

ABE 591: Big Data Analytics and HPC
Machine Learning and High-Performance
Computing for Digital Agriculture and Biological
Engineering

Part 1: Algorithms, resilient data lakes, and
analytics at the edge

Somali Chaterji, Assistant Professor, ABE



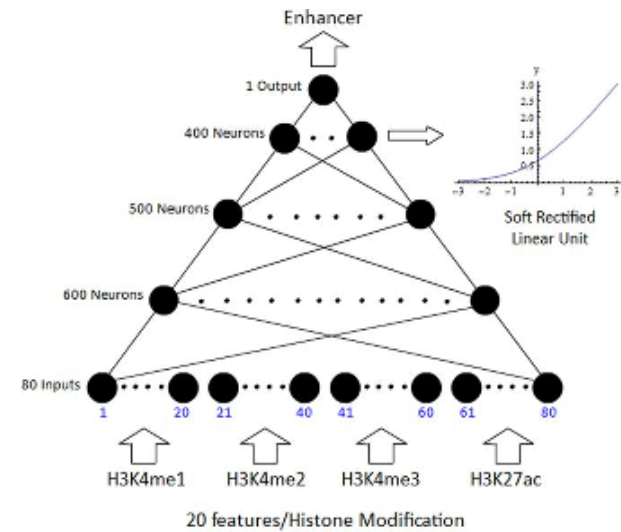
@somalichaterji

About

- **Where & When:** WALC 3121, March 9th – May 2nd; Wed 6:00 pm - 7:50 pm
- PI of **ICAN: Innovatory for Cells and (Neural) Machines**; More at: <https://schaterji.io>; @somalichaterji
- Prerequisite: STAT 30100 OR ABE 20500 OR CHE 32000 OR Graduate Standing; a more advanced statistics course is also acceptable as long as the student has taken the course for credit.
- Auditing is not encouraged.
- This course is part of a Purdue initiative that aims to deliver stackable one credit courses to create a custom data-science curriculum.
- The course will be offered online too through lecture capture during live lectures for future digitized offerings.

Algorithmic Foundations

This course is about **algorithms** that provide the **conceptual bases** for **machine learning** (ML) and its **application** to the **domains of genomics, digital agriculture, and IoT**. These are some of the domains that are generating terabytes of data, both diverse and available in large volumes. These data sets can be used for “**deciphering the rules of life**” or to extract actionable information from the **ubiquitous IoT sensors**. Overall, this course is meant for both **advanced undergraduate or graduate students** and does not assume any prior knowledge of ML.



How to Deploy Algorithms at Scale

- Use-inspired
- Distributed learning
- Scalable databases and data lakes
- Heterogeneous platforms
 - **On-device** computing (in-sensor analytics)
 - **Edge** computing
 - **Cloud** computing

Topics Snapshot

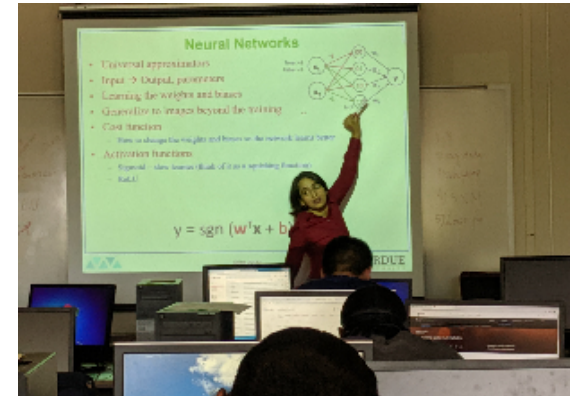
- “Genomical” and IoT data: concepts and storage
- Data immersion (AI/ML stories) [Part 1]
- How to classify?
- Regression
- Interpretable ML
- Data immersion (AI/ML stories) [Part 2]
- A practitioner's guide to ML (includes use of real-world datasets)
- Shallow versus deep networks
- Reinforcement learning

Special topics [based on latest research]

- Any **Two** of:
 - In-sensor analytics
 - Edge computing
 - Data lakes and scalable databases
 - Acceleration for ML:- GPU, TPU, ASICs
 - Epigenomics and regulatory genomics
 - Distributed ML on supercomputers

Course Flyer and Course Updates

- This interactive course will feature active learning. The instructor will customize the difficulty level to the cohort of students to make data science and engineering accessible to all students through varying difficulty of quizzes and course content.
- URL: <http://schaterji.io/teaching.html>
- @somalichaterji [Twitter]



Somali Chaterji,

schaterji@purdue.edu;

**@somalichaterji [Follow for Lab and
Course Updates]**

[How to say my name: SHOH-MAH-LEE CHA-ter-
JEE]

**Assistant Professor [ICAN; Digital Agriculture,
Computational Genomics]; Department of
Agricultural & Biological Engineering**