Akash Alok Mahajan

+1-(650)-546-5305 akashmjn@stanford.edu akashmjn.github.io LinkedIn/akashmjn

SUMMARY

Curious & enjoy wearing different hats. Background in applied statistics & signal processing.
Interested in ML/data products especially audio/speech/NLP. Graduating in June 2018.

- Teaching Assistant (TA) at Stanford for Machine Learning (CS229) and Deep Learning (CS230)
- Built a custom deep learning model on radio signals, under evaluation for deployment at SETI
- Built an ECG annotation model comparable to inter-expert deviation on a public dataset 2016
- Initial data science team member at Tiger Global-funded smart vehicle startup in Bangalore 2015-16

Languages : Python, R, Scala, C/C++, SQL, MATLAB
Libraries & Tools : Tensorflow, Keras, Apache Hadoop, Spark, AWS EMR/S3, Shiny, Processing

EDUCATION

Stanford University, Management Science & Engineering Stanford, CA
MS, Applied Statistics & Optimization **GPA : 3.65** Sep 2016-June 2018
Courses - ML : Small Data (MS&E226), Machine Learning, AI (CS229, 221), Data Mining (CS246)
CS : Databases, Algorithms (CS145, 106), Computer Systems (CS107)
Ongoing : NLP with Deep Learning* (CS224n), Digital Signal Processing (DSP)* (EE264) (*current)
Teaching Assistant : CS230 Deep Learning*, CS229 Machine Learning **(co-taught by Andrew Ng)**

Indian Institute of Technology, Madras Chennai, India
B.Tech., Chemical & Control Systems Engineering **GPA : 8.78/10** July 2011-July 2015
Courses/Projects : Time Series Analysis, Kalman Filters, Modern Control Theory

RESEARCH

Attention, I'm Trying to Speak: Text-to-Speech Synthesis Stanford, CA
NLP Project for CS224n, Mentors : Richard Socher, Ziang Xie Ongoing*

- Implementing a single-speaker convolution & attention-based model [1]/[2], on the LJ Speech dataset
- Plan to explore transfer learning for synthesis in different languages from smaller datasets

ATA Radio Signal Classification, SETI Institute+IBM Watson Stanford, CA
Identifying signals from very low SNR, Advisor : Prof. Jeffrey Ullman [report] [git] Mar-June 2017

- Built an ensemble model - custom CNN architecture + optimization based signal tracing (Python/Keras)
- Model under evaluation to be deployed at SETI (6-class Accuracy 80%, 2-class F1 96%)

PROJECTS

Real-time DSP Implementation on iOS *(C++)* EE264 Project*

- Implementing Discrete Multi-tone (DMT) communication through the iPhone audio jack

AI-based Music Generation from Google Magenta *(Python/Tensorflow) [git]* CS221 Project

- Implemented Markov chains, RNNs language models with beam search decoding for inference

Dynamic Memory Allocator - Implementing `malloc`**,** `realloc`**,** `free` *(C)* CS107 Project

- Implemented a segregated explicit free list, exceeding benchmark utilization and throughput targets

EXPERIENCE

Ather Energy, Data Scientist Bangalore, India
Building intelligence on smart electric scooters, part of the initial team of 2 Jul 2015-Jun 2016

Worked on initial feature roadmap & led 3 prototypes - used at the product unveiling.

- Systems to detect drivetrain damage, locate speed bumps, and profile riding styles from sensor data
- Infrastructure - CAN data parsers, initial Postgres schema, internal R/Shiny libraries
- Riding style visualization projects used to engage the early-adopter community [link]

INTERNSHIPS

Salesforce - Coolan (acquired in 2016), Data Science Intern San Francisco, CA
Datacenter hardware monitoring : Assisting Hadoop data-pipeline migration Jun-Sep 2017

- Built an S3 data cataloguing tool *(python/boto3)* and setup a pilot Spark+S3 cluster on Elastic Mapreduce (EMR). Learnt Scala, Spark and Hadoop tools over the summer
- Built a pilot Spark ETL job to structure compressed JSON backups on S3, in use for migration

Predible Health, Deep Learning Engineer Intern Bangalore, India
PoC for automated QT interval annotation of heart ECG waveforms using CNNs Jun-July 2016

- Built a custom 1-D convolution based CNN architecture on MIT-Physionet dataset
- Performance comparable to human inter-expert deviation on dataset (Mean +/- SD : 18 +/- 19.6ms)