

Zusammenfassung

Eine Modifizierung der Lernmethode der Support Vector Machines (SVMs) ist die Lernmethode der lokalisierten SVMs, bei welcher SVMs auf jeder Zelle einer Partition des Eingaberaums gelernt werden. Neben dem verbesserten Rechenaufwand dieser lokal lernenden Verfahren ist es von entscheidendem Vorteil, dass auf jeder betrachteten Zelle unterschiedliche Kern- und Regularisierungsparameter gewählt werden können. In der vorliegenden Arbeit nutzen wir diese Eigenschaft, um globale Klassifikations-Lernraten für lokalisierte SVMs mit Gaußkernen und der hinge-Verlustfunktion zu erhalten. Die erzielten Raten sind unter geeigneten Voraussetzungen besser als die bereits bekannten Klassifikationsraten für lokalisierte SVMs, globale SVMs oder andere betrachtete Klassifikationsverfahren. Dabei setzen wir eine schwache geometrische Bedingung an die Entscheidungslinie voraus und treffen übliche Annahmen über das Verhalten der Daten-erzeugenden Verteilung in der Nähe der Entscheidungslinie. Die Parameter, die letzteres beschreiben, spiegeln sich in den Parametern der lokalisierten SVMs, sowie in der Lernrate wieder und wir zeigen, dass eine Trainings- und Validierungsmethode adaptiv die gleichen Raten erzielt.

Die statistische Analyse beruht auf einer einfachen Technik, die den Eingaberaum in sich zwei überlappende Mengen partitioniert. Dabei wird einmal die Menge betrachtet, deren Zellen nahe zur Entscheidungslinie liegen, sowie die Menge, deren Zellen einen ausreichenden Abstand zur Entscheidungslinie besitzen. Das Überschussrisiko wird separat auf diesen Mengen, deren Trennung durch einen Parameter $s > 0$ beschrieben wird, abgeschätzt. Um diese Analyse und ihre Einflüsse zu verdeutlichen, wenden wir die Technik zunächst auf die einfache Lernmethode der Histogramm-Regel an und erhalten unter geeigneten Annahmen Lernraten, die sogar besser sind als die der komplexen SVMs. Für die lokalisierten SVMs zeigen die separaten Analysen, welchen Einfluss die unterschiedlichen Kern- und Regularisierungsparameter, sowie die Parameter welche die Verteilung beschreiben, auf das lokale Lernverhalten besitzen. Eine geeignete Wahl des Trennungsparmeters s führt schließlich zu einer globalen Lernrate für lokalisierte SVMs mit Gaußkernen und der hinge-Verlustfunktion.

Abstract:

One of the main characteristics of localized support vector machines, which solve SVMs on many spatially defined small chunks is, besides the computational benefit compared to global SVMs, the freedom of choosing arbitrary kernel and regularization parameters on each cell. In the present work, we take advantage of this observation to derive global learning rates for localized SVMs with Gaussian kernels and hinge loss for classification. The rates we obtain outperform known classification rates for localized SVMs, for global SVMs, and other learning algorithms under a suitable set of assumptions. They are achieved under a mild geometric condition on the decision boundary and under a set of margin conditions that describe the behavior of the data-generating distribution near the decision boundary, where no assumption on the existence of a density is made. It turns out that a margin condition that relates the location of noise to the decision boundary is crucial to obtain improved rates. These margin parameters appear in the chosen parameters for localized SVMs and in the learning rates. Nevertheless, we show that a training validation procedure learns with the same rates adaptively, that is, without knowing these rates.

The statistical analysis relies on a simple partitioning based technique, which analyzes the excess risk separately on sets that are close to the decision boundary and on those, which are sufficiently far away. These sets depend on a splitting parameter $s > 0$. To illustrate and to understand the mechanisms of that technique we first apply it to the simple histogram rule and derive even for this simple method learning rates, which outperform rates for global SVMs under suitable assumptions.

For localized SVMs with Gaussian kernel and hinge loss, we are able to derive local learning rates that demonstrates how kernel and regularization parameter, as well as different margin parameter affect the learning rates on the considered sets. For an appropriately chosen splitting parameter, we finally derive global learning rates for localized SVMs with Gaussian kernels and hinge loss.