Ninad Khargonkar

CURRICULUM VITAE

Sunderland, Massachusetts – 01375
nkhargonkar@umass.edu | ninad.khargonkar@gmail.com
Webpage: ninception.github.io
LinkedIn: linkedin.com/in/ninadkhargonkar
Phone: 413-345-9601

EDUCATION

University of Massachusetts Amherst
Master of Science, Computer Science   *Expected May' 19*
GPA: 4.0/4.0

Indian Institute of Technology Kanpur
Bachelor of Science, Mathematics and Scientific Computing   *Graduated June' 17*
GPA: 8.4/10

INTERESTS

Machine Learning and Probabilistic Modelling and its applications in Medicine/Healthcare, Natural Language and Speech Processing.

AWARDS & ACHIEVEMENTS

Recipient of **Inspire** scholarship awarded by Dept. of Science and Technology (Govt. of India).
Awarded **Mitacs Globalink** scholarship for summer research internship in Canada.
Secured a percentile score of **97.7** in JEE (Advanced)–2013 out of 150,000 candidates.
Secured a percentile score of **99.8** in JEE (Main)–2013 out of about 1.3 million candidates.
Merit Certificates from the Central Board Of Secondary Education (CBSE) – 2011, 2013.

RESEARCH EXPERIENCE

Chronic Disease Modelling
Supervisor : Prof. Peter Haas — UMass Amherst   *Jun '18 - Aug '18*

- Worked on the problem of learning a distribution from sparse data set of disease prevalence and also modelling the transition of disease states across a year for 9 diseases.
- The principle of maximum entropy was used to infer the distribution with the marginals for individual diseases along with some top (ranked) feature pairs acting as the constraints to the optimization problem.
- The pipeline of data loading, feature selection and optimization for the data set was implemented in Python.

Space Efficient Dynamic Programming for LIS
Supervisor : Prof. Barna Saha — UMass Amherst   *Apr '18 - Jun '18*

- Designed, analyzed and implemented space efficient algorithms (approximate but deterministic) in dynamic programming for the longest increasing subsequence problem.
- Using the approach of selective memory/forgetting, I tried to design a non-randomized approximate algorithm to improve the space and time bound from an earlier work by Prof. Saha.
- For the approximate algorithms, I proved algorithmic guarantees about an improved approximation factor and implemented the deterministic and randomized algorithms using Numpy in Python.

Globalink Research Internship
Prof. Saman Muthukumarana — University of Manitoba, Winnipeg   *May '16 - Jul '16*

- Studied the problem of sampling from large graphs and extracting relevant graph statistics like the clustering coefficient in R.
- Scale down sampling was implemented on graphs by using random walk variants like Metropolis-Hastings random walk and Jump-random walks and their performane analyzed.
- Statistical models like ERGM were used for producing model fits and simulating random networks for testing significance of network substructures like triangles.

Bayesian Estimation and Sampling
Prof. Debasis Kundu — IIT Kanpur <div align="right">*Jan '16 - Apr '16*</div>

- Studied the usage of Bayes theorem and the difficulties in Bayesian inference versus Frequentist analysis along with numerical techniques for sampling and bayes estimation from a weibull distribution.
- Fundamentals of sampling like Monte Carlo methods and Importance sampling were studied along with the techniques for random variable generation for a probability distribution.

COURSE
PROJECTS

Part of Speech Tagger for Hindi
Prof. Brendan O'Connor — University of Massachusetts, Amherst <div align="right">*Sep '17 - Dec '17*</div>

- Developed a part of speech (POS) tagger for Hindi by learning a word level translation between Hindi & English using word embeddings and a parallel corpora between the 2 languages.
- No supervised part of speech information was provided during training and the 300 dimensional fast-Text mono lingual word embeddings were used for training.
- The tagger also improved the performance by 12% on the proxy task of domain classification of documents in Hindi when the pos labels were added as an additional feature to an LSTM based model.

Modelling Uncertainty in Deep Learning
Prof. Erik Learned-Miller — University of Massachusetts, Amherst <div align="right">*Sep '17 - Dec '17*</div>

- Analyzed the problem of modelling uncertainty in the results from the outputs/scores of a deep neural network.
- Coded a neural network in Keras (Python) for the uncertainty value calculation on cifar-10 and mnist datasets.
- The uncertainty estimates also improved the performance of a linear classifier by around 2% in detecting out of training set images (proxy for anomaly detection) when added as additional features.

