Brian Gereke, Ph.D.

Portland • OR | 520.256.9613 | bgereke89@gmail.com
bgereke.github.io | github.com/bgereke | linkedin.com/in/brian-gereke

Experience

Insight Data Science Fellowship | **San Francisco Bay Area** **2020**

Postdoctoral Researcher | **Lawrence Berkeley National Lab** **2018–2020**

- Developed an object-oriented, multi-process, multi-module, gpu-accelerated data acquisition system using Arduino, Python (e.g., tkinter, multiprocessing, opencv, numpy, etc.), and cuda-enabled ffmpeg. The system tracked all experimental metadata, persisted config settings across experiments, minimized experiment setup time, and reduced experimenter errors. All code documented w/ sphinx on readthedocs.
- Analytically derived robot forward/inverse kinematics, developed state-space sampling routines for data-driven control calibration, and a novel nonlinear PID controller using generalized additive models and shaped-constrained additive models deployed on Arduino. The solution achieved submillimeter position accuracy and provided an interpretable, maintainable mvp.
- Managed team that benchmarked a markerless behavior 3D tracking system using multi-view video data, DeepLabCut (tensorflow package), and DLT (camera calibration and 3D reconstruction).
- Designed a 3-dof pneumatically actuated robotic system for automating complex reaching tasks in rodents.

Graduate Student Researcher | **University of Texas at Austin** **2012–2018**

- Led lab's transition to open source HPC resources, eliminating previous commercial licensing costs. Performed batch analyses (i.e., slurm jobs) on large scale rodent hippocampal neural spiking, calcium imaging, and local field potential data, mostly using generalized additive mixed models in R. Optimized and benchmarked memory usage locally to further optimize compute costs on the cluster.
- Revised lab-wide codebase for crucial processing routines such as kernel density estimators, continuous wavelet transforms, confidence interval estimators, and multivariate connectivity metrics.
- Developed novel conditional Tweedie point process models for sparse time-frequency features with applications to hippocampal field potential data.

Skills

- **Python (5 yrs):** object-oriented + functional | modules | pep8 | numpy | pandas | scikit-learn | statsmodels | scipy | pyspark | opencv | faiss | tkinter | flask | fastapi | requests | multiprocessing | sphinx | unittest | pytest
- **R (5 yrs):** object-oriented + functional | mgcv | glmnet | dplyr | ggplot2 | rgl | r markdown | dbi
- **Other tools:** matlab (10 yrs) | arduino (3 yrs) | sql | git | databricks | xgboost | tensorflow | pytorch | ffmpeg | docker | kubeflow | aws (iam, ec2, s3, eks) | javascript | node.js | vue.js | webpack | html | css | vuetify | bootstrap | oauth2 | axios | fetch | postman | regex | chrome extensions | github pages | jekyll liquid
- **Methods:** time-frequency analysis | time series analysis | gamm | gradient boosted trees | convolutional networks | numerical optimization | dimensionality reduction | clustering | approximate nearest neighbors | point processes | computer vision | regularization techniques | bootstrap, permutation methods

Open Source Projects

The R Inside Airbnb **2020**

- Makes "insideairbnb.com" queryable by building PostgreSQL database(s) on-the-fly. Scrapes data using regex and piped bash curl commands directly from postgres to minimize redundant data transfer and improve query performance.
- Provides a library for cleaning and preprocessing scraped data (i.e., dealing w/ missing values, coding of numerical/categorical variables, etc.). All functions documented w/ examples using Roxygen2.
- Performs exploratory data analysis (i.e., variable distributions, correlations, etc.), temporal and spectral clustering of listing price dynamics using tadpole (pam + dtw), and predictive modeling of listing prices and occupancies using xgboost. A detailed r markdown notebook walkthrough is provided.

Candlelight Visual Search **2020**

- A Vue.js chrome extension that performs screen capture, image cropping, and reverse image search in the browser similar to Pinterest's "flashlight" feature. Reverse image search can be performed using the google cloud vision api (oauth2 required), bing visual search api, or custom candlelight api (in progress).
- Achieved 2-3% improvement in recall@1 over Pinterest's published state-of-the-art image embedder (pretrained resnet50 + layer normalization + class balanced minibatch sampling) with similar latency and memory costs by replacing the backbone with the newer resnest50.

Education

PhD, University of Texas at Austin | **Neuroscience** **2012-2018**

BS, University of Arizona | **Applied Mathematics** **2007-2012**