Simao Eduardo

Graduating PhD in **Machine Learning** from **University of Edinburgh.** Interests in Generative Models, Bayesian Inference, Deep Learning, Robust ML. Past work in Optimization, Image & Signal Processing.

Portuguese Citizen (EU)
Settled Status (UK)
Mail: simao.eduardo.ist@gmail.com
Website: sfme.github.io
Github: github.com/sfme

EXPERIENCE

ML Research Intern - *Hazy, UK (Remote in Portugal)*

October 2022 - February 2023

Machine Learning engineering tasks, mostly on data preprocessing pipeline of the product. **Skills**: Python; Pandas; Numpy; Scikit-Learn; PyTorch. **ML theory:** Generative Models; Differential-Privacy; Metrics. **Tools**: pytest (testing); CircleCI (continuous integration); Git / GitHub. **References: Georgi Ganev (Supervisor), Carl Tishler (HR).**

PhD Student Researcher - *University of Edinburgh, UK*

October 2016 - March 2023

School of Informatics. **Machine Learning PhD Thesis: 'Data Cleaning with Variational Autoencoders'**. **ML knowledge:** Generative Models, Bayesian Inference, Deep Learning. **Skills:** Python, SQL, JAVA, PyTorch, Scipy, Numpy, Pandas, Matplotlib, Git. **References: Charles Sutton (Supervisor).**

Tutor / Marker - *University of Edinburgh, UK*

October 2016 - December 2018

Taught postgraduate courses: tutorials & labs. Machine Learning and Pattern Recognition, Machine Learning Practical (Deep Learning), Extreme Computing (MapReduce, HDFS). **Skills:** Python, JAVA.

Research Assistant Fellow - *Instituto de Telecomunicações at Instituto Superior Tecnico (IST), Lisbon PT*

December 2014 - August 2015

Researched Proximal Algorithms (Optimization) for Inverse Problems in Imaging: **Segmentation and Deblurring in Super-Resolution Images**. Image Processing tasks. **Skills:** MATLAB, Latex. **References: Mario Figueiredo.**

Junior GNSS Software Engineer, *GMV Innovating Solutions, Lisbon PT*

May 2013 - May 2014

Developed **GNSS receiver simulator for assessing robust carrier phase techniques** (**Digital Signal Processing**, DSP) for the **European Space Agency**. Integrated novel **DSP and statistical methods** (time-series). Analysis and reporting. **Skills:** MATLAB, C, MS Word. **Reference: Teresa Ferreira.**

EDUCATION

SKILLS / KNOWLEDGE

Programming: Python, SQL, MATLAB, C, JAVA, Bash

Framework: PyTorch, Scipy, Numpy, Git, Pandas, SQLite, Tensorflow (basic), Scikit-Learn, Jupyter Notebook, AWS, Matplotlib, Seaborn

Math: Machine Learning, Optimization, Signal Processing, Time-Series (Kalman Filters, Autoregressive models), Bayesian Inference (MCMC, Variational Inference), Deep Learning, Outlier Detection, NLP (very basic), Decision Trees, Random Forests, Clustering, Classification, Regression

AWARDS

EPSRC CDT Scholarship, funding 2015-2019, EP/L106427/1

Research Assistant Fellowship, PAConvex project, 2014-2015

Erasmus Scholarship, MSc Thesis at Aalborg University, 2011-2012

MSc(R) Data Science (Distinction), University of Edinburgh

September 2015 - September 2016

Thesis: **'Data Cleaning in Tabular Datasets: Error Detection and Robust Rule Estimation'**, Supervisors: **Charles Sutton**

MSc Electrical and Computer Engineering (A- top 7%; 17/20), *Instituto Superior Tecnico, University of Lisbon*

September 2010 - January 2013

Thesis:**'Cell load balancing in heterogeneous scenarios: A 3GPP LTE case study'** at ***University of Aalborg***, Supervisors: **Antonio Rodrigues** (IST), Albena Mihovska (University of Aalborg)

BSc Electrical and Computer Engineering (A- top 6%; 16/20), *Instituto Superior Tecnico, University of Lisbon*

September 2007 - August 2010

REFERENCES

Georgi Ganev - georgi@hazy.com

Carl Tishler - carl@hazy.com

Dr. Charles Sutton - csutton@inf.ed.ac.uk / charlessutton@google.com

Prof. Mario Figueiredo - mtf@lx.it.pt

Teresa Ferreira - teresa.ferreira@gmv.com

PUBLICATIONS / REPORTS

Repairing Systematic Outliers by Learning Clean Subspaces in VAEs - *preprint (arXiv)*

2022

Simao Eduardo, Kai Xu, Alfredo Nazabal, Charles Sutton
https://arxiv.org/pdf/2207.08050.pdf

Robust Variational Autoencoders for Outlier Detection in Mixed-Type Data - *AISTATS 2020*

v1: 2019, v2: 2020

Simao Eduardo, Alfredo Nazabal, Chris Williams, Charles Sutton
https://arxiv.org/pdf/1907.06671.pdf

Data Cleaning using Probabilistic Models of Integrity Constraints - *NIPS 2016 Workshop on Artificial Intelligence for Data Science (AI4DataSci 2016)*

2016

Simao Eduardo, Charles Sutton
http://workshops.inf.ed.ac.uk/nips2016-ai4datasci/papers/NIPS2016-AI4DataSci_paper_12.pdf

Proximal Algorithms in Inverse Imaging Problems: First and Second-Order Splitting - *Technical Report*

2015

Simao Eduardo; Supervised by: Mario Figueiredo; JM Bioucas-Dias
https://sfme.github.io/TechReport_Proximal_Algos_2015_IST.pdf

Cell load balancing in heterogeneous scenarios: A 3GPP LTE case study - *Wireless VITAE 2013, IEEE*

2013

Simao Eduardo, Antonio Rodrigues, Albena Mihovska, Neeli R. Prasad
https://ieeexplore.ieee.org/document/6617063?arnumber=6617063