

jacobschreiber

<http://homes.cs.washington.edu/~jmschr/>

about

jmschr@cs.washington.edu
github.com/jmschrei

programming

python, cython, sql, bash,
git

academic interests

machine learning
probabilistic modelling
deep learning
bayesian networks
bioinformatics

education

- since 2014 **ph.d.** student in computer science University of Washington
Advisor: Dr. William Noble
- 2009–2013 **b.s. cum laude** University of California, Santa Cruz
biomolecular engineering
bioinformatics minor

experience

- 2016 **Data Science Team, AspenTech** Research Intern
Reference: Mike Noskov Mike.Noskov@aspentech.com
As a member of the AspenTech data science team, my focus was on implementing scalable and efficient probabilistic models for the analysis of large amounts of data.
- 2015 **CEA Saclay, INRIA** Software Engineering Intern
Reference: Olivier Grisel olivier.grisel@inria.fr
As a software engineering intern on the scikit-learn team, my focus was on speeding up the gradient boosting implementation. My work ended up speeding up most tree-based estimators in addition to gradient boosting.
- 2013-2014 **Nanopore Group, University of California Santa Cruz** Junior Specialist
Reference: Kevin Karplus +1 831 459 4250 karplus@soe.ucsc.edu
As a computational specialist in the UCSC Nanopore Group, I used machine-learning methods to automate the analysis of large volumes of data. I was involved with writing several papers based on my findings and grant which got accepted from the NIH.
- 2011-2013 **Nanopore Group, University of California Santa Cruz** Research Assistant
Reference: Mark Akeson +1 831 459 1038 makeson@soe.ucsc.edu
I performed nanopore experiments in the UCSC nanopore lab and all of the wetlab preprocessing steps for the components which went into them.

awards

- 2015 **IGERT Big Data Fellowship**
Two year fellowship awarded to students pursuing a topic involving big data at the intersection of two fields. My fellowship was awarded to pursue scalable machine/deep learning techniques to analyze the 3D architecture of the genome.
- 2015 **NSF GFRP Scholarship Honorable Mention**
Runner up award for the prestigious NSF GFRP fellowship to signify an impressive application, but an unfortunate lack of resources to support their studies.
- 2013 **Dean's Award**
Top 10 senior research thesis in the School of Engineering.

publications

- 2017 **Nucleotide sequence and DNaseI sensitivity are predictive of 3D chromatin architecture** bioRxiv
Jacob Schreiber, Maxwell Libbrecht, Jeffrey Bilmes, and William Noble
- 2015 **Segmentation of Noisy Signals Generated By A Nanopore** bioRxiv
Jacob Schreiber and Kevin Karplus
- 2015 **Analysis of Nanopore Data using Profile HMMs** Bioinformatics
Jacob Schreiber and Kevin Karplus
- 2014 **Discrimination among Protein Variants Using an Unfoldase-Coupled Nanopore** ACS Nano
Jeffrey Nivala, Logan Mulroney, Gabriel Li, Jacob Schreiber, Mark Akeson
- 2014 **Nanopores Discriminate among Five C5-Cytosine Variants in DNA** JACS
Zachary Wescoe, Jacob Schreiber, Mark Akeson
- 2013 **Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands** PNAS
Jacob Schreiber, Zachary Wescoe, Robin Abu-Shumays, John Vivaian, Baldandorj Baatar, Kevin Karplus, Mark Akeson

projects

- 2015 **scikit-learn** python/cython
scikit-learn is a machine learning package for python with cython optimized implementations. I am currently a developer working on optimizing tree based decision models, with gradient boosting in mind.
- 2015 **pomegranate** python/cython
Graphical models for Python, focusing on hidden Markov models and probability distributions, but also including discrete Bayesian networks, finite state machines, and general mixture models, with efficient cython implementations.
- 2015 **tangerine** python/sql
tangerine is a lightweight embedded database system for Python which applications can interact with through an object-relational mapping or raw SQL commands. Two major features are (1) that complex queries are represented as left-deep plans where tuples are streamed through commands one at a time for significant gains in speed and smaller memory footprint and (2) that all commands are inherently in transactions and all changes to the database in memory must be explicitly committed to disk.
- 2014 **yahmm** python/cython
Feature rich implementation of hidden Markov models for Python, implemented in Cython for speed. Merged into pomegranate
- 2014 **PyPore** python/cython/sql
Data analysis, visualization, and storage of nanopore data. Written for Python with computationally heavy aspects implemented in cython and native integration with yahmm.
- 2014 **antz** python
Basic sequence analysis and visualization tools in Python for beginner bioinformaticians. Built to support new programming students taking BME 160: Introduction to Programming for Biochemists, where I was a TA twice.