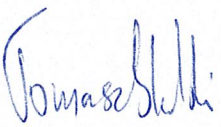
Summary of the doctoral dissertation

Geometric and Combinatorial Aspects of Statistical Models — Tomasz Skalski

This dissertation concerns new applications of discrete geometry and combinatorics in modern statistics. First of them focuses on one of widely used remedies to the inevitable growth of data, that is the use of penalized linear regresion methods. With an aim to recover the needed properties possessed by the vector of regression coefficients, we start our discussion with the Sorted ℓ_1 Penalized Estimator (SLOPE), which was proposed almost a decade ago. Especially we examine the notion of the SLOPE pattern, which maintains the information about the support, sign and ranking between the regression coefficients. In particular, it preserves the clusters of coefficients with the same absolute value. In Chapter 3 we provide the conditions, under which SLOPE recovers the set of relevant covariables and the clusters when the design matrix is orthogonal. We also derive new results on the strong consistency of the SLOPE estimator and its pattern. Chapter 4 extends the discussion on SLOPE to a general class of fixed design matrices. We provide the SLOPE irrepresentability condition, which is necessary and sufficient for the pattern recovery in the noiseless case and illustrate it geometrically. Later on, we consider the case of asymptotic growth of the number of explanatory variables and of the incremental error. In Chapter 5 we study the wider class of penalized estimators, called the polyhedral gauges. It allows one to use the notions from the geometry of polyhedra to generalize the notion of the pattern and the results on its recovery. Chapter 6 is articulated around the existence of the Maximum Likelihood Estimator (MLE) for discrete exponential familics. We give its new characterization based on the notion of the set of uniqueness. Later on, we inspect the size of independent identically distributed samples which is needed to ensure its existence with high probability. For that reason we use the notions from the analysis of discrete hypercubes and apply our results in the environment of random graphs. Last of the chapters connects the theory of graphical models in statistics with the notion of graph Laplacian matrices and discretized Wiener processes. The thesis is based on three already published articles and two preprints, which are available on-line.



Streszczenie rozprawy doktorskiej

Geometryczne i kombinatoryczne zagadnienia modeli statystycznych
Tomasz Skalski

Niniejsza rozprawa poświęcona jest nowym zastosowaniom geometrii dyskretnej i kombinatoryki w nowoczesnej statystyce. Pierwsze z nich skupione jest na jednym z popularniejszych rozwiązań na radzenie z ciągłym przyrostem danych, jest nim penalizowana regresja liniowa. Mając na celu odtworzenie potrzebnych nam własności wektora współczynników regresji, rozpoczynamy dyskusję od estymatora SLOPE (Sorted ℓ_1 Penalized Estimator), który został wprowadzony w poprzedniej dekadzie. Szczególną uwagę poświęcamy pojęciu wzorca SLOPE, który zachowuje informację o nośniku, znaku i rankingu między współczynnikami regresji. Informuje on również o klastrach współczynników o tej samej wartości bezwzględnej. W rozdziale trzecim podajemy warunki, dla których SLOPE poprawnie odtwarza nośnik oraz klastry wektora współczynników regresji przy ortogonalnej macierzy eksperymentu. Przy tym założeniu wyprowadzamy też nowe wyniki dotyczące mocnej zgodności estymatora SLOPE i jego wzorca. Rozdział czwarty rozszerza dyskusję na temat SLOPE, pomijając założenie o ortogonalności macierzy eksperymentu. Wprowadzamy warunek niereprezentowalności dla SLOPE, który jest konieczny i dostateczny do odtworzenia wzorca w przypadku braku szumu, po czym ilustrujemy ten warunek geometrycznie. Następnie rozważamy przypadek asymptotycznego przyrostu liczby zmiennych objaśniających i szumu rosnącego inkrementalnie. W rozdziale piątym omawiamy szerszą klasę penalizowanych estymatorów zwaną polyhedral gauges. Pozwala ona na wykorzystanie twierdzeń z geometrii wielościanów do uogólnienia pojęcia wzorca i wyników dotyczących jego odtwarzania. Rozdział szósty dotyczy istnienia estymatora największej wiarogodności (MLE) w dyskretnych rodzinach wykładniczych. Podajemy jego pełną charakteryzację za pomocą pojęcia zbioru jednoznaczności. Następnie badamy rozmiar próby niezależnych zmiennych losowych o tym samym rozkładzie, która zapewnia istnienie MLE z wysokim prawdopodobieństwem. W tym celu wykorzystujemy narzędzia z analizy hipersześcianów dyskretnych i stosujemy otrzymane wyniki w modelach wykładniczych grafów losowych. Ostatni z rozdziałów skupiony jest na połączeniu między teorią modeli graficznych w statystyce, a pojęciami laplasjanu grafu oraz dyskretyzacji procesów Wienera. Rozprawa jest oparta na trzech opublikowanych artykułach oraz dwóch preprintach dostępnych on-line.

Tomasz Skalski