

Syamantak Kumar

Austin, Texas • +1-737-336-5714
syamantak@utexas.edu • linkedin.com/in/syamantak-kumar

Education

University of Texas at Austin <i>Ph.D. in Computer Science</i>	Aug 2023 - May 2027 GPA : 3.97/4.0
University of Texas at Austin <i>Master of Science in Computer Science</i>	Aug 2021 - May 2023 GPA : 3.97/4.0
Indian Institute of Technology Bombay <i>Bachelor of Technology in Computer Science with Honours</i>	July 2016 - May 2020 GPA : 9.35/10.0

Research Interests

Broadly, my research interests lie at the intersection of statistics and machine learning. I am interested in designing machine learning algorithms with provable guarantees of convergence and correctness. I also want to work on quantifying the computational and statistical aspects of state-of-the-art deep learning methods. I deeply enjoy studying the applications of high-dimensional statistics, optimization, and probability theory to real-world problems. I am currently advised by Prof. Purnamrita Sarkar.

Work Experience

Microsoft, Mountain View | *Applied Scientist Intern* | *Core Search and AI* **May 2022 - Aug 2022**

- Worked on Dense Information Retrieval Models for Bing Search
- Experimented with early and late fusion architectures to perform Transformer Mixing of Numerical features such as PageRank with Text features for query and document encoders to enhance recall by 0.5 points
- Explored query topic annotation features along with numerical features and performed statistical analysis of the model scores to derive key insights into the predictions

Google, Bengaluru | *Software Engineer* **Aug 2020 - July 2021**

- Designed algorithms to generate key signals to detect Hijack and Disguise of genuine businesses on Google Maps
- Trained lattice regression models for abuse classification based on signals provided by Google Local Search
- Performed extensive SQL based analyses to derive key insights and statistically verify proposed algorithms
- Implemented multiple data processing pipelines in C++ using a MapReduce framework for productionization
- Led discussions with cross-functional teams comprising of 10+ members each, to increase the quality of manual operator labels for abuse classification, with a direct revenue impact

Google, Bengaluru | *Software Engineering Intern* **May 2019 - July 2019**

- Built a generative model to predict the probability of users making a contribution to Maps
- Modeled user-contributions as a poisson process and performed estimation with poisson regression
- Inferred labels on unlabelled user edits by aggregating future contributions on each place on Maps

Publications

- Streaming PCA for Markovian Data [paper]
Syamantak Kumar, Purnamrita Sarkar
Spotlight (Top ~ 3% of accepted papers) Presentation at NeurIPS 2023, New Orleans, Louisiana
- A comparison of open source libraries ready for 3D reconstruction of wounds [paper]
Syamantak Kumar, Dhruv Jaglan, Nagarajan Ganapathy and Thomas Deserno
Oral Presentation at Proceedings of SPIE Medical Imaging 2019, San Diego, California
- From Machine Translation to Code-Switching: Generating High-Quality Code-Switched Text [paper]
Ishan Tarunesh, **Syamantak Kumar** and Preethi Jyothi
Oral Presentation at Proceedings of ACL-IJCNLP 2021, Bangkok, Thailand

Research Experience

Feature Extraction for Gaussian Process Regression **July 2019 - July 2020**
Guide: Prof. Suyash Awate | Bachelor's Thesis *IIT Bombay*

- Worked on combining the rich feature-extraction capabilities of deep neural networks with the explainability and flexibility of Gaussian Process Regression models to create an end-to-end training framework.
- Compared the framework's advantages over other techniques on various regression problems available in the medical domain, such as cell-counting in whole-slide images and survival prediction using brain MRI scans
- Achieved an MAE of 8 on VGG cells dataset for cell-counting, and an accuracy of 40% on the MICCAI BRaTS 2020 dataset

Machine Translation for Code-Switching | Research Project

Guide: Prof. Preethi Jyothi | Research Project

Jan-July 2020

IIT Bombay

- Our work proposed a curriculum for training a transformer based machine translation model to generate code-switched text
- Implemented and developed metrics for evaluation of model-generated code-switched text for perplexity, diversity and text-translation quality
- Experimented with mBERT embeddings for hindi-english code-switched text, pre-training on a multilingual corpus and finetuning on smaller code-switched text to achieve an improvement of ~40 per word in language model perplexity

3D Reconstruction of Wounds | Assessment and Diagnosis

Guide: Prof. Thomas Deserno | Summer Internship

May-July 2018

TU Braunschweig, Germany

- Developed an android application to extract 3D point clouds by scanning objects using ARCore, Google's platform for building augmented reality experiences on smartphones
- Compared the results obtained from ARCore with the Structure from Motion (SfM) technique to review efficacy of smartphones in generating 3D models for wound assesment

Key Projects

- **Adversarial Examples for Keyword Spotting** : Used AdvGAN to generate adversarial examples for keyword spotting and enhance its robustness, with a 2% lift in classification accuracy
- **Top-k Tournament Ranking** : Developed an algorithm for fully sequential sampling for tournament ranking by modelling it as a stochastic multi-armed bandit problem
- **Artistic Style Transfer** : Implemented an expectation-maximization based algorithm for style transfer between images
- **RL for Cache admission and Replacement** : Modelled cache admission as a Markov Decision Process and used standard Actor-Critic models with tile-coding to achieve a gain of 36% in the overall byte hit rate. Modelled cache replacement as an adversarial bandit problem instance and demonstrated optimality of LFU for cache traces drawn from zip-f distributions.
- **A Probabilistic Formulation of Unsupervised Text Style Transfer** : Performed a reproducibility study to implement a deep latent sequence model that performs unsupervised style transfer. By hypothesizing a parallel latent sequence that generates each observed sequence, the model learned to transform sequences from one domain to another in a completely unsupervised fashion.

Teaching Experience

- **Graduate Teaching Assistant** -
 - CS 361S - Network Security and Privacy, with Professor Seth Nielson at UT Austin for Fall 2021
 - EE 461P - Data Science Principles, with Professor Joydeep Ghosh at UT Austin for Spring 2022
- **Undergraduate Teaching Assistant** - Among the **20 students** selected across all batches for teaching a class of **47** first-year students for the undergraduate course on Quantum Physics. Coordinated with the Physics Dept. to conduct regular tutorial sessions & evaluate exam papers
- **Teaching Assistant** - Selected as a Teaching Assistant for the Sandbox@Alphabet team's "Linear Algebra for Machine Learning" course, engaging in a chat room of 700+ participants and leading small independent break-out sessions with attendees to discuss problems and exercises.

Technical Skills & Coursework

Languages : C++, C, R, Python, Java, Bash, MATLAB, Racket/Scheme, Prolog, SQL

Developer Tools : Tensorflow, Pytorch, Google Colab, Jupyter Notebooks, Android Studio, Git, Kaldi

Relevant Courses : AI and Machine Learning, Automatic Speech Recognition, Natural Language Processing, Reinforcement Learning, Advanced Machine Learning, Information Retrieval and Web Mining, Digital Image Processing