Technometrics, ISSN 0040-1706
Volume 62, number 4 (2020)

Ridge Regression: A Historical Context P. 420-425
Roger W. Hoerl

Abstract
Two classical articles on Ridge Regression by Arthur Hoerl and Robert Kennard were published in *Technometrics* in 1970, making 2020 their 50th anniversary. The theory and practice of Ridge Regression, and of related biased shrinkage estimators, have been extensively developed over the years. Further, newer shrinkage estimators, such as the Lasso and the Elastic Net, have become popular more recently. These newer developments have led to renewed interest in the original 1970 articles. What has perhaps been lost since 1970 is the context of these classic articles. That is, who were Art Hoerl and Bob Kennard, and what led two statisticians working in the private sector to develop Ridge Regression in the first place? What are the origins of Ridge Regression? Where did the name come from? The purpose of this article is to provide this historical context by discussing the men involved, their work at DuPont, and their approach to methodological development. As Art Hoerl was my father, this is admittedly a personal viewpoint.

Ridge Regularization: An Essential Concept in Data Science P. 426-433
Trevor Hastie

Abstract
Ridge or more formally l2 regularization shows up in many areas of statistics and machine learning. It is one of those essential devices that any good data scientist needs to master for their craft. In this brief *ridge fest*, I have collected together some of the magic and beauty of ridge that my colleagues and I have encountered over the past 40 years in applied statistics.

Generalized Principal Component Analysis: Projection of Saturated Model Parameters P. 459-472
Andrew J. Landgraf & Yoonkyung Lee

Abstract
Principal component analysis (PCA) is very useful for a wide variety of data analysis tasks, but its implicit connection to the Gaussian distribution can be undesirable for discrete data such as binary and multi-category responses or counts. We generalize PCA to handle various types of data using the generalized linear model framework. In contrast to the existing approach of matrix factorizations for exponential family data, our generalized PCA provides low-rank estimates of the natural parameters by projecting the saturated model parameters. This difference in formulation leads to the favorable properties that the number of parameters does not grow with the sample size and simple matrix multiplication suffices for computation of the principal component scores on new data. A practical algorithm which can incorporate missing data and case weights is developed for finding the projection matrix. Examples on simulated and real count data show the improvement of generalized PCA over standard PCA for matrix completion, visualization, and collaborative filtering.

P. 473-485

Sparse Principal Component Analysis Based on Least Trimmed Squares
Yixin Wang & Stefan Van Aelst

Abstract
Sparse principal component analysis (PCA) is used to obtain stable and interpretable principal components (PCs) from high-dimensional data. A robust sparse PCA method is proposed to handle potential outliers in the data. The proposed method is based on the least trimmed squares PCA method which provides robust but non-sparse PC estimates. To obtain sparse solutions, our method incorporates a regularization penalty on the loading vectors. The principal directions are determined sequentially to avoid that outliers in the PC subspace destroy the sparse structure of the loadings. Simulation studies and real data examples show that the new method gives accurate estimates, even when the data are highly contaminated. Moreover, compared to existing robust sparse PCA methods the computation time is reduced to a great extent. Supplementary materials providing more simulation results and discussion, and an R package to compute the proposed method are available online.

Sequential Model-Based Optimization for Continuous Inputs with Finite Decision Space P. 486-498
Qiong Zhang & Youngdeok Hwang

Abstract
Optimization using stochastic computer experiments is commonplace in engineering and industry. This article addresses the problem of optimization, in which the input space of stochastic computer model is continuous, whereas the decision space in the real problem is restricted to be finite. We propose a sequential model-based optimization method to tackle this problem. The knowledge gradient based on this restricted decision space is used as the criterion to choose new design points, and the fixed rank kriging or Gaussian process is used as the surrogate. This combination takes advantage of the continuous search space to decrease the uncertainty over the finite decision space. We demonstrate the benefit of our proposed methodology compared with existing sequential model-based optimization methods.

Adaptive Minimum Confidence Region Rule for Multivariate Initialization Bias Truncation in Discrete-Event Simulations P. 499-512
Jianguo Wu, Honglun Xu, Feng Ju & Tzu-Liang (Bill) Tseng

Abstract
Initialization bias truncation is critically important for system performance assessment and warm-up length estimation in discrete-event simulations. Most of the existing methods are for univariate signals, while multivariate truncation has been rarely studied. To fill such gap, this article proposes an efficient method, called adaptive minimum confidence region rule (AMCR) for multivariate initialization bias truncation. It determines the truncation point by minimizing the modified confidence volume with a tuning parameter for the mean estimate. An elbow method is developed for adaptive selection of the tuning parameter. Theoretical properties of the AMCR rule for both data with and without autocorrelations have been derived for justification and practical guidance. The effectiveness and superiority of the AMCR rule over other existing approaches have been demonstrated through thorough numerical studies and real application.

