Gurpreet Singh

(646) 541-8879 | gurpreet.s@columbia.edu | github/ffs97 | linkedin/ffs97 | 505 West 122nd Street, New York, NY

Education

COLUMBIA UNIVERSITY *New York, NY*
MASTER OF SCIENCE IN COMPUTER SCIENCE (MACHINE LEARNING TRACK) | GPA 4.0 / 4.0 *Sep 2019 - Dec 2020 (Expected)*
Coursework: Optimization Methods, Artificial Intelligence, Probabilistic Programming
- Course Assistant: Optimization Methods in Machine Learning (*Fall 2019*), and Program Languages and Translators (*Fall 2019*)
- Research Assistant: Prof. Michael D. Parrott (*Spring 2020*); Worked on building ONNX models using common ML frameworks

INDIAN INSITITUTE OF TECHNOLOGY KANPUR *Kanpur, India*
BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE | GPA 9.3 / 10.0 *Jul 2015 - May 2019*
Coursework: Probabilistic Modeling and Inference, Learning Theory, Machine Learning, Natural Language Processing, Computer Vision
- Tutor: Fundamentals of Computing (*Spring 2019*), and Course Mentor: Introduction to Machine Learning (*Fall 2018*)
- Awarded the Academic Excellence Award at IITK for two consecutive years from 2016–18
- Secured an All India Rank (AIR) of 174 among 1.5 million candidates in JEE Advanced, 2015

Technical Skills

Machine Learning PyTorch, Tensorflow, Pyro, Keras, Scikit-Learn, NumPy
Programming/Scripting Python, C/C++, Bash, Octave/MATLAB

Work Experience

NUTANIX *San Jose, CA*
MACHINE LEARNING ENGINEERING INTERN *May 2020 — Aug 2020*
- The objective was to implement robust Machine Learning models for anomaly detection in operations performed on large file servers
- Built a variant of the Holt-Winters Model for temporal based anomaly detection on per hour operation counts
- Implemented and used Collaborative Filtering for learning user behaviour using peer comparison for file access patterns

GOLDMAN SACHS *Bangalore, India*
SUMMER ANALYST *May 2018 — Jul 2018*
- Built a lite calculator for customer margin allocation for proper internalization based on a waterfall logic, taking various parameters into consideration and a greedy strategy to allocate margin per stock, improving the run time of the allocation logic by tenfolds
- Introduced changes in existing models for asset liability gap management identifying bugs in pre-existing code and augmenting features for proper AL Gap Management for deposits and clearing house initial margin

INMOBI *Bangalore, India*
DATA SCIENCE INTERN *May 2017 — Jul 2017*
- Extracted image features from Ad creative images using OpenCV (in python) and Google Cloud Vision API to explain ad CTR
- Analysed Pearson Correlation, performed regression tests, and used variable selection (Weka) to detect explainable features
- Created a python server to handle feature extraction and prediction for building suggestions for possible Ad enhancements based on CTR

Projects

BAYESIAN ML FOR PREDICTING UNDERGROUND WATER LEVELS *COMS 6998: Probabilisitic Programming, CU*
DR. ALP KUCUKELBIR (FERO LABS) *Sep 2019 — Nov 2019*
- Modeled underground water levels in a sub-region in Rajasthan (India) based on readings from sparse observatory wells
- Used a HMM-modulated kernel regression to model temporal and spatial patterns dependent on satellite observations of farmland

IMPLEMENTING AND ANALYSIS OF ADAPTIVE LEARNING OF NEURAL NETWORKS *COMS 4995: Optimization Methods, CU*
DR. SATYEN KALE (GOOGLE RESEARCH) *Sep 2019 — Nov 2019*
- Implemented AdaNet using PyTorch and analysed its performance with different settings of hyperparameters
- Improved hyperparameter sensitivity by adaptively changing the subnetwork width without compromising on performance

DISCRETE VAES AND STOCHASTIC BLOCK MODELS *Undergraduate Project, IITK*
PROF. PIYUSH RAI *Aug 2018 — Jan 2019*
- Surveyed continuous relaxations to discrete latent variables and implemented GumBolt relaxation for RBM prior using tensorflow
- Augmented GVAEs with binary latent embeddings to offer interpretable latent representations, imitating mixed membership models
- Employed the resultant model for link prediction on graph datasets (Citeseer and Cora) and achieved superior results to baseline models

INCREMENTAL NEURAL NETWORKS TRAINING *CS777: Learning Theory, IITK*
PROF. PURUSHOTTAM KAR *Jan 2018 — Apr 2018*
- Employed the concept of training ensembles using gradient boosting for training two layer networks by representing two layer networks as an ensemble of single hidden node networks affording definite theoretical convergence guarantees
- Studied the convergence analysis of incremental training under various constraints and assumptions
- Applied incremental training as pre-training, along with backpropagation for fine-tuning, and observed remarkably better convergence