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 – GPA: 4.0/4.0   *Expected May '19*

Indian Institute of Technology Kanpur
Bachelor of Science, Mathematics and Scientific Computing   *Jul '13 - Jun '17*

RESEARCH INTERESTS

Machine Learning, Probabilistic Modeling and Approximate Bayesian Inference. Motivated by applications in healthcare, natural language processing and knowledge representation.

ACADEMIC AWARDS

- Recipient of **Inspire** scholarship (awarded by Govt. of India) during undergraduate education for consistent academic performanc in the Mathematics department.
- Awarded **Mitacs Globalink** scholarship for summer research internship in Canada in 2016.
- Merit Certificates from the Central Board Of Secondary Education (CBSE) in 2011 and 2013.

RESEARCH EXPERIENCE

Generative models for document embeddings
Supervisor : Prof. Benjamin Marlin — UMass Amherst   *Sep '18 - Dec '18*

- Working on the problem of learning a generative model for document level embeddings/vectors in an unsupervised manner. Using Edward/Tensorflow (Python) for the coding the probabilistic models.
- Current direction focuses on trying to improve the performance on hamiltonian monte carlo in inference of the document vectors by adopting better negative example sampling techniques.
- The next part includes tweaking the overall inference scheme to improve the performance on downstream nlp task of sentiment analysis in the multi-domain setting.

Chronic Disease Modelling
Supervisor: Prof. Peter Haas — UMass Amherst   *Jul '18 - Oct '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 using the maximum entropy principle for learning.
- Piece-wise likelihood method was employed to partition the feature space graph to approximate the exact inference for the distribution with constraints picked from market basket analysis over the data.
- Constraints given by market-basket were also compared with top ranked feature pairs found out by L-measures (measuring statistical dependence from data) and the pipeline of data loading, feature selection and optimization was implemented in Python using Scipy, Numpy and Pandas.

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
Supervisor: 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
Supervisor: Prof. Debasis Kundu — IIT Kanpur   *Jan '16 - Apr '16*

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

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, *UMass Amherst*

Responsible for grading of assignments, exams and setting up the submission platform along with answering course queries for the following graduate level courses:
- **CS 590M Simulation** (Spring '18), Instructor: Prof. Peter J. Haas
- **CS 646 Information Retrieval** (Fall '18), Instructor: Prof. James Allan & Hamed Zamani

Student Guide, *Counselling Service | IIT Kanpur*

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

CONFERENCE PUBLICATIONS

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

TECHNICAL SKILLS

Languages: Python, R, Matlab, SQL, Java. | **Other Tools:** LaTeX, Git, Vim, Bash, Linux
Frameworks/Libs: Pandas, Scikit-learn, PyTorch, Keras, AWS.

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, Bioinformatics.

Part of Speech Tagger for Hindi

Prof. Brendan O'Connor — University of Massachusetts, Amherst   *Sep '17 - Dec '17*

- 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 fastText 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   *Sep '17 - Dec '17*

- Studied the problem of modelling uncertainty in the results from the (softmax) outputs of deep neural networks trained on cifar-10 and mnist data sets using Keras libaray (Python).

- The uncertainty estimates 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.

Topic Modelling of Documents – LDA

Prof. Piyush Rai — IIT Kanpur   *Jan '17 - Apr '17*

- Implemented a document-topic categorization model on the Reuters news agency data set using the gensim library in Python and surveyed extensions of Variational Inference (VI) for scalability.

- Latent Dirichlet Allocation (LDA) statistical model was used for categorization and the stochastic extension of VI was used for scaling up the inference process for the 7700 documents in training set.

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.