ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
DEPARTMENT OF STATISTICS

ΚΥΚΛΟΣ ΣΕΜΙΝΑΡΙΩΝ ΣΤΑΤΙΣΤΙΚΗΣ ΔΕΚΕΜΒΡΙΟΣ 2017

Konstantinos Perrakis
DZNE: German Center for Neurodegenerative Diseases, Bonn

Scalable Bayesian regression in high dimensions with multiple data sources

ΠΑΡΑΣΚΕΥΗ 22/12/2017
13:15

ΑΙΘΟΥΣΑ 607, 6ος ΟΡΟΦΟΣ,
ΚΤΙΡΙΟ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
(ΕΥΕΛΠΙΔΩΝ & ΛΕΥΚΑΔΟΣ)

ΠΕΡΙΛΗΨΗ

Many current applications of high-dimensional regression involve multiple sources of covariates. We propose methodology for this setting, motivated by biomedical applications in the "wide data" regime with very large total dimensionality p and sample size n << p. As a starting point, we formulate a flexible ridge-type prior with shrinkage levels that are specific to data type or source. These multiple shrinkage levels are set automatically in a data-driven manner using empirical Bayes. Importantly, all the proposed estimators can be formulated in terms of outer-product data matrices of size n x n, rendering computation fast and scalable in the wide data setting, and are free of user-set tuning parameters. We extend the approaches towards sparse solutions via constrained minimization of a certain Kullback-Leibler divergence, including a relaxed variant that scales to large p, allows adaptive and source-specific shrinkage and has a closed-form solution. The proposed methods are compared to standard high-dimensional methods in a simulation study based on biological data. We present also results from a case study in Alzheimer's disease involving millions of predictors and multiple data sources.

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ DEPARTMENT OF STATISTICS

AUEB STATISTICS SEMINAR SERIES DECEMBER 2017

Konstantinos Perrakis

DZNE: German Center for Neurodegenerative Diseases, Bonn

Scalable Bayesian regression in high dimensions with multiple data sources

FRIDAY 22/12/2017

13:15

ROOM 607, 6th FLOOR,
POSTGRADUATE STUDIES BUILDING
(EVELPIDON & LEFKADOS)

ABSTRACT

Many current applications of high-dimensional regression involve multiple sources of covariates. We propose methodology for this setting, motivated by biomedical applications in the "wide data" regime with very large total dimensionality p and sample size n << p. As a starting point, we formulate a flexible ridge-type prior with shrinkage levels that are specific to data type or source. These multiple shrinkage levels are set automatically in a data-driven manner using empirical Bayes. Importantly, all the proposed estimators can be formulated in terms of outer-product data matrices of size n x n, rendering computation fast and scalable in the wide data setting, and are free of user-set tuning parameters. We extend the approaches towards sparse solutions via constrained minimization of a certain Kullback-Leibler divergence, including a relaxed variant that scales to large p, allows adaptive and source-specific shrinkage and has a closed-form solution. The proposed methods are compared to standard high-dimensional methods in a simulation study based on biological data. We present also results from a case study in Alzheimer's disease involving millions of predictors and multiple data sources.