Zelong/Alan Wang

Contact

Address

La Jolla, CA, 92122

Phone

858-539-5564

E-mail

zew013@ucsd.edu

Personal/GitHub/LinkedIn

zew013.github.io

Test Taken

GRE Verbal: 162

GRE Quantitative: 169

Scholarship & Honor

Two-year Provost Honors

Skills

- Python — Proficient
- Stata — Very Good
- R — Good
- Microsoft Office Suit — Good
- SQL/PostgreSQL — Good
- Java — Average
- HTML — Average
- Git/Github — Average
- Tableau — Average
- LaTex — Average
- AWS — Some Experience
- Heroku — Some Experience
- MATLAB — Some Experience
- Mandarin — Excellent

Python Packages

Pandas, NumPy, bs3 Matplotlib, SciPy, Plotly, Folium, Scikit-Learn, PyTorch, Flask, **and a lot more**.

Education

The University of California San Diego | 2020-2024 (expected) | GPA:3.98

B.S | First Major - Data Science | Double Major - Joint Major in Math and Economics

Data Science Related: Python, Java, HTML, Data Structures, Data Science Algorithms, Web Scraping, Data Mining, Visualization, Machine/Deep Learning, Recommender System, Network Science & Graph Theory, Data Base System.
Note: Will learn big data and systems such as MapReduce/Hadoop and Spark systematically this Winter.

Math & Econ Related: Calculus, Linear Algebra, Numerical Analysis, Optimization, Probability, Statistics, Micro & Macroeconomics, Econometrics (t, z, Chi-squared, F test, Regressions, panel data, etc.), Operations Research.

Experience

Student Researcher | Professor Dale Squires and Professor Richard Carson | July 2022 – Present

- Explore membership assessment of international orgs. and marginal utility of consumption to measure ethical preference.
- Read historical documents and meeting reports to filter relevant data. Built pipeline to automatically extract and prepare data from low-resolution UN reports (1950-2022) using Python and AWS lambda and S3 bucket.

Tech Consulting Intern | IBM Consulting, China | Aug. 2022 – Sep. 2022

- Provided reports about a consolidated cement company for the team. Collaborated with cross-functional stakeholders, developed scalable solution and worked towards consensus across the organization. Built metrics to inform performance.

CSE-PACE Program Designer | UCSD CSE Department | Apr. 2022 – Sep. 2022

- Searched and reviewed a miscellaneous assortment of papers, articles, programs, and even games for course topics.
- Designed CSE cohort programs and built 7 courses on various Computer Science/Data Science related topics.

Research Assistant | UCSD Professor Richard Carson | Dec. 2021-Present

- Did literature review and gained background knowledge on discrete choice models and related topics.
- Arranged and corrected research data to create representative graphs intended for academic paper. (approx. 60 drafts)
- Coded Monte Carlo Simulation, conditional logit (max/minimum Gumbel assumption). Performed qualitative analysis.

Data Analyst/Tech VP | Lumnus Consulting (Student Enterprise) | Nov. 2021-Present

- Collaborated with other leaders to identify and prioritize problems. Organized team events and fostered team communication. Responsible for website maintenance using React.js and Heroku.
- Built prediction models based on historical Instagram and Twitter data. Created visualizations and analyses for the team. Delivered project presentations on how we should present our organization on social media.

Tutor & Math Team Leader | 2019 – 2021, Data Manager & Robot Builder | Robotics Team | 2018 – 2020

Self-Driven Projects (more detailed description and examples can be found on LinkedIn project session)

Basic Multi-Layer Neural Network | NumPy, Git

- Use only **NumPy** to implement **Backprop**, **Mini-Batch Gradient Descent**, **Cross Validation**. Added **Image-Normalize**, **Early Stop**, **Momentum**, **Regularization** to improve. Support sigmoid, tanh, ReLU, softmax as activation.

More Deep learning: Image Captioning using CNNs and LSTMs on COCO Dataset
Amazon Massive Dataset Intent Classification with Transformers and deployed API on the website

Forage Data Analytics Virtual Experience Programs | Python, Tableau, Excel, Git

- Automatic EDA with **SweetViz**. Assigned price sensitivity to each customer. Tested hypothesis by **permutation test**, **K-S test**, and heat map. Predicted customer churn by **Random Forest** and **XGBoost**. Storytelling with **Tableau**.

Analysis of Power Outage Status in the Continental U.S. | Python, Excel

- Went through the full process of questioning, researching, data cleaning & EDA, missingness assessment, hypothesis test, baseline model, **scikit-learn** ML pipelines, final model, and fairness analysis.

Indian Crop Production and Indian Climate Analysis | Python, Slides

- Read reports about Indian agriculture, pastoralism, and climates. Reported findings and provided suggestions.
- Data gathering, data mining, interactive visualization using **bs4**, **JSON**, **Geopandas**, **Altair**, **Folium**, **Plotly**.

Statistical Language Models using Public Domain Books | Python

- **Web scraping** for specific books from a public book website. Tokenize corpora.
- Created an **N-Gram Language Model** that can generate paragraphs resembling the style of an author or a book.

Wealth Prediction based on 1991 Survey of Income and Program Participation (SIPP) | R

- Applied EDA, feature engineering. Incorporated and compared **polynomials** and **splines** in **GAM**. Tested **full OLS**, **Lasso**, **Ridge**, **Stepwise regression model** and **Random Forest model**.