Depth from Monocular Images
Prof. Subhransu Maji — University of Massachusetts, Amherst <div align="right">*Sep '17 - Dec '17*</div>

- Depth map prediction at test time in monocular images was targeted. An unsupervised approach to train using concepts from stereo vision was applied.
- Stereo image pairs were used during training to learn a disparity and the depth map was inferred from the disparity map. No ground truth depth data was used while training.

Topic Modelling of Documents – LDA
Prof. Piyush Rai — IIT Kanpur <div align="right">*Jan '17 - Apr '17*</div>

- Implemented a document-topic categorization model on the Reuters news agency data set in Python (gensim-lib).
- Latent Dirichlet Allocation (LDA) statistical model was used for categorization and Variational Inference along with its stochastic extension was used for scaling up the inference process for the 7700 documents in training set.

Automated Image Colorization
Prof. Piyush Rai — IIT Kanpur <div align="right">*Jul '16 - Nov '16*</div>

- Implemented an automatic colorization model by predicting the per pixel colour histogram using the idea of hyper columns from a Convolutional Neural Network (VGG-16 pretrained on ImageNet).
- The network was modified for gray scale image by averaging over colour channels and l_2 regularized loss function was chosen between network output and ground truth. TensorFlow and Keras (Python) were used for coding.

High Dimensional Anomaly Detection
Prof. Purshottam Kar — IIT Kanpur <div align="right">*Jul '16 - Nov '16*</div>

- Anomaly detection in high dimensional data was studied using classification and clustering based methods like K-nn and 1-SVM. Matlab package DeeBNet was used for experiments.
- Implemented a hybrid model using a Deep belief network to extract features and then training one class SVM on them. Explored model extensions for streaming data in online setting.

Football tournament simulation model

Prof. Saman Muthukumarana — University of Manitoba, Winnipeg   *Jun '16 - Jul '16*

- A simulation model for a football tournament was also developed using simple belief networks to analyze any bias or faults in the tournament group structure.

- Some biases were found in the new tournament structure which gave a higher probability of qualification to next rounds for certain teams. Past game data was used to simulate the outcome of football matches.

OTHER EXPERIENCE

Programmer – BioStats Lab

Dept. of Biostatistics — University of Massachusetts, Amherst   *Sep '17 - Nov '17*

Worked with Prof. Alkema's lab on a project on statistical indicators for family planning in developing countries and assisted in coding the simulation exercises and modelling the experiments in R programming language.

Course assistant

Prof. Peter Haas — University of Massachusetts, Amherst   *Jan '18 - May '18*

Grader for a graduate level Simulation course by Prof. Peter Haas. Responsible for grading of assignments and exams, setting up the submission platform and clearing any doubts in the assignments.

Student Guide

Counselling Service — IIT Kanpur   *Jun '14 - Jul '15*

Mentored a group of 6 freshmen and helped them in their induction into the institute. Volunteered in conducting talks and workshops for guidance in academic and cultural issues. Supervised the orientation and registration process of about 850 new students with the counselling service.

CONFERENCE PUBLICATIONS

Saman Muthukumarana, **Ninad Khargonkar**, "Modeling and Simulation of UEFA tournament" in *Section on Statistics in Sports of Statistical Modeling for Sports Science and Applications activity of JSM 2017 Online program*, Baltimore, USA.

TECHNICAL SKILLS

Languages: Python, Java, C/C++, R, Matlab, SQL
Frameworks/Libs: Apache Spark, AWS, , Numpy, Scikit-learn, PyTorch
Other Tools: LaTeX, Git, Vim, Bash, Linux

RELEVANT COURSEWORK

Graduate: Machine Learning, Optimization, Neural Networks, Bayesian Learning, Information Theory, Secure Distributed Systems, Computer Vision, Database Design and Implementation.

Undergrad: Data Structures & Algorithms, Probability & Statistics, Numerical Computation, Linear Algebra, Real Analysis, Mathematical Logic, Group theory, Bio-informatics.

EXTRA CURRICULAR

– Football: Playing on indoor court/turf and Fantasy Premier League.
– Proficiency in German (1A) through the Foreign Language Program of IIT Kanpur.
– Indian Classical Music, playing Tabla and Harmonica, pencil sketching.
– Solving Project Euler and other mathematical problems using programming.