



## education

### phd | machine learning

uc berkeley | '17-'22

research: interpretable ml

advisor: bin yu

collaborators:

s. upadhyayula

a. kornblith

### ms | data science

uc berkeley | '17-'21

research: statistical biology

### bs | cs & math

university of virginia | '14-'17

double major

## skills

machine learning

data cleaning • deep learning

pytorch • scikit-learn • jax

aws ec2 • s3 • sagemaker

keras • mllib • tensorflow

experienced

python • java • matlab

proficient

r • c/c++ • web basics

human languages

english • spanish • hindi

## awards

berkeley grad slam semifinalist '19

outstanding teaching award '18

uva rader research award '17

uva undergrad symposium winner '17

raven honor society '16-'17

icpc regional qualification '14-'16

1st place microsoft code jam '16

3rd place google games uva '17

2nd place apt puzzle competition '17

funding awards

pdsoros fellowship finalist '19

ircn workshop travel award '19

vidya shelat fund award '16

rodman scholarship '14-'17

## experience

### berkeley | interpretable ml research (bin yu lab %)

fall '17 - present

- developed interpretation methods for ml models (e.g. neural nets)
- developed interpretable models in medicine, biology, and computer vision

### paige ai | ai data scientist

summer '21

- interpretation and neural network modeling for medical imaging in oncology

### aws | research internship (pietro perona lab %)

summer '20

- testing for bias with causal matching using GANs
- interpreting semantic directions in generative models

### response4life | volunteer data scientist

spring '20

- helped develop, integrate, and deploy models to forecast covid-19 severity

### pacmed ai | interpretable ml internship

summer '19

- developed techniques to interpret machine-learning models for healthcare
- integrated interpretability techniques for predicting icu re-admission

### facebook | computer vision internship

summer '17

- investigated unsupervised deep learning for segmentation of satellite imagery
- implemented crfs for segmentation post-processing

### uva | ml research (yanjun qi lab %)

fall '16 – spring '17

- developed novel weighted- $\ell_1$ , multi-task gaussian graphical model
- analyzed large-scale functional brain connectivity with graphical models

### hhmi | ml research (srini turaga lab %)

summer '15, winter '15, summer '16

- extended cnns and watershed algorithms for neural image segmentation
- implemented distributed random forests for image segmentation

### uva | comp. neuroscience research (william levy lab %)

fall '14 - fall '16

- developed detailed biophysical models of neural computation
- analyzed energy efficiency, noise, and variability in stochastic neurons

### hhmi scientific computing | comp. neuroscience research

summer '14

- analyzed backpropagating action potentials via biophysical simulations

### research innovations inc. | web dev + android internship

summer '13 - spring '14

- developed web/mobile app for task coordination with qr codes

## coursework

### computation

machine learning  
computer vision  
structure learning  
algorithms  
artificial intelligence  
deep learning  
learning theory  
ai in graphics  
cs theory  
data structures  
software dev. I & II  
information retrieval  
computer architecture

### stat/math






statistical models  
probability  
statistics  
optimization  
linear algebra  
info theory  
real analysis  
linear models  
stochastic processes  
chaos theory I & II  
multivariate calculus  
discrete mathematics  
differential equations  
abstract algebra

### neuroscience




neural coding  
neural network models  
neurobiology  
visual neuroscience  
cognitive science

## selected publications





### interpretable deep learning

- adaptive wavelet distillation from neural networks through interpretations: ha, **cs**, et al. *neurips* '21 
- interpretations are useful: penalizing explanations to align neural networks with prior knowledge: rieger, **cs**, murdoch, & yu, *icml* '20 
- hierarchical interpretations for neural network predictions: **cs**\*, murdoch\*, & yu, *iclr* '19 
- transformation importance with applications to cosmology: **cs**\*, ha\*, lanusse, boehm, liu & yu, *iclr* '20 *workshop (spotlight talk)* 
- interpretable machine learning: definitions, methods, and applications: murdoch\*, **cs**\*, kumbier, abbasi-asl, & yu, *pnas* '19 




### interpretable rule-based modeling

- fast interpretable greedy-tree sums (figs): tan\*, **cs**\*, nasseri, agarwal, & yu *arxiv* '22 
- post-hoc regularization for decision trees and random forests *in prep* 
- disentangled attribution curves for interpreting random forests and boosted trees: devlin, **cs**, & yu *arxiv* '19 



### interpretable data-science

- curating a covid-19 data repository and forecasting county-level death counts in the united states: altieri, barter, ..., **cs**\*, ..., & yu\* *harvard data science review* '20  
- developing reliable clinical decision rules: a case study in identifying blunt abdominal trauma in children: kornblith\*, **cs**\*, et al. *seam abstract* 
- interpretable deep learning for accurate molecular partner prediction in clathrin-mediated endocytosis: **cs**\*, li\* et al. *in prep* 



### software packages

- imodels: a python package for interpretable modeling: **cs**\*, nasseri\*, tan, tang, & yu, *journal of open source software* '21  
- veridical-flow: a python package for facilitating stable data analysis: duncan\*, kapoor\*, agarwal\*, **cs**\*, & yu, *journal of open source software* '22 

### statistical neuroscience

- large scale image segmentation with structured-loss-based deep learning for connectome reconstruction: funke et al. *tpami* '18 
- a weighted- $\ell_1$ , multi-task graphical model with applications to heterogeneous brain connectivity: **cs**, wang, & qi, *neurips* '17 *amlicd workshop* 


### misc: causal inference, ml theory

- matched sample selection with GANs for mitigating attribute confounding: **cs**, balakrishnan, & perona *cvpr* '21 *civ workshop* 
- revisiting complexity and the bias-variance tradeoff: dwivedi\*, **cs**\*, yu, & wainwright *topml workshop* '21 

## teaching

### berkeley | student instructor

summer 2018

machine learning: cs 189/289 

lectures to class of 80+ students

fall 2019


artificial intelligence: cs 188 

## mini-projects

notes, blog, & slides 

hummingbird tracking 

news balancer django app 

java mini-games 

## service

basis education volunteering '19-'22

bair undergrad mentoring '18-'22

neurips reviewer '21

iclr workshop reviewer '21

cvpr reviewer '21

aaai xai workshop reviewer '21

neurips ml4h workshop reviewer '20

computer literacy volunteering '15-'17