

# Ninad Khargonkar

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

## CURRICULUM VITAE

---

EDUCATION	<b>University of Massachusetts Amherst</b> <i>Master of Science, Computer Science</i> GPA: 4.0/4.0	<i>Expected May' 19</i>
	<b>Indian Institute of Technology Kanpur</b> <i>Bachelor of Science, Mathematics and Scientific Computing</i> GPA: 8.4/10	<i>Graduated June' 17</i>

---

INTERESTS	Machine Learning, Probabilistic Modelling and Scalable Bayesian Inference motivated by health-care, natural language processing and knowledge representation.	
-----------	---	--

AWARDS & ACHIEVEMENTS	Recipient of <b>Inspire</b> scholarship awarded by Dept. of Science and Technology (Govt. of India). Awarded <b>Mitacs Globalink</b> scholarship for summer research internship in Canada. Secured a percentile score of <b>97.7</b> in JEE (Advanced)–2013 out of 150,000 candidates. Secured a percentile score of <b>99.8</b> in JEE (Main)–2013 out of about 1.3 million candidates.	
-----------------------	---	--

---

RESEARCH EXPERIENCE	<b>Chronic Disease Modelling</b> <i>Supervisor : Prof. Peter Haas — UMass Amherst</i>	<i>Jun '18 - Aug '18</i>
	<ul style="list-style-type: none"><li>- 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.</li><li>- 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.</li><li>- The pipeline of data loading, feature selection and optimization for the data set was implemented in Python.</li></ul>	
	<b>Space Efficient Dynamic Programming for LIS</b> <i>Supervisor : Prof. Barna Saha — UMass Amherst</i>	<i>Apr '18 - Jun '18</i>
	<ul style="list-style-type: none"><li>- Designed, analyzed and implemented space efficient algorithms (approximate but deterministic) in dynamic programming for the longest increasing subsequence problem.</li><li>- 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.</li><li>- For the approximate algorithms, I proved algorithmic guarantees about an improved approximation factor and implemented the deterministic and randomized algorithms using Numpy in Python.</li></ul>	
	<b>Globalink Research Internship</b> <i>Prof. Saman Muthukumarana — University of Manitoba, Winnipeg</i>	<i>May '16 - Jul '16</i>
	<ul style="list-style-type: none"><li>- Studied the problem of sampling from large graphs and extracting relevant graph statistics like the clustering coefficient in R.</li><li>- Scale down sampling was implemented on graphs by using random walk variants like Metropolis-Hastings random walk and Jump-random walks and their performance analyzed.</li><li>- Statistical models like ERGM were used for producing model fits and simulating random networks for testing significance of network substructures like triangles.</li></ul>	

## Bayesian Estimation and Sampling

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.

---

## COURSE PROJECTS

### 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 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

Sep '17 - Dec '17

- 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

Sep '17 - Dec '17

- 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

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.
-

OTHER EXPERIENCE	<b>Programmer – BioStats Lab</b> <i>Dept. of Biostatistics — University of Massachusetts, Amherst</i> <i>Sep '17 - Nov '17</i> <p>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.</p>
	<b>Course assistant</b> <i>University of Massachusetts, Amherst</i> <i>Jan'18 - Dec'18</i> <p>Responsible for grading of assignments, exams and setting up the submission platform along with solving course queries for the following graduate level courses:</p> <ul style="list-style-type: none"> <li>- <b>CS 590M Simulation</b> (Spring'18), Instructor: Prof. Peter Haas</li> <li>- <b>CS 646 Information Retrieval</b> (Fall'18), Instructor: Prof. James Allan &amp; Hamed Zamani</li> </ul>
	<b>Student Guide</b> <i>Counselling Service — IIT Kanpur</i> <i>Jun '14 - Jul '15</i> <p>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.</p>
CONFERENCE PUBLICATIONS	<p>Saman Muthukumarana, <b>Ninad Khargonkar</b>, “Modeling and Simulation of UEFA tournament” in <i>Section on Statistics in Sports, Statistical Modeling for Sports Science and Applications activity of JSM 2017 Online program</i>, Baltimore, USA.</p>
TECHNICAL SKILLS	<p><b>Languages:</b> Python, R, Matlab, SQL, Java, C++  <b>Frameworks/Libs:</b> Pandas, Scikit-learn, PyTorch, Keras, Apache Spark, AWS  <b>Other Tools:</b> LaTeX, Git, Vim, Bash, Linux</p>
RELEVANT COURSEWORK	<p><b>Graduate:</b> Machine Learning, Optimization, Neural Networks, Bayesian Learning, Information Theory, Secure Distributed Systems, Computer Vision, Database Design and Implementation.</p> <p><b>Undergrad:</b> Data Structures &amp; Algorithms, Probability &amp; Statistics, Numerical Computation, Linear Algebra, Real Analysis, Mathematical Logic, Group theory, Bio-informatics.</p>