ECE Academy of Machine Learning

June 2019 Update

Motivation

- Paradigm shift towards data-centric solutions; often referred to as "Data Science".

- Machine Learning (ML) plays a central role in this new paradigm.

- Many academic institutions have initiated programs for training and equipping undergraduate students with ML skillsets.

  - *MIT's minor in Statistics and Data Science;*
  - *University of Michigan's undergraduate program in Data Science*;
  - *UC San Diego's undergraduate minor in Machine Learning and Controls*.

- Most existing programs are hosted in CS and/or Statistics departments.

  - *UMD's Data Science Specialization for CS majors*.

Rationale

- Our ECE students are already being rigorously trained in areas
  - E.g. probability and statistics, computer programing, signal & systems.
- Many data come as signals from sensors (including info.-rich media data, IoT etc.)
- Many opportunities from industry to support training programs and university collaborations remain unleveraged.
- Disadvantage of lack of a targeted training program centered around ML
  - Makes our students less competent/appealing in job market trend and graduate application, compared to their peers with "Data Science" degrees.
- This **ECE citation in ML** addresses these issues
  - A streamlined ML training via leveraging the rich course offerings, expertise of the ECE faculty, and industry support.

Updates: approved by Univ.; pre-launch with offering of key courses in '19-'20; First cohort anticipated in 2020/2021.

Academy of Machine Learning

Students will receive a citation notation on their official transcript

Admission Requirements (Tentative)

- ENEE150 *Intermediate Programming Concepts for Engineers* (3) (prereq: ENEE140) or CMSC216 *Introduction to Computer Systems* (4).
- MATH141 *Calculus II* (4)
- ENEE222 *Elements of Discrete Signal Analysis* (4)

Main Requirements

- ENEE324 *Engineering Probability* (3) or STAT 400 *Applied Probability and Statistics I* (3)
- ENEE351 *Algorithms and Data Structures* (4) or CMSC351 *Algorithms* (3) or ENEE469O Intro to Optimization
- ENEE436/439M *Foundations of Machine Learning* (3)
- ENEE437 *Design Projects in Machine Learning (3)*

Consortium Plan

- Support from industry through membership (Silver, Gold, Platinum)
- Distinguished fellowship programs
- Summer internships for students
- Computing/Lab equipment and data sharing for the design course

Main Training Program

Probability and Stats + Algorithms + Machine Learning + Hands-on Design

ENEE324 *Engineering Probability*

- Probabilistic Models
- Conditional Probability
- Independence
- Counting Discrete random variables
- probability mass functions
- Functions of random variables
- Expectation, mean, and variance
- Joint pmfs of multiple random variables
- Conditioning for discrete random variables
- Independent random variables
- Continuous random variables and pdfs
- Cumulative distribution functions
- Normal random variables
- Joint pdfs of multiple random variables
- Conditioning of random variables
- Derived distributions
- Covariance and correlation
- Transforms
- Markov and Chebyshev inequalities
- Weak Law of Large Numbers
- Convergence in probability
- Central Limit Theorem
- Strong Law of Large Numbers
- Bernoulli process
- Poisson process
- Selected topics on Markov chains
- Selected topics on second order processes

Main Training Program

Probability and Stats + Algorithms + Machine Learning + Hands-on Design

ENEE351 *Algorithms and Data Structures*

- Mathematical induction
- Recursion
- Combinatorics (counting)
- Discrete probability;
- Recurrence relations
- Analyzing algorithmic performance:
  Work-depth, asymptotic notation, worst case, randomized and probabilistic complexity.
- Stacks, Queues, Graphs, Trees, B-trees, Binary-search trees
- Hash tables, Dictionaries, Heaps.
- Sorting Algorithms and their Analysis:
  Insertion sort, Merge sort, Quicksort, Radix sort
- Graph algorithms:
  Depth-first search, Breadth-first search, Shortest path, Minimum spanning tree, Topological sort.
- Fast Fourier Transform (FFT).
- Algorithmic approaches:
  brute-force algorithms, greedy algorithms, divide-and-conquer, dynamic programming.
- Advanced Topics: Advanced (Tree) Data Structures, Max-flow/Min-cut, NP-Completeness, Parallel Computing

