JAMES BUENFIL

buenfil@uw.edu ◇ Personal Website

Research key-words: Dimensionality reduction, High-dimensional statistics, Feature learning, Unsupervised learning, Variable selection, Functional data analysis, Geometric data analysis

EDUCATION

University of Washington, Seattle *Oct. 2020 - Present*
PhD Statistics

Relevant Coursework:
Theory: Advanced Probability Theory (Math 521, 522), Advanced Theory of Statistical Inference (STAT 581, 582, 583)

Methods: Statistical Learning (STAT 535), Multivariate Analysis (STAT 542), Advanced Regression Methods for Independent and Dependent Data (STAT 570, 571), Statistical Consulting (STAT 599)

University of Wisconsin, Madison *Aug. 2016 - May 2020*
B.S. Applied Mathematics, Engineering, and Physics

Relevant Coursework:
Computer Science: Intro. to Optimization (CS 524), Intro. to Data Structures (CS 367), Intro. to Algorithms (CS 577), Numerical Linear Algebra (CS 513), Numerical Analysis (CS 514)

Statistics: Intro. to Stochastic Processes (STAT 632), Intro. to Statistical Inference (STAT 610)

EXPERIENCE

Research Assistant, with Professor Marina Meila
Statistics Dept. UW-Seattle *March 2023 - Present*

- Designing approach to use supervised learning embeddings (e.g. layer of a neural network) of high dimensional data to define what the important geometry of data is.
  - Subsequently constructing an interpretable embedding of the data using user-defined dictionary functions that matches the important geometry.
- Working with the National Security Agency (NSA) on hypothesis testing problems related to rankings from recommender systems.

Research Assistant, with Assistant Prof. Eardi Lila
Biostatistics Dept. UW-Seattle *May 2022 - Present*

- Developing theoretically guaranteed novel feature learning and variable selection techniques to find features which are maximally correlated between datasets of different modalities.
  - One dataset consists of high dimensional multivariate data, while the other consists of time-dependent curves which lie on a manifold (e.g. the set of positive definite matrices).
  - Applying to investigate relationships between patient lifestyle/ demographic variable data and Diffusion-MRI brain scan data. R package on CRAN coming soon.

Teaching Assistant for DATA 556, Intro. to Statistics and Probability
Master of Science Data Science Program, UW-Seattle *Sept. 2022 - Dec. 2022*

- Duties include holding discussion sections for homework assignments, grading homework assignments and exams, managing the course website, and regularly answering student questions.

Research Assistant, with Professor Marina Meila
Statistics Dept. UW-Seattle *April 2021 - June 2021*

- Developed a novel molecular dynamics enhanced sampling method, "Tangent Space Least Adaptive Clustering"
  - Resulted in first-author paper accepted for poster presentation at the ICML 2021 Workshop on Unsupervised Reinforcement Learning.

Independent Study with Garvesh Raskutti (then Assistant Professor)
Statistics Dept. UW-Madison *May 2019 - Jan 2020*

- Applied the PUlasso algorithm, a variable selection approach designed to take advantage of unlabeled data, along with other classification methods, to a large amino-acid sequence dataset.

Independent Study with Benjamin Peherstorfer (then Assistant Professor)
Mechanical Engineering Dept. UW-Madsion *Sep. 2017 - June 2018*

- Ran finite element method simulations of a Navier-Stokes fluid flow problem.
- Transformed a system of partial differential equations describing the chemical equilibrium of a tubular reactor, leading to more efficient numerical simulation via reduced-order modeling.

Research Assistant, Mcdermott Lab
Physics Dept. UW-Madison *Feb. 2017 - Aug. 2017*

- Assisted with testing the performance of black-box optimization algorithms.
- Performed a theoretical calculation of the impedance of a Josephson junction transmission line.

PUBLICATIONS

1. **Buenfil, James,** Samson J. Koelle, and Marina Meila. "Tangent Space Least Adaptive Clustering." ICML 2021 Workshop on Unsupervised Reinforcement Learning. 2021.
2. **Buenfil, James,** Eardi Lila. "Hybrid canonical correlation analysis of Riemannian and high-dimensional data." In preparation.
3. **Buenfil, James,** Marina Meila. "Supervised learning to learn the important geometry of high dimensional data; Unsupervised learning to interpret it." In preparation.

SKILLS AND OTHER

Programming Languages:
Highly proficient: Python, R, MATLAB
Familiar: Java, Julia, PyTorch

Foreign Languages:
English (native), Spanish (advanced), French (basic)

Co-organizer of the Geometric Data Analysis Reading Group at UW-Seattle.

ACADEMIC AWARDS

2020-22 GO-MAP Graduate Excellence Award
2020-22 ARCS Foundation Scholar