Learning with High-Dimensional Data

Abstract: This thesis is divided into three parts. In the first part we introduce a framework that allows us to investigate learning scenarios with restricted access to the data. We use this framework to model high-dimensional learning scenarios as an infinite-dimensional one in which the learning algorithm has only access to some finite-dimensional projections of the data. Finally, we provide a prototypical example of such an infinite-dimensional classification problem in which histograms can achieve polynomial learning rates.

In the second part we present some individual results that might by useful for the investigation of kernel-based learning methods using Gaussian kernels in high- or infinite-dimensional learning problems. To be more precise, we present log-covering number bounds for Gaussian reproducing kernel Hilbert spaces on general bounded subsets of the Euclidean space. Unlike previous results in this direction we focus on small explicit constants and their dependence on crucial parameters such as the kernel width as well as the size and dimension of the underlying space. Afterwards, we generalize these bounds to Gaussian kernels defined on special infinite-dimensional compact subsets of the sequence space ℓ_2. More precisely, the considered domains are given by the image of the unit ℓ_∞-ball under some diagonal operator.

In the third part we contribute some new insights to the compactness properties of diagonal operators from ℓ_p to ℓ_q for $p \neq q$.

Lernen mit hochdimensionalen Daten

Zusammenfassung: Diese Arbeit ist in drei Teile geteilt. Im ersten Teil stellen wir ein Framework vor, das es uns erlaubt Lernszenarien mit eingeschränktem Zugriff auf die Daten zu untersuchen. Wir verwenden dieses Framework um hochdimensionale Lernszenarien als unendlich dimensionale Szenarien zu modellieren, bei denen der Lernalgorithmus nur Zugang zu endlich dimensionalen Projektionen der Daten hat. Abschließend stellen wir ein Modellbeispiel für ein solches unendlich dimensionales Klassifikationsproblem vor, in dem Histogramme polynomielle Lernraten erreichen können.

Im zweiten Teil stellen wir Ergebnisse vor, die zur Untersuchung von kernbasierten Lernmethoden in hoch- oder unendlich dimensionalen Lernszenarien nützlich sein könnten. Genauer präsentieren wir Schranken an die log-Überdeckungszahlen für Gauß'sche reproduzierende Kern-Hilberträume auf allgemeinen beschränkten Teilmengen des euklidischen Raums. Im Gegensatz zu bisherigen Ergebnissen in diesem Gebiet konzentrieren wir uns auf kleine explizite Konstanten in den Abschätzungen und deren Abhängigkeit von wichtigen Parametern wie der Kernweite sowie der Größe und Dimension des zugrunde liegenden Raums. Anschließend verallgemeinern wir diese Schranken auf Gaußkerne, die auf speziellen unendlich dimensionalen kompakten Teilmengen des Folgenraums ℓ_2 definiert sind. Genauer gesagt betrachten wir Gebiete, die als das Bild der ℓ_∞-Einheitskugel unter einem Diagonaloperator gegeben sind.

Im dritten Teil liefern wir einen Beitrag zu den Kompaktheitseigenschaften von Diagonaloperatoren von ℓ_p nach ℓ_q für $p \neq q$.