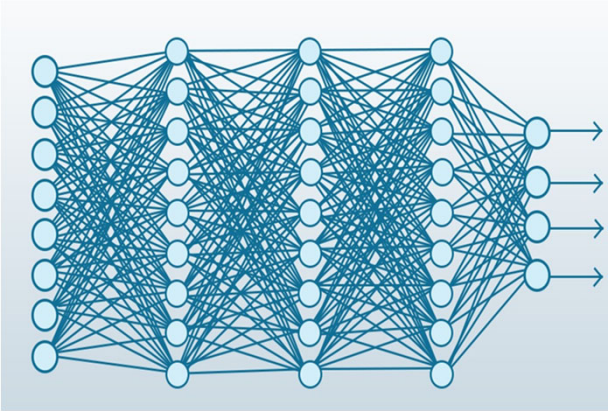
Main Training Program

Probability and Stats + Algorithms + Machine Learning + Hands-on Design

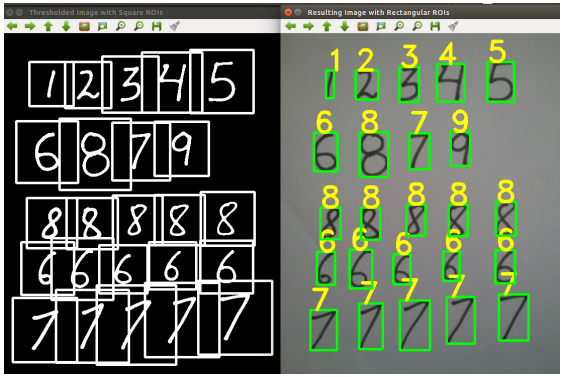
ENEE436/439M *Foundations of Machine Learning*

- Bayes decision theory
- Bayesian classifiers: Gaussian case
- Maximum likelihood estimation
- Principal component analysis
- Fisher's linear discriminant
- Nearest neighbor rule and error bounds
- The Perceptron algorithm
- Convex optimization primer
- Support vector machines
- Neural networks
- Deep learning networks
- Introduction to Caffe/Pytorch
- Unsupervised learning
- Spectral clustering
- Expectation maximization
- IsoMap, LLE and Laplacian eigenmaps
- Hidden Markov models
- Applications to speech
- Applications to images and videos

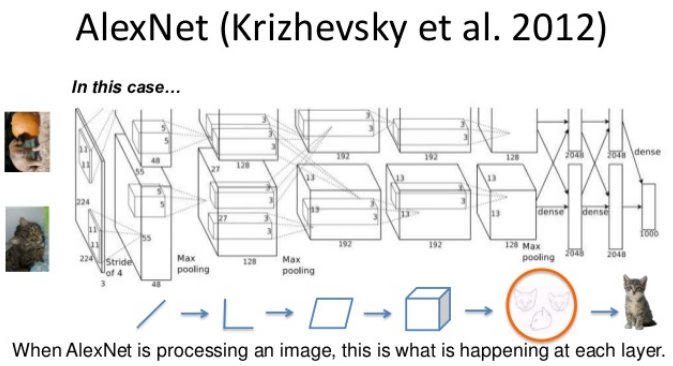
Two hands-on projects:



Handwritten digit recognition



Deep learning for image classification



Main Training Program

Probability and Stats + Algorithms + Machine Learning + Hands-on Design

ENEE437 ***Design Experience in Machine Learning***

- Put theory/techniques into team-based project training
- Build on an existing multimedia Capstone course (ENEE408G)
  - Developed in 2001 by Prof. Liu & Wu, taught by four SP faculty on four media types (image, video, speech, audio), and material updated in 2011 and 2019
  - Pattern recognition already incorporated for most modalities in addition to media compression technologies
  - Team projects mimic tech start-ups, using the then forward-looking "PocketPC" platform
  - Well liked by students and industry recruiters
- Develop ENEE437 course material to gear the focus of labs & projects to learning & analytics
  - Set pre-requisite to the "Intro ML" course
- Students continue a regular Capstone requirement

Main Training Program

Probability and Stats + Algorithms + Machine Learning + Hands-on Design

Under dev **ENEE437** ***Design Experience in Machine Learning***

- Weekly lecture & hands-on labs: First 1-1.5 months
  Choose some of these modules based on project focuses
  - Image, Video & Learning (2-3 weeks)
    Leverage experienced from EE101, 439M, 631 & 633
  - Speech & Natural Language Processing (2-3 weeks)
    Course material updated by Prof. Espy-Wilson for S'19
  - Learning Using GPU & on Data Beyond Media (2-3 weeks)
    Financial record, health data, social media feed, IoT sensing, etc.
- Team Project
  - Start around 1 month; full force in the final 1-2 months
  - Topic brainstorm: with a pool of topic suggestions from industry partners, research & student initiatives
  - Project proposal, planning, and feedback
  - Tech implementation, Milestone report+feedback
  - Final presentation, report and showcase

Current Status & Next Steps

So far:

- Initiated the revision of the capstone design course among SP/ML faculty
- Technical considerations: hidden prerequisites, admission criteria, etc.
- Positive feedback from both ECE faculty and advisory board
- Campus approval received in Spring 2019!

What to do next: Prepare for a formal launch

- Develop and pilot the design course; Line up faculty for the foundation course
- Securing initial support from industry (in progress)
- Learn and refine the program requirements: e.g. senior optimization course

What can Industry help: $ participate in consortium; proj. ideas/data + mentors

Thank you! Questions & Comments?