Arkajyoti Bhattacharjee

arkab-ds.github.io | bhattacharjee.57@osu.edu | +1 614 989 7490 | ArkaB-DS | arkajyoti-aj

EDUCATION

The Ohio State University, Department of Statistics — Columbus, OH, USA
Doctor of Philosophy (Ph.D.) in Statistics — *Aug '22 - Ongoing*

- Graduate Teaching Assistant: *Aug '22 - Ongoing*
  - Responsible for grading, tutoring, and assisting in R labs for STAT 2480, STAT 3470.

Indian Institute of Technology Kanpur, Department of Mathematics and Statistics — Kanpur, UP, India
Master of Science in Statistics — *Sep '20 – Jun '22*

- Relevant Coursework: Statistical & AI Techniques in Data Mining, Time Series Analysis, Inference I & II, Multivariate Analysis, Statistical Simulation & Data Analysis, MCMC.
- Awards: Academic Excellence Award (2020, 2021) (for the top 10% of the batch and having a CPI > 8.5)
- Leadership: Student Nominee in Departmental Undergraduate Committee; STAMATICS Coordinator

Presidency University, Department of Statistics — Kolkata, WB, India
Bachelor of Science (Honours) in Statistics — *Aug '16 – Jun '19*

PROFESSIONAL EXPERIENCES

The R Project for Statistical Computing, Google Summer of Code, 2021 — Remote
Student Developer — *Jun – Aug '21*
Mentors: Prof. John Nash (Univ. of Ottawa), Dr. Heather Turner (Univ. of Warwick)

- Developed an **R package nlsCompare** that compares the accuracy of model characteristics derived using existing or new R functions for nonlinear least-squares.
- Wrote **tests** for nlsj, an interim R package developed to improve *nls()* functionalities, to ensure robust code structure.

View Project: https://summerofcode.withgoogle.com/projects/5154479671869440

Accenture Solutions Pvt. Ltd. — Remote
Data Science Analyst Intern — *May – Jul '21*

- Developed an **AutoML**-based tool with **GUI**, using Python and Dash, that allows users to:
  - use multiple data pre-processing methods and select among classical and state-of-the-art algorithms;
  - use **model optimization** and **model blending**, interactively visualize **model diagnostics** and make predictions based on the best-optimized model.

SKILLS

- Programming Languages and Software: R, Python, C, LaTeX

PROJECTS

A Brief Review of Sparse Principal Components Analysis (SPCA) — *Mar – Apr '22*

- Reviewed SPCA and General Adaptive SPCA as **dimension reduction** techniques and applied them to simulated and real data.

Efficient High-Dimensional Robust Variable Selection via Rank-based LASSO Methods — *Mar – Apr '22*

- Reviewed the properties of **Rank-LASSO** and modified Rank-LASSO as a robust variable selection method.
- Demonstrated their superior performance over regular **LAD-LASSO** under **high-dimensional** settings via simulations.

Spectral Clustering (SC): Theory and Applications — *Mar – Apr '22*

- Reviewed similarity graphs, graph Laplacians, three popular SC algorithms, and different views of SC.
- Applied SC for **image segmentation**, compared it with **k-means clustering,** and applied it to the **Iris** dataset.

Understanding Nonparametric Multimodal Regression (NMR) via Kernel Density Estimation (KDE) — *Feb '22*

- Reviewed estimation using the **mean-shift algorithm,** geometry and consistency in NMR.
- Constructed confidence sets using **bootstrap** and bandwidth selection of the KDE using prediction sets' sizes.

Understanding Confidence Intervals in Adaptive Markov Chain Monte Carlo (AMCMC) — *Aug – Nov '21*

- Reviewed the literature on the kernel estimators of asymptotic variance and confidence intervals in AMCMC.
- Verified conditions under which a **CLT** holds via simulation examples in R.