(JASA)

1. **Online Covariance Matrix Estimation in Stochastic Gradient Descent**

   Comment: Covariance matrix를 추정하는 한 방법을 소개합니다. SGD를 개량한 averaged SGD (ASGD)를 소개하고 asymptotic estimate와 statistical inference를 계산합니다.

   Abstract: This article aims at conducting statistical inference of SGD-based estimates in an online setting. In particular, we propose a fully online estimator for the covariance matrix of averaged SGD (ASGD) iterates only using the iterates from SGD. We formally establish our online estimator's consistency and show that the convergence rate is comparable to offline counterparts. Based on the classic asymptotic normality results of ASGD, we construct asymptotically valid confidence intervals for model parameters. Upon receiving new observations, we can quickly update the covariance matrix estimate and the confidence intervals.

2. **What are the Most Important Statistical Ideas of the Past 50 Years?**

   Comment: 제목에서 보면 알 수 있듯이, 무언가 발견한 것은 아니고 조사 & 리뷰 논문입니다. 저자는 Andrew Gelman 교수님입니다. 상당히 유명한 분인 것 같고, 논문 조회수가 4000을 넘었습니다. 내용도 괜찮은 것 같아서 추천 드립니다.

   Abstract: We review the most important statistical ideas of the past half century, which we categorize as: counterfactual causal inference, bootstrapping and simulation-based inference, overparameterized models and regularization, Bayesian multilevel models, generic computation algorithms, adaptive decision analysis, robust inference, and exploratory data analysis. We discuss key contributions in these subfields, how they relate to modern computing and big data, and how they might be developed and extended in future decades.

3. **A Normality Test for High-dimensional Data based on a Nearest Neighbor Approach**

   Comment: nearest neighbor information 을 이용한 정규성 판정 방법을 소개합니다. High-dimensional에도 성립할뿐더러 parameter를 사용하지 않기 때문에 상당히 좋은 방법으로 추정됩니다. (즉, nonparametric normality test). 논문이 친절한 것도 덤입니다.

   Abstract: In this work, we propose a novel nonparametric test that utilizes the nearest neighbor information. The proposed method guarantees the asymptotic type I error control under the high-dimensional setting. Simulation studies verify the empirical size performance of the proposed test when the dimension grows with the sample size and at the same time exhibit a superior power performance of the new test compared with alternative methods.

(BIOMETRIKA)

1. **On order determination by predictor augmentation**

   Comment: 통계와 머신러닝과 같은 주제를 접할 때 언제나 curse of dimensionality를 피할 수 없습니다. 특히 이 논문은 어떤 matrix를 추정할 때 최소의 차원을 찾는 적절한 방법을 제시해 줍니다. 핵심 아이디어는 predictor에 노이즈를 추가하는 것입니다.

   Abstract: In this article, we propose a novel and highly effective order-determination method based on the idea of predictor augmentation. We show that if the predictor is augmented by an artificially generated random vector, then the parts of the eigenvectors of the matrix induced by the augmentation display a pattern that reveals information about the order to be determined.

2. **A parsimonious personalized dose-finding model via dimension reduction**

   Comment: covariate되어있는 변수가 많을 때 아주 작은 집합으로부터 생성되는 공간을 estimate하는 것은 기존의 방법으로 어렵다고 합니다. 이 경우 bootstrap estimator는 효과적임을 알려주는 논문입니다. Introduction이 꽤나 흥미롭습니다.

   Abstract: We propose a dimension reduction framework that effectively reduces the estimation to an optimization on a lower-dimensional subspace of the covariates. We exploit the fact that the individualized dose rule can be defined in a subspace spanned by a few linear combinations of the covariates to obtain a more parsimonious model. Owing to direct maximization of the value function, the proposed framework does not require the inverse probability of the propensity score under observational studies. Under mild regularity assumptions, results on the asymptotic normality of the proposed estimators are established. We also derive the consistency and convergence rate of the value function under the estimated optimal dose rule.

3. **Learning Block Structures in U-statistic Based Matrices**

   Comment: Block-structured 행렬은 어디서나 볼 수 있는 주제입니다. 이 논문은 U-statistic과 large dimensional random matrix theory를 이용한 clustering을 제시합니다. 그 결과로 variable selecting을 성공적으로 할 수 있는 결과를 예시로 보여줍니다.