Uncertainty Quantification for High-Dimensional Sparse Nonparametric Additive Models P. 513-524
Qi Gao, Randy C. S. Lai, Thomas C. M. Lee & Yao Li

Abstract
Statistical inference in high-dimensional settings has recently attracted enormous attention within the literature. However, most published work focuses on the parametric linear regression problem. This article considers an

important extension of this problem: statistical inference for high-dimensional sparse nonparametric additive models. To be more precise, this article develops a methodology for constructing a probability density function on the set of all candidate models. This methodology can also be applied to construct confidence intervals for various quantities of interest (such as noise variance) and confidence bands for the additive functions. This methodology is derived using a generalized fiducial inference framework. It is shown that results produced by the proposed methodology enjoy correct asymptotic frequentist properties. Empirical results obtained from numerical experimentation verify this theoretical claim. Lastly, the methodology is applied to a gene expression dataset and discovers new findings for which most existing methods based on parametric linear modeling failed to observe.

Transformation and Additivity in Gaussian Processes P. 525-535

Li-Hsiang Lin & V. Roshan Joseph

Abstract

We discuss the problem of approximating a deterministic function using Gaussian processes (GPs). The role of transformation in GP modeling is not well understood. We argue that transformation of the response can be used for making the deterministic function approximately additive, which can then be easily estimated using an additive GP. We call such a GP a transformed additive Gaussian (TAG) process. To capture possible interactions which are unaccounted for in an additive model, we propose an extension of the TAG process called transformed approximately additive Gaussian (TAAG) process. We develop efficient techniques for fitting a TAAG process. In fact, we show that it can be fitted to high-dimensional data much more efficiently than a standard GP. Furthermore, we show that the use of the TAAG process leads to better estimation, interpretation, visualization, and prediction. The proposed methods are implemented in the R package *TAG*.

Statistical Modeling of Multivariate Destructive Degradation Tests With Blocking P. 536-548

Qiuzhuang Sun, Zhi-Sheng Ye & Yili Hong

Abstract

In degradation tests, the test units are usually divided into several groups, with each group tested simultaneously in a test rig. Each rig constitutes a rig-layer block from the perspective of design of experiments. Within each rig, the test units measured at the same time further form a gauge-layer block. Due to the uncontrollable factors among test rigs and the common errors incurred for each measurement, the degradation measurements of the test units may differ among various blocks. On the other hand, the degradation should be more homogeneous within a block. Motivated by an application of emerging contaminants (ECs), this study proposes a multivariate statistical model to account for the two-layer block effects in destructive degradation tests. A multivariate Wiener process is first used to model the correlation among different dimensions of degradation. The rig-layer block effect is modeled by a one-dimensional frailty motivated by the degradation physics, while the gauge-layer block effect at each measurement epoch is captured by a common additive measurement error. We develop an expectation-maximization algorithm to obtain the point estimates of the model parameters and construct confidence intervals for the parameters. A procedure is proposed to test significance of the block effects in the degradation data. Through a case study on an EC degradation dataset, we show the existence of the two-layer block effects from the test. By making use of the proposed model, decision makers can readily make risk assessment of each contaminant and determine the minimal water treatment time for removal of the contaminants.

A Note on Cross-Validation for Lasso Under Measurement Errors P. 549-556

Abhirup Datta & Hui Zou

Abstract

Variants of the Lasso or $\ell 1$-penalized regression have been proposed to accommodate for presence of measurement errors in the covariates. Theoretical guarantees of these estimates have been established for some oracle values of the regularization parameters which are not known in practice. Data-driven tuning such as cross-validation has not been studied when covariates contain measurement errors. We demonstrate that in the presence of error-in-covariates, even when using a Lasso-variant that adjusts for measurement error, application of naive leave-one-out cross-validation to select the tuning parameter can be problematic. We provide an example where such a practice leads to estimation inconsistency. We also prove that a simple correction to cross-validation procedure restores consistency. We also study the risk consistency of the two cross-validation procedures and offer guideline on the choice of cross-validation based on the measurement error distributions of the training and the prediction data. The theoretical findings are validated using simulated data.

Technometrics, ISSN 0040-1706
Volume 62, number 4 (2020)