On low dimensional models for functions in high dimensions

Author(s):
Tyagi, Hemant

Publication Date:
2016

Permanent Link:
https://doi.org/10.3929/ethz-a-010666821

DISS. ETH NO. 23395

On low dimensional models for functions in high dimensions

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES OF ETH ZURICH
(DR. SC. ETH ZURICH)

presented by

Hemant Tyagi

MSc EPFL Communication Systems
born on 4 September 1983
citizen of India

accepted on the recommendation of

Prof. Dr. Bernd Gärtner, examiner
Prof. Dr. Andreas Krause, co-examiner
Prof. Dr. Jared Tanner, co-examiner

2016

Abstract

Many problems in science and engineering involve an underlying function that is high dimensional, i.e., a function depending on d variables, with d typically large. In this thesis, we study two problems involving such functions namely: continuum armed bandits and function approximation. Provided only classical smoothness assumptions are made on the functions, both these problems are known to suffer from the so called curse of dimensionality. We tackle these problems in the setting where the underlying functions additionally possess an unknown low dimensional structure with $k \ll d$ degrees of freedom.

In the first part of the thesis, we study the continuum armed bandit problem. This is a type of online optimization problem where an algorithm chooses a strategy from a compact subset of $\mathbb{R}^d$, over a sequence of consecutive rounds. Associated to each strategy is a reward that corresponds to the evaluation of a reward function defined over the strategy set, and which can potentially change across rounds. The performance of the algorithm is measured in terms of the regret defined as the difference between the total reward of the best fixed strategy and that obtained by the player. Assuming the reward functions to be merely smooth implies that the (worst case) per round regret for any algorithm converges to zero at a rate at least exponentially slow in d. We first consider the setting where the reward functions depend on an unknown, fixed subset, of the coordinate variables, of size $k \ll d$. We derive a simple algorithm that achieves nearly optimal regret bounds, with the rate depending only on k and not d. This is also extended to the scenario where the subset of active coordinate variables can possibly change across time. The algorithm is anytime i.e., does not require to know the number of rounds in advance and can handle both stochastic, as well as adversarial choices of reward functions. Next, we consider a generalization of this setting wherein the reward functions are ridge functions, i.e., effectively depend on an unknown k-dimensional subspace of $\mathbb{R}^d$. Assuming the number of rounds to be known in advance and the reward functions to be generated stochastically, we derive an algorithm with nearly

optimal regret bounds. As before, the regret rate depends only on k and not d.

In the second part of the thesis, we study the problem of approximating an unknown d-variate function from its samples, over some known compact subset of $\mathbb{R}^d$. This problem has been studied extensively in: *statistics* (non-parametric regression) where the samples are typically noisy and assumed to be generated beforehand, and *approximation theory* where one assumes freedom to sample the function, at any location within its domain. It is well known that if mere smoothness assumptions are made on the function, then in the worst case, any algorithm needs at least exponential in d many samples to suitably approximate the function. This is usually avoided by assuming the function to be intrinsically low dimensional. We study one such class of functions referred to as sparse additive models (SPAMs), whererin the function is the sum of a sparse number of univariate components, each depending on some unknown coordinate variable. To this end, we derive a randomized sampling scheme, and an efficient algorithm, that exactly recovers the unknown support as well as the underlying univariates. Specifically, we derive non-asymptotic sampling bounds along with uniform approximation guarantees for the components. The results are also extended to the setting where the samples are corrupted with either arbitrary bounded noise, or stochastic noise. Next, we consider a generalized SPAM which also allows for the presence of a sparse number of pairwise *interaction terms*, represented by bivariate functions. As before, we derive an efficient scheme along with non-asymptotic sampling bounds for exact recovery of the univariate, and bivariate supports. Furthermore, the results are also extended to the aforementioned noisy settings.

Zusammenfassung

Vielen Fragestellungen in Wissenschaft und Technik liegt eine hochdimensionale Funktion zugrunde, d. h. eine Funktion in d Variablen, wobei d sehr gross ist. In dieser Arbeit behandeln wir zwei solche Fragestellungen: das *continuum-armed-bandits*-Problem und die Approximation von Funktionen. Beschränkt man sich auf klassische Glattheitsannahmen, so leiden beide Fragestellungen unter dem sogenannten Fluch der Dimensionalität. Hier machen wir die Annahme, dass die zugrundeliegenden Funktionen eine niedrigdimensionale, aber unbekannte, Struktur mit $k \ll d$ Freiheitsgraden besitzen.

