Plan Overview

A Data Management Plan created using DMPTool

DMP ID: https://doi.org/10.48321/D1KS9K

Title: DMSP for "CDSE: Quantization-Based Methods for Optimal Nonparametric Inference"

Creator: Cristo Yanez leon - **ORCID:** 0000-0002-0930-0179

Affiliation: New Jersey Institute of Technology

Principal Investigator: Zuofeng Shang

Contributor: Cristo Leon

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: PD 23-8084

Grant: https://new.nsf.gov/funding/opportunities/computational-data-enabled-science-engineering-3

Template: NSF-DMS: Mathematical Sciences

Project abstract:

With rapid development in science and technology, massive data are ubiquitous. The intrinsic highdimensional/functional structure in massive data often requires more sophisticated statistical modeling and data processing tools such as deep neural networks. Meanwhile, due to limited data storage capacity, a large portion of the massive data was either ephemeral or temporarily cached and subsequently overwritten with newer data. Limited data storage capacity, as well as the growing degree of complexity in statistical models, severely challenge standard nonparametric inferential theory and applications. A long-term goal of the PI's research is to promote advanced nonparametric methods to overcome modern massive data challenges. The specific aim of this project is to explore nonparametric inferential procedures when data are quantized, and models are highdimensional/functional involving complicated interaction effects. Statistical optimality of the procedures, in the presence of data and modeling challenges, forms the core of this proposal. Theoretical insights gained from analyzing the proposed algorithms are beneficial for real-world problems ranging from large scale data to functional data.

Start date: 06-01-2024

End date: 05-31-2027

Last modified: 01-19-2024

Copyright information:

DMSP for "CDSE: Quantization-Based Methods for Optimal Nonparametric Inference"

The project’s principal investigator, Zuofeng Shang, will ultimately be responsible for all of the data management. Cristo Leon’s role as compliance assisted with policies.

The project will not generate data, samples and physical collections. Computing software such as R packages for implementing the proposed algorithms shall be made available to the public through Github at the end of the project period. Certain components of the research outputs shall be used as curriculum materials.

The project doesn't produce data.

The project doesn't produce data.

The project doesn't produce data.

The project doesn't produce data.
A paper will be created as an output.

Dublin core is a basic domain-agnostic standard that can be easily understood and implemented and will be used for metadata standards. Following the APA citation format.

Planned Research Outputs

Data paper - "Optimal quantization methods for nonparametric inference"

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Optimal quantization methods for nonparametric inf ...	Data paper	Unspecified	Open	None specified		None specified	Dublin Core	No	No