zu ziehen eine simple und intuitive Methode zu sein. Genau auf dieser Idee basieren Support Vector Machines (SVM). Jedoch sind Daten nicht immer perfekt trennbar oder es liegt keine lineare Trennlinie vor. Wie SVM funktionieren, wie sie mit solchen Situationen umgehen, wie sie unter vorgegebenen Bedingungen performen und welche Schlussfolgerungen für die praktische Anwendung gezogen werden können, soll in dieser Arbeit beleuchtet werden.

Die Grundlage für eine optimal separierende Hyperebene wurde bereits 1964 von Alexej Chervonenkis und Vladimir Vapnik gelegt und die Methode wurde anschließend von mehreren Autoren stetig erweitert (Vapnik, 2006). Seitdem ist ihre Leistung, trotz großer Fortschritte im Bereich neuronaler Netze, vor allem im Bereich binärer Klassifikationsprobleme unumstritten und sie gelten dabei als eine der meistgenutzten Methoden (Soofi & Awan, 2017). Dabei bieten SVM eine große Anzahl an Anwendungsmöglichkeiten, wie beispielsweise, in der Bildklassifikation, in der Bioinformatik (Krebsklassifikation) oder im Aufdecken von Kreditkartenbetrug (Cervantes et al., 2020).

In dieser Arbeit wollen wir zuerst die Funktionsweise der SVM näher betrachten. Dazu wird zunächst gezeigt wie die Konstruktion dieser separierenden Hypereben im Falle linear trennbarer Daten funktioniert. Daraufhin wird die Idee eines Soft Margin Classifier aufgegriffen, der sich dem Problem nicht trennbarer Daten annimmt. Zuletzt geht es um die Anwendung von Kernels, die es ermöglichen, nicht lineare Entscheidungsgrenzen herzustellen.

In Kapitel 3 werden dann die Vor- und Nachteile der Methode betrachtet. Anhand dieser Informationen werden Schlussfolgerungen zur Leistung der SVM gezogen. Darauf aufbauend wird im folgenden Kapitel der Aufbau unseres Experiments beschrieben. Wir generieren uns zum anschließenden Vergleich neun verschiedene Datenszenarien, welche in ihrer Dimensionalität und Komplexität der Entscheidungsgrenze variieren. Für die Evaluierung der Klassifikationsleistung werden neben den SVM weitere Klassifikationsmethoden herangezogen.

Im Kapitel Hypothesen wurden auf Basis von Literatur, in der bereits ähnliche Versuche durchgeführt wurden, Vermutungen aufgestellt, wie die einzelnen Klassifikationsmethoden im Vergleich abschneiden werden. Es folgen die Ergebnisse, in denen die Durchführung des Experiments beschrieben und die Leistungen anhand verschiedener Maßzahlen evaluiert werden. Darüber hinaus wird überprüft, ob sich die Hypothesen die wir zuvor aufgestellt haben bewahrheitet haben. Im letzten Abschnitt wird ein Fazit gezogen, indem die Ergebnisse diskutiert und Weiterentwicklungsmöglichkeiten für unsere Arbeit aufgearbeitet werden.

Literatur

- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A Comprehensive Survey on Support Vector Machine Classification: Applications, Challenges and Trends. *Neurocomputing*, 408, 189–215. <https://doi.org/10.1016/j.neucom.2019.10.118>
- Soofi, A. A., & Awan, A. (2017). Classification Techniques in Machine Learning: Applications and Issues. *Journal of Basic & Applied Sciences*, 13, 459–465.
- Vapnik, V. (2006). *Estimation of Dependences Based on Empirical Data*. Springer. <https://doi.org/10.1007/0-387-34239-7>