Ross Altman

+1-516-375-4505 | ross.e.altman@gmail.com | github.com/knowbodynos | linkedin.com/in/ross-altman | Publications

Summary

Experienced data/ML scientist, physicist, and engineer. Lifelong learner, bad pun teller. Driven to understand how things work. Solving puzzles using data, algorithms, hardware, and research fundamentals.

Work Experience

Inari Agriculture *Cambridge, MA*

Data Scientist *Jan 2019 - Present*

- Co-led a four-person team to develop a BERT-based protein language model to discover novel stability-enhancing mutations for improving editing efficiency of CRISPR-Cas proteins in plant cells.
- Developed epistasis-aware DNA and protein language models to identify high-impact deleterious mutations in sequenced maize varieties and prioritize them for correction via gene editing.
- With a three-person team, developed a method for discovering causal interventions for trait improvement in soybean by integrating gene regulatory network analysis, AlphaFold2, and protein language models.
- Mentored 8+ junior colleagues and interns, attended multiple conferences, and maintained regular communication with external scientific advisors.
- Co-invented three patents.

Insight Data Science *Boston, MA*

Fellow *Sept 2018 - Dec 2018*

- Built a https://www.wikontext.us/ to enhance Wikipedia page previews by intelligently displaying more relevant information in contrast to the baseline naive page summary.
- Implemented a Word2Vec-based sentence embedding model, and externally validated it using knowledge of bi-directional clickstream information with up to 17% more accuracy than baseline.
- Served the model using a jQuery/Flask/AWS stack to achieve high performance on real-time, client-side data.

Northeastern University *Boston, MA*

Graduate Research Assistant *Sept 2011 - May 2017*

- Designed a modular, distributed algorithm in C++/Python to systematically compute type-IIB string theory vacuum states on a HPC cluster.
- Generated and analyzed the world's largest string theory dataset of 10^5 vacuum spaces, resulting in the discovery of many new candidate universes extending the Standard Model.
- Built and maintained a queryable UI with MongoDB backend based on feedback from stakeholders in the string theory community.
- Estimated the number of vacuum spaces in the string landscape multiverse ($> 10^{(10^4)}$) using deep neural networks implemented in Keras using specialized "equation learner" layers.
- Co-founded the String Data group for ML applied to string theory, leading to international conferences and industry collaborators.

Education

Northeastern University *Boston, MA*

PhD in Physics *Sept 2011 - May 2017*

- **Focus:** Computational Methods, High-Dimensional Geometry, String Theory.

Cornell University *Ithaca, NY*

MEng in Applied Physics *Sept 2010 - May 2011*

Cornell University *Ithaca, NY*

BSc in Applied and Engineering Physics *Sept 2005 - May 2009*

Skills

Programming	Python (Pandas, PyTorch, NumPy, Scikit-learn. etc.), C/C++, Shell, HTML/CSS, JavaScript, Bioinformatic Tools.
Machine Learning	PyTorch, TensorFlow, Scikit-learn, JAX, Hugging Face.
Engineering	Docker, AWS, Kubernetes, Terraform, MongoDB, Airflow, CI/CD.
Soft Skills	Leadership, Teamwork, Literature Review, Rapid Prototyping, Self-Motivation, Scientific Communication, Documentation.

References available upon request.