

A Hierarchical Bayesian Approach to Regression and its Application to Predicting Survival Times in Cancer

Lars Kaderali

Ein Hierarchischer Bayes'scher Ansatz zur Regression mit
Anwendungen in der Überlebenszeit-Vorhersage bei
Krebserkrankungen

Zusammenfassung

In dieser Arbeit wird ein hierarchischer Bayes'scher Ansatz entwickelt, der Regression in hochdimensionalen Vektorräumen mit einer geringen Anzahl von Datenpunkten ermöglicht, wenn zusätzlich angenommen werden kann, dass viele der Eingabe-Messwerte irrelevant sind. Das Modell wird zunächst für ein einfaches lineares Regressionsproblem dargestellt, um anschließend für ein nichtlineares Vorhersageproblem fortentwickelt zu werden: die Vorhersage von Überlebenszeiten von Tumorpatienten aus Genexpressionsdaten.

Zwei Methoden zur Rechnung und Vorhersage mit dem Modell werden vorgestellt: Ein Maximum-a-Posteriori-Ansatz basierend auf einem Gradientenabstiegsverfahren, sowie ein Monte-Carlo Ansatz mittels Markov-Ketten. Letzterer Ansatz ermöglicht nicht nur die Vorhersage von Regressionsschätzwerten und die Beurteilung der Relevanz einzelner Eingabe-Messwerte, sondern kann darüber hinaus zur Berechnung von Konfidenzintervallen und Visualisierung von Verteilungen über sowohl Regressionsschätzwerte als auch über Relevanzen von Eingabemesswerten herangezogen werden.

Die entwickelten Verfahren werden anschließend auf simulierten sowie realen Daten getestet. Die Methoden werden verwendet zur Vorhersage von Überlebenszeiten auf zwei öffentlich zugänglichen klinischen Studien zum B-Zell Lymphom sowie zum Adenokarzinom der Lunge, und es wird im Vergleich mit bestehenden Verfahren sowie klinischem Staging gezeigt, dass die entwickelten Verfahren verbesserte Vorhersagen ermöglichen. Auf einen vorangehenden Dimensionsreduktions-Schritt kann dabei weitgehend verzichtet werden, und im Gegensatz zu den meisten bestehenden Verfahren werden tatsächliche Überlebenszeiten in Jahren und Monaten vorhergesagt, und nicht nur Risiko-Klassen zugeordnet. Durch den wahrscheinlichkeitstheoretischen Ansatz können darüber hinaus Konfidenzintervalle zu den Überlebenszeiten angegeben werden.