





jacobschreiber

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

whoami

jmschreiber91@gmail.com 
@jmschrei 
@jmschreiber91 
jmschreiber91 

software

scikit-learn (core developer)
pomegranate (author)
PyPore (author)

programming

python, cython, sql, git

academics

machine/deep learning
probabilistic modeling
bioinformatics
bayesian networks
chromatin architecture
nanopore sequencing

awards

NSF IGERT Big Data Fellowship
NSF GFRP Honorable Mention
Dean's Award (UCSC Thesis)

education

since 2016 **Ph.D.** Computer Science University of Washington
Advisor: Dr. William Noble
2014-2016 **M.S.** Computer Science University of Washington
2009-2013 **B.S. Cum Laude** Biomolecular Engineering University of California, Santa Cruz

experience

2017 **Research Intern, Autopilot Maps, Tesla**
Reference: Nathan Jones <najones@tesla.com>
This internship focused on exploring new ways that machine learning can improve Tesla AutoPilot, and involved processing terabytes of fleet data, doing exploratory data analysis, and building working machine learning prototypes.

2016 **Research Intern, Data Science Team, AspenTech**
Reference: Mike Noskov <Mike.Noskov@aspentech.com>
This internship focused on implementing scalable and efficient probabilistic models for the analysis of large amounts of data.

2015 **Software Engineering Intern, Neurospin, INRIA**
Reference: Olivier Grisel <olivier.grisel@inria.fr>
This internship focused on speeding up the gradient boosting implementation in scikit-learn. My work ended up speeding up most tree-based estimators in addition to gradient boosting.

2011-2014 **Junior Specialist, Nanopore Group, University of California Santa Cruz**
Reference: Kevin Karplus <karplus@soe.ucsc.edu>
This position focused on using machine-learning methods to automate the analysis of large volumes of data, resulting in several published papers and a grant that got accepted from the NIH.

talks

2017	pomegranate: probabilistic modeling in python ODSC East, Tesla APM	INVITED
2017	Deep Learning Meets Chromatin Architecture UCSF	INVITED
2013-2016	Introduction to Machine Learning UCSC	INVITED
2017	Rambutan: Predicting 3D Genome Structure in Python GLBIO2017, scipy2017	
2016-2016	pomegranate: probabilistic modeling in python UW eScience, PyData Chicago, Moore-Sloan Data Science Summit, Seattle DAML Meetup, Data Intelligence, scipy2017	
2016	Large Scale HMMs For Nanopore Data Analysis UW eScience, UW ACMS Research Seminar	

publications

Schreiber, J. and Noble, W. Finding the optimal Bayesian network given a constraint graph. *PeerJ Computer Science*, 3:e122, 2017.

Schreiber, J., Libbrecht, M., Bilmes, J. and Noble, W. Nucleotide sequence and DNaseI sensitivity are predictive of 3D chromatin architecture. *Bioinformatics (Under Review)*, 2017. (1 citation as of 8/11/17)

Schreiber, J. and Karplus, K. Segmentation of noisy signals generated by a nanopore. *bioRxiv*, 2015. (3 citations as of 8/11/17)

Schreiber, J. and Karplus, K. Analysis of nanopore data using profile HMMs. *Bioinformatics*, 31(12) pp. 1897-1903, 2015. (10 citations as of 8/11/17)

Nivala, J., Mulrone, L., Li, G., **Schreiber, J.**, and Akeson, M. Discrimination among protein variants using an unfoldase-coupled nanopore. *ACS Nano*, 8(12) pp. 12365-12375, 2014. (20 citations as of 8/11/17)

Wescoe, Z., **Schreiber, J.**, and Akeson, M. Nanopores discriminate among five C5-cytosine variants in DNA. *JACS*, 136(47) pp. 16582-16587, 2014. (36 citations as of 8/11/17)

Schreiber, J., Wescoe, Z., Abu-Shumays, R., Vivian, J., Baldandorj, B., Karplus, K., and Akeson, M. Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. *PNAS*, 110(47) pp. 18910-18915, 2013. (82 citations as of 8/11/17)

software

pomegranate

pomegranate is a python package for flexible probabilistic modeling, implementing models such as Bayesian networks, hidden Markov models, and mixture models. Some major features include multi-threaded parallelism, GPU training, mini-batch training, semi-supervised learning, and out-of-core learning. I am the author and primary developer of pomegranate. (1,016 GitHub stars as of 8/11/17)

scikit-learn

scikit-learn is a popular python package for classic machine learning. In addition to implementing a wide variety of estimators, it also implements many data preprocessing methods and model evaluation techniques. I was invited to be a core developer for scikit-learn on 2/14/17. (20,502 GitHub stars as of 8/11/17)

PyPore

PyPore is a python package for the loading, processing, and visualization of nanopore generated data. It was written to support the research done in the UCSC Nanopore Lab. . I am the author of PyPore, though I do not actively develop it anymore. (15 GitHub stars as of 8/11/17)

Rambutan

Rambutan is a python package for the prediction of 3D genome structure at high resolution using a deep convolutional neural network. I am the author of Rambutan. (16 GitHub stars as of 8/11/17)

community

2017	SciPy 2017 Program Committee Member	SciPy
2017	Software Carpentry Instructor	UW eScience
2016	Google Summer of Code Mentor	scikit-learn
2016	Admissions Committee & CompBio Scheduling Chief	UW CSE
2015	Prospective Student Committee Co-Chair	UW CSE