

# Jean Feng

---

---

## Position

July 2020–**Assistant Professor In-Residence**, Department of Epidemiology and Biostatistics,  
Present *University of California, San Francisco*

---

## Education

Sept 2015– **PhD, Biostatistics**, *University of Washington*, Seattle, WA.  
June 2020 Advisors: Noah Simon, Frederick Albert Matsen IV  
2012–2013 **MS, Computer Science**, *Stanford University*, Stanford, CA.  
2009–2013 **BS, Computer Science**, *Stanford University*, Stanford, CA.

---

## Publications

Jean Feng, William S DeWitt, Aaron McKenna, Noah Simon, Amy Willis, and Frederick A Matsen. Estimation of cell lineage trees by maximum-likelihood phylogenetics. *Annals of Applied Statistics*, 2021.

Jean Feng. Learning to safely approve updates to machine learning algorithms. *Proceedings of the Conference on Health, Inference, and Learning*, 2021.

Brian D Williamson and Jean Feng. Efficient nonparametric statistical inference on population feature importance using shapley values. *International Conference on Machine Learning*, 2020.

Jean Feng and Noah Simon. An analysis of the cost of hyper-parameter selection via split-sample validation, with applications to penalized regression. *Statistica Sinica*, 2020.

Jean Feng, Scott Emerson, and Noah Simon. Approval policies for modifications to machine Learning-Based software as a medical device: A study of bio-creep. *Biometrics*, 2020.

Jean Feng, David A Shaw, Vladimir N Minin, Noah Simon, and Frederick A Matsen, IV. Survival analysis of DNA mutation motifs with penalized proportional hazards. *Ann. Appl. Stat.*, 2019.

Kristian Davidsen, Branden J Olson, William S DeWitt, 3rd, Jean Feng, Elias Harkins, Philip Bradley, and Frederick A Matsen, 4th. Deep generative models for T cell receptor protein sequences. *Elife*, 2019.

Jean Feng, Brian Williamson, Noah Simon, and Marco Carone. Nonparametric variable importance using an augmented neural network with multi-task learning. *International Conference on Machine Learning*, 2018.

Jean Feng and Noah Simon. Gradient-based regularization parameter selection for problems with nonsmooth penalty functions. *J. Comput. Graph. Stat.*, 2018.

## Pre-prints

Jean Feng and Noah Simon. Ensembled sparse-input hierarchical networks for high-dimensional datasets. *arXiv*, 2020.

Jean Feng, Arjun Sondhi, Jessica Perry, and Noah Simon. Selective prediction-set models with coverage guarantees. *arXiv*, 2019.

Jean Feng and Noah Simon. Sparse-Input neural networks for high-dimensional nonparametric regression and classification. *arXiv*, 2019.

## Funding History

2020-2022 UCSF-Stanford CERSI Program; Role: PI  
“Safe algorithmic change protocols for modifications to AI/ML-based Software as a Medical Device.”  
\$100,946 in direct costs

## Presentations

### Invited Oral Presentations and Seminars

- 2021 *Bayesian logistic regression for online recalibration and revision of risk prediction models with guarantees*, ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop
- 2021 *Variable Selection and Architecture Search for Neural Networks*, ASA Statistical Learning and Data Science Webinar
- 2021 *Bayesian logistic regression for online recalibration and revision of risk prediction models with guarantees*, Western North American Region (WNAR) Annual Meeting
- 2021 *Learning to safely approve updates to machine learning algorithms*, ACM Conference on Health, Inference, and Learning
- 2021 *Safe approval policies for continual learning systems in healthcare*, University of Waterloo
- 2020 *Efficient nonparametric statistical inference on population feature importance using Shapley values*, International Conference on Machine Learning
- 2020 *Training Procedures and Regulatory Policies for Safe Machine Learning Models in Healthcare*, University of California, San Francisco
- 2020 *Training Procedures and Regulatory Policies for Safe Machine Learning Models in Healthcare*, The University of Texas, MD Anderson Cancer Center
- 2020 *Training Procedures and Regulatory Policies for Safe Machine Learning Models in Healthcare*, University of California, Irvine
- 2020 *Approval policies for modifications to Machine Learning-Based Software as a Medical Device: A study of Bio-creep*, International Conference on Health Policy Statistics
- 2019 *Sparse-Input Neural Networks for High-dimensional Nonparametric Regression and Classification*, Western North American Region (WNAR) Annual Meeting
- 2018 *Nonparametric variable importance using an augmented neural network with multi-task learning*, International Conference on Machine Learning