   Abstract: We introduce a conceptually simple, efficient and easily implemented approach for learning the block structure in a large matrix. Using the properties of U-statistic and large dimensional random matrix theory, the group structure of many variables can be directly identified based on the eigenvalues and eigenvectors of the scaled sample matrix.

4. **Dimension Reduction for Covariates in Network Data**

   Comment: 여러 사람들과 그들의 정보 관계를 weighted graph having multiple edges로 나타낼 수 있을 것 입니다. 이러한 수 많은 정보를 low-dimensional space로 투영하는 방법을 제시합니다. 이는 network-supervised dimension reduction 라는 방법이 적용됩니다.

   Abstract: A problem of major interest in network data analysis is to explain the strength of connections using context information. To achieve this, we introduce a novel approach named networksupervised dimension reduction by projecting covariates onto low-dimensional spaces for revealing the linkage pattern, without assuming a model. We propose a new loss function for estimating the parameters in the resulting linear projection, based on the notion that closer proximity in the low-dimension projection renders stronger connections

(PMLR)

1. **Accelerating Distributed SGD for Linear Regression using Iterative Pre-Conditioning**

   Comment: distributed linear least squares problem이 무엇인지 소개하고 이를 해결하는 것에 있어서 stochastic algorithms을 도입합니다. 기존의 Iteratively Pre-conditioned Gradient-descent (IPG) method (2020) 보다 좋은 성능을 보입니다.

   Abstract: We extend the idea of iterative pre-conditioning to the stochastic settings. We also show that our proposed Iteratively Pre-conditioned Stochastic Gradient-descent (IPSG) method converges linearly in expectation to a proximity of the solution.

2. **Learning without Knowing: Unobserved Context in Continuous Transfer Reinforcement Learning**

   Comment: 강화 학습에서 보상을 사전에 알 수 없는 경우 어떤 선택이 효과적인지 보여줍니다. 기존의 방법이 왜 안되는지 효과적인 예제와 함께 시작하기 때문에 동기부여가 잘 되는 논문입니다.

   Abstract: In this paper, we consider a transfer Reinforcement Learning (RL) problem in continuous state and action spaces, under unobserved contextual information. For example, the context can represent the mental view of the world that an expert agent has formed through past interactions with this world. We assume that this context is not accessible to a learner agent who can only observe the expert data. Then, our goal is to use the context-aware expert data to learn an optimal context-unaware policy for the learner using only a few new data samples.

3. **Exploiting Sparsity for Neural Network Verification**

   Comment: fully-connected feed forward neural network에 Chordal Graphs의 특성과 이론을 적용하여 딥러닝의 일부분을 이론적으로 설명하려는 논문입니다. 본 논문에 자세한 정의, 정리, 레퍼런스가 있으므로 Chordal Graph가 몰라도 될 것 같습니다.

   Abstract: We investigate the sparsity that arises in a recently proposed semi-definite programming framework to verify a fully connected feed-forward neural network. We show that due to the intrinsic cascading structure of the neural network, the constraint matrices in the semi-definite program form a block-arrow pattern and satisfy conditions for chordal sparsity. We reformulate and implement the optimisation problem, showing a significant speed-up in computation, without sacrificing solution accuracy.

(Others)

1. **Latent Neural Differential Equations for Video Generation**

   Comment: Neural ODE 및 SDE를 적용한 최근 논문입니다. (메인은 아님) 이전 발표에서 재미있게 들어서 남깁니다.

   Abstract: We study the effects of Neural Differential Equations to model the temporal dynamics of video generation. The paradigm of Neural Differential Equations presents many theoretical strengths including the first continuous representation of time within video generation. In order to address the effects of Neural Differential Equations, we investigate how changes in temporal models affect generated video quality. Our results give support to the usage of Neural Differential Equations as a simple replacement for older temporal generators.

2. **A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis**

   Comment: matrix의 eigenvector를 구할 때 사용하는 power method의 constraint 버전이라고 보면 됩니다. 아주 널리 사용되는 듯 하여 추천합니다.

   Abstract: We present a penalized matrix decomposition (PMD), a new framework for computing a rank-K approximation for a matrix. This results in a regularized version of the singular value decomposition. In addition, we show that when the PMD is applied to a cross-products matrix, it results in a method for penalized canonical correlation analysis (CCA). We apply this penalized CCA method to simulated data and to a genomic data set consisting of gene expression and DNA copy number measurements on the same set of samples.