Im ersten Teil der Arbeit behandlen wir das *continuum-armed-bandits*-Problem. Hierbei handelt es sich um ein Online-Optimierungsproblem, worin der Algorithmus in jeder Runde eine Strategie aus einer kompakten Teilmenge des $\mathbb{R}^d$ auswählt. Eine Gewinnfunktion, welche sich von Runde zu Runde ändern kann, ordnet jeder Strategie einen Gewinn zu. Die Güte des Algorithmus wird durch den *Regret* gemessen, d. i. die Differenz des erreichten Gewinns zu dem mit einer festen Strategie optimal möglichen Gewinn. Nimmt man die Gewinnfunktionen lediglich als glatt an, so wird der worst-case Regret je Runde für jeden beliebigen Algorithmus gegen Null konvergieren, wobei die Konvergenzrate exponentiell von d abhängt. Wir treffen nun zunächst die Annahme, dass die Gewinnfunktionen nur von einer unbekannten, aber festen, Teilmenge der Variablen abhängt; k ist hier die Grösse dieser Menge von „aktiven" Variablen. Wir beschreiben einen einfachen Algorithmus mit fast optimalen Regret-Schranken und mit einer Konvergenzrate, die nur von k, nicht aber von d abhängt. Auch erweitern wir den Algorithmus, so dass sich die aktiven Variablen von Runde zu Runde ändern dürfen. Der Algorithmus ist *anytime*, d. h. die Anzahl der Runden muss nicht im Vorhinein bekannt sein, und sowohl stochastisch als auch von einem Gegenspieler bestimmte Gewinnfunktionen werden berücksichtigt. Dann betrachten wir eine allgemeinere Aufgabenstellung, worin die Gewinnfunktionen Ridge-Funktionen sind, d. h. sie hängen nur von einem (unbekannten) k-dimensionalen Teilraum des $\mathbb{R}^d$ ab. Wir beschreiben

einen Algorithmus mit fast optimalen Regret-Schranken, wenn die Rundenzahl im Vorhinein bekannt ist und die Gewinnfunktionen zufällig erzeugt werden. Wie vorher hängt die Konvergenzrate nur von k und nicht von d ab.

Im zweiten Teil der Arbeit behandeln wir die Aufgabe, eine unbekannte Funktion in d Variablen näherungsweise aus Stichprobenwerten zu rekonstruieren, die aus einer bekannten kompakten Teilmenge des $\mathbb{R}^d$ stammen. Dieses Problem gehört einerseits zur Statistik (nichtparametrische Regression: hier sind die Stichprobenwerte gewöhnlich fehlerbehaftet und a propri gegeben) und andererseits zur Approximationstheorie (hier darf der Algorithmus die Stichproben an beliebigen Stellen selbst nehmen). Nimmt man die Funktion lediglich als glatt an, so ist bekannt, dass die für eine genügende Approximation schlimmstenfalls nötige Anzahl Stichproben exponentiell von d abhängt. Man umgeht dieses Problem, indem man wiederum annimmt, dass die Funktion ein verborgenes niedrigdimensionales Verhalten aufweist. Wir behandeln hier speziell *sparse additive models* (SPAMs); die Funktionen lassen sich in diesem Modell als Summe weniger univariater Funktionen schreiben, wobei die beteiligten Variablen unbekannt sind. Wir beschreiben einen randomisierten Stichprobenplan und einen effizienten Algorithmus, welche die beteiligten Variablen bestimmen sowie die univariaten Funktionen approximieren. Für die Anzahl der Stichproben geben wir eine nicht-asymptotische Schranke, für die Güte der Approximationen eine gleichmässige Schranke. Überdies erweitern wir unsere Ergebnisse auf fehlerbehaftete Stichprobenwerte (entweder beliebige beschränkte Fehler, oder stochastische Fehler). Weiters betrachten wir ein verallgemeinertes SPAM, welches durch bivariate Funktionen modellierte paarweise Korrelationen berücksichtigen kann. Für dieses Modell beschreiben wir einen effizienten Algorithmus mit einer nicht-asymptotischen Schranke für die Stichprobenzahl, welcher die univariaten Funktionen sowie die korrelierten Paare exakt berechnet. Wiederum verallgemeinern wir unsere Ergebnisse für die oben beschriebenen Fehlermodelle.