- 2018 *Sparse-Input Neural Networks for High-dimensional Nonparametric Regression and Classification*, University of Washington Biostatistics Colloquium
- 2018 *Sparse-Input Neural Networks for High-dimensional Nonparametric Regression and Classification*, Joint Statistical Meetings
- 2017 *Sparse-Input Neural Networks for High-dimensional Nonparametric Regression*, ICML Workshop on Principled Approaches to Deep Learning
- 2011 *Haptic Belt with Pedestrian Detection*, Neural Information Processing Systems
- Contributed Oral Presentations
- 2019 *Uncertainty-Aware Black-Box Predictors with Coverage Guarantees*, Joint Statistical Meetings

## Awards

- 2020 University of Washington Thomas R. Fleming Excellence in Biostatistics Award
- 2020 International Conference on Health Policy Statistics, Student Travel Award  
For manuscript: *Approval policies for modifications to Machine Learning-Based Software as a Medical Device*
- 2018 Joint Statistical Meetings Section on Statistical Learning and Data Science, Student Paper Award  
For manuscript: *Sparse-input neural networks for high-dimensional nonparametric regression and classification*
- 2018 University of Washington Biostatistics Donovan J. Thompson Award for Best Combined Performance on Ph.D. Theory and Applied Qualifying Examinations
- 2015–2017 Big Data for Genomics and Neuroscience Training Grant

## Software

- EASIER-Net Python and R packages for fitting neural networks for high-dimensional data  
[https://github.com/jjfeng/easier\\_net](https://github.com/jjfeng/easier_net)  
[https://github.com/jjfeng/easier\\_net\\_R](https://github.com/jjfeng/easier_net_R)
- GapML Python package for analyzing cell-lineage tracing data from GESTALT  
<https://github.com/matsengrp/gestaltamania>
- SPINN Python package for estimating sparse-input neural networks  
<http://github.com/jjfeng/spinn>
- samm Python package for estimating somatic hypermutation rates of nucleotide motifs  
<http://github.com/matsengrp/samm>

## Teaching

- Spring 2021 Guest lecture, DATSCI 225: Machine Learning for the Biomedical Sciences II, UCSF
- Winter 2021 Instructor, Biostat 216: Machine Learning in R for the Biomedical Sciences, UCSF
- 2020 Instructor, Supervised statistical learning, 6th Seattle Symposium in Biostatistics
- 2020 Instructor, Machine Learning Boot Camp: Analyzing Biomedical and Health Data, Columbia University

- 2018 Teaching Assistant, Biostat 524: Design of Medical Studies, University of Washington
- 2017 Guest lecture, Biostat 561: Computational Skills for Biostatistics I, University of Washington
- 2016–2017 Teaching Assistant, Unsupervised Methods for Statistical Machine Learning, University of Washington Summer Institute in Statistics for Big Data
- 2016 Teaching Assistant, Supervised Methods for Statistical Machine Learning, University of Washington Summer Institute in Statistics for Big Data
- 2011 Section Leader, CS106A: Programming Methodology, Stanford University

## Service

### Referee Service

- International Conference on Machine Learning
- International Conference on Learning Representations
- Journal of Computational and Graphical Statistics
- Annals of Statistics
- Statistics in Medicine
- Neural Networks
- ASA Section on Statistical Learning and Data Science Student Paper Committee

### Session Chair

- Joint Statistical Meetings
- Western North American Region (WNAR) Annual Meeting

### University Service

- UCSF Initiative for Digital Transformation in Computational Biology & Health Grant Review Committee, September 2021
- Department of Epidemiology and Biostatistics, Digital Health Initiative Steering Committee, 2020–Present

## Work Experience

- 2019 **Research Intern**, *Insitro*, South San Francisco, CA.  
Developed statistical models of genomic data.
- 2012–2015 **Software engineer**, *Coursera*, Mountain View, CA.  
Built the professional certificate program and payment system. Technical lead on projects with 3-5 people. Mentored interns and junior engineers.