13 Jan 2019

Canvas Technology
1855 South 57th Court
Boulder, CO, USA

SUBJECT: Application for Machine Learning Scientist (Intern)

What summary statistics are available from point cloud data sets, when one is forced to make a minimal interpretation of the point cloud? In my doctoral research of topological data analysis (TDA), I am concerned with this genre of *agnostic model selection problems*. Questions in computer vision are related. What statistics are available from noisy visual signals, when one wants a model that will generalize well? Consider the relationship between neural nets and TDA in this workflow.

- Reason about possible signals with TDA and computational geometry (pre-processing),
- Mung, de-noise, and classify signals in a trained multi-layer neural network (post-processing).
- Compare classifications to summary statistics from pre-processing. Iterate.

I am motivated to intern at Canvas because, just as autonomous vehicle safety spurned on computational geometry research in the late 1990's, it continues to motivate developments in TDA. My goal, then, as a Machine Learning Intern, is to contribute to the design and effective implementation of neural architectures.

To outline my qualifications to design and deploy neural architectures:

1. **Technical sophistication.** My basic computing toolkit includes version control (git, duplicity), high level (R, Python, Haskell) and low level (C++) programming languages, and scientific markup (`ipynb`, `Rmd`, XML, HTML, pandoc markdown, TeX). I am accustomed to both Unix-like (MacOS, Ubuntu/Debian) and Windows operating systems. I am comfortable performing computations on the Google Cloud Platform, and I have introductory knowledge of XSEDE science gateways from participating in SGCI webinars.

2. **Project-specific interest.** This fall I contributed to CU Boulder's StatOptML (Statistics, Optimization, and Machine Learning) seminar. I applied TDA to consider "is depth needed for deep learning?" following Guss and Salakhutdinov's empirical study [1]. Using persistent homology to catalog feed forward neural architectures, I introduced myself to TensorFlow and Ripser. I concluded neural networks of fixed width and varying depth exhibit statistically significant topological phase transitions on training data with any homological complexity.

3. **Career goals.** Statistical and computational experience would support my career goal to do topological data analysis. There are limited opportunities in my department (pure mathematics) to train with powerful computational tools. I believe that interning at Canvas would complement my theoretical strengths.

4. **Unique background.** After my undergrad, I took two years to perform stipended service work. For a year in Houston, I developed scalable resources for refugee social services, including a crowd-sourced map of primary care clinics and languages spoken. I wrote bug reports for the implementation three SQL databases, and, when Texas cut funding for Refugee Medical Assistance, I contributed to a data management plan for refugees transitioning from state to federal medical care. For a year in Olympia, I served as a community organizer at a 24/7 homeless shelter. I relied on distributed version control, and became a staunch advocate for deploying "early, often, and with redundant backups".

In all, I have technical expertise, project-specific interests, compatible career goals, and unique background experience.

[1] W. H. Guss and R. Salakhutdinov, "On Characterizing the Capacity of Neural Networks using Algebraic Topology," *arXiv:1802.04443 [cs, math, stat]*, Feb. 2018.

Thank you,

Colton Grainger