

Warum eine gerade Ebene manchmal die beste Möglichkeit ist, um Daten in zwei Klassen aufzuteilen? Darum soll es in dieser Arbeit gehen. Wir wollen hier die Methode der Support Vector Machines, die mit dieser Annahme arbeitet, erläutern und gleichzeitig auf den Prüfstand stellen. Denn auch wenn die Grundidee, vielleicht etwas zu simpel klingt, ist ihre Leistung was vor allem binäre Klassifikationsprobleme angeht unumstritten. 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). Zuerst wollen wir die Funktionsweise der SVM näher beleuchten. Dazu wird zuerst gezeigt wie die Konstruktion dieser separierenden Hyperebenen in dem Fall funktioniert, in dem die Daten tatsächlich linear trennbar sind. Danach soll gezeigt werden wie eine Soft Margin Klassifizier funktioniert, der auch funktioniert bei nicht perfekt linear trennbaren Daten und zuletzt soll es um die Anwendung von Kernels gehen, die es dann auch ermöglichen, nicht lineare Entscheidungsgrenzen herzustellen. Im Kapitel 3 wollen wir dann die Vor und Nachteile der Methode uns anschauen. Anhand dieser Informationen konnten wir uns dann bereits einige Gedanken zur Leistung der SVM in unserem Datenexperiment machen. Wie wir dieses Experiment aufbauen soll im nachfolgenden Kapitel beschrieben werden. Wir generieren uns zum anschließenden Vergleich neun verschiedene Datenszenarien, welche in ihrer Dimensionalität und Komplexität der Entscheidungsgrenze variieren. Für den Vergleich wollen wir dann Klassifikationsleistung von SVM und anderen Klassifikationsmethoden vergleichen. Im Teil Hypothesen haben wir aufgrund von Literatur, die bereits ähnliche Versuche durchgeführt haben, Vermutungen aufgestellt, wie die einzelnen Classifier im Vergleich abschneiden werden. Im Ergebnis-Teil wollen wir dann das Experiment durchführen und mithilfe von verschiedenen Maßzahlen die Leistung evaluieren. Ebenfalls wollen wir überprüfen, ob die Hypothesen die wir zuvor aufgestellt haben sich bewahrheiten. Im letzten Teil soll ein Fazit gezogen werden, in dem wir die Ergebnisse noch einmal diskutieren und Defizite in unserer Arbeit aufarbeiten.

Literatur

Vapnik, V. (2006). *Estimation of Dependences Based on Empirical Data*. Springer. <https://doi.org/10.1007/0-387-34239-7>