Alexander Loftus

Data Scientist & Machine Learning Engineer

Linkedin: https://www.linkedin.com/in/alex-loftus/ alexloftus2004@gmail.com — 206-484-1563
Github: https://github.com/loftusa 4222 Moraga Ave, San Diego, CA

Summary of Qualifications

Wrote a textbook on network machine learning: Publishing contract with Cambridge University Press. Contributed 120+ pages of text and data visualization including over 90 figures. Currently editing final draft. One co-author.

1st place ranking in the $100,000 Vesuvius Kaggle competition: Detected ink from ancient scrolls using a two-stage depth-invariant UNETR/Segformer projection model. First place prize of $25,000, competing against 1249 teams. Presented on our solution to a group of 60 people (link)

Best workshop poster, NeurIPS: Won Best Poster award for our saliency clustering paper at NeurIPS 2023.

7+ years of Python experience with beefy open-source contributions in major libraries: Primary contributor to microsoft network statistics package graspologic

Experienced with cloud computing: Handled over $10,000 of AWS research credits for use on a brain network estimation pipeline, resulted in a Nature Methods paper. Years of experience training models in distributed systems in AWS, Google Cloud.

Experienced with CI/CD and solid+agile software development practices: Developed, deployed, and maintained multiple unit testing infrastructures & data pipelines.

Experienced manager: Led a team of three to develop an object detection augmentation algorithm; a team of five to contribute to a brain network estimation pipeline; assistant director managing 8-12 instructors

Implemented numerous algorithms and methods: including GAN/diffusion-based super-resolution models, out-of-sample embedding, covariate-assisted spectral embedding, multiscale manifold learning, saliency clustering models, and variational volumetric autoencoders for molecule representation.

Skills Summary

Languages: Python, R, Rust, Bash, CSS, Mojo. Expertise in Python.

Tools & Frameworks: pytorch, pytorch-lightning, tensorflow, jax, numpy, scipy, pandas, polars, sklearn, seaborn, matplotlib, docker, AWS, google cloud (GCP), photoshop, SQL, weights & biases, mlflow, kubernetes, linux

Areas of Expertise: Linear algebra, probability & statistics, deep learning, information theory, transformers, diffusion models, convolutional autoencoders, embeddings, GPUs and cuda, public speaking, leadership & management, teaching, natural language processing, computer vision

Education

Johns Hopkins University Baltimore, MD
MSE Biomedical Engineering: Machine Learning & Data Science Focus 2019-2021
dean's list, highest honors, GPA 3.95/4.0

Rank #1 Biomedical Engineering department in the world, US News.
Paid Teaching Assistant for neural data science and bioinformatics.
Research focus on spectral methods applied to networks.

Western Washington University Bellingham, WA
BS Behavioral Neuroscience 2014-2018
Founder & President, Computational Neuroscience Club
Vice President, Neuroscience Club

Chemistry & Philosophy minors.
Built computational neuroscience club from scratch, taught weekly seminars. At least two of my students are now pursuing PhDs in data science or machine learning.

Experience

Creyon Bio San Diego, CA
Data Scientist 2023

ESP Embeddings: Feature representation learning for electrostatic potential data. Contrastive learning approach resulted in 10x improvement over previous method in l^2-norm reconstruction accuracy.

Neuron Toxicity Detection: Built deconvolution and segmentation pipeline to detect toxicity in neurons.

Molecule Diffusion: Find toxicity direction in the latent space of a denoising VAE trained to generate molecule representations.

Blue Halo Rockville, MD
Research Engineer 2021-2023

Conditional Image Generation with Generative Adversarial Networks: Augment datasets with diffusion images. Received commendations for my work.

Detecting Objects with Enhanced Yolo and Knowledge Graphs: Object detection & natural language processing project. Predicted network semantic properties of objects in videos.

Geometric Multi-Resolution Analysis: Natural language processing project. Used a low-dimensional manifold on news data to create a hierarchically clustered semantic space.

Johns Hopkins University Baltimore, MD

Research Assistant & MS Researcher — Dr. Joshua Vogelstein 2018-2021

Hands-On Network Machine Learning with Scikit-Learn and Graspologic: Textbook author. Publishing contracts offered by both Springer Publishing and Cambridge University Press. Wrote 120 pages, generated 91 figures, edited into final draft.

Open-Source Contributor to Microsoft network ML package Graspologic: Development and implementation of dimensionality reduction models on networks. Found and addressed a fundamental & difficult problem with a paper I implemented.

Primary maintainer & Developer of m2g, an open-source brain network estimation pipeline: Diffusion MRI to graphs pipeline. AWS cloud-computing integration with pytest CI/CD infrastructure. Eliminated over 1000 lines of code and halved computation time.

University of Washington Seattle, WA

Assistant Director, iD Tech Camps 2014-2018 summers

Asssistant Director: Administrator for a STEM education camp which taught C++, Python, Java, game design, and robotics at the University of Washington. Managed 8-12 instructors with 80-120 students per week.

Curriculum Designer: Conceptualized, designed, and built an online curriculum teaching a Dota 2 modding course. The curriculum was used across 50 camp locations in the United States by tens of thousands of students.

Instructor: Taught courses for three years in summers during college, generally focused on game development. Regularly received feedback from students that I was the best instructor in the camp.

Western Washington University Bellingham, WA

Research Assistant 2015-2018

Worm behavior research: Studied a protein in the worm *c. elegans* called CAMKII involved in learning. Biological assays: PCR, microscopy, cell culture growth. Presented research at the Society for Neuroscience, 2017.

Built data pipeline: Brought over 4 days of manual work down to 5 minutes of running code. Scripts still providing value today, five years later.

Talks & Publications

Hands-on Network Machine Learning with Scikit-Learn and Graspologic: Textbook, 2023, Cambridge University Press publishing contract. Dual authorship, contributed roughly half the pages and figures. Old draft available online.

A Saliency-based Clustering Framework for Identifying Aberrant Predictions: Paper, 2023, NeurIPS (Workshop paper, accepted). Won best poster. Second author, wrote half the paper, designed poster, did literature review.

A low-resource reliable pipeline to democratize multi-modal connectome estimation and analysis: Paper, 2022, Nature Methods, under review. Second author, wrote infrastructure for the codebase. Preprint on biorxiv. Code on github.

State of the Art in Knowledge Editing: Presentation, 2023, for 30 people. Presented on current techniques in knowledge localization and editing in both attention-based and diffusion models, vision and text data

1st Place Solution - Vesuvius Ink Competition: Presentation, 2023, for 60 people. Presenting on our winning solution to a $100,000 Kaggle competition, part of the $1,000,000 Vesuvius competition.

ICML Conference Highlights: Talk, 2023, about machine learning techniques in drug discovery and medicine at ICML 2023

Linear Algebra, from Dot Products to Neural Networks: I created a YouTube tutorial series on the fundamentals of linear algebra for machine learning. Created from tutoring sessions given to a friend.

Working with LLMs: Talk, 2023, for 100 people at the AIML San Diego meetup

Effects of an unc-43 (CaMKII) Gene Deletion on Short-Term Memory for Associative Conditioning in C. elegans: Talk, presented at Psychfest, 2017, Bellingham, WA.

Role of CAMKII in Associative Conditioning and GLR-1 Expression in C. Elegans: Poster, presented at Society for Neuroscience, 2017, Washington, DC. Later author, conducted most of the later experiments.