COMP 5630/6630:Machine Learning

Course Administration

- Instructor
  - Dr. Tin Nguyen
  - Email: tinn@auburn.edu
  - Office hours: after class or by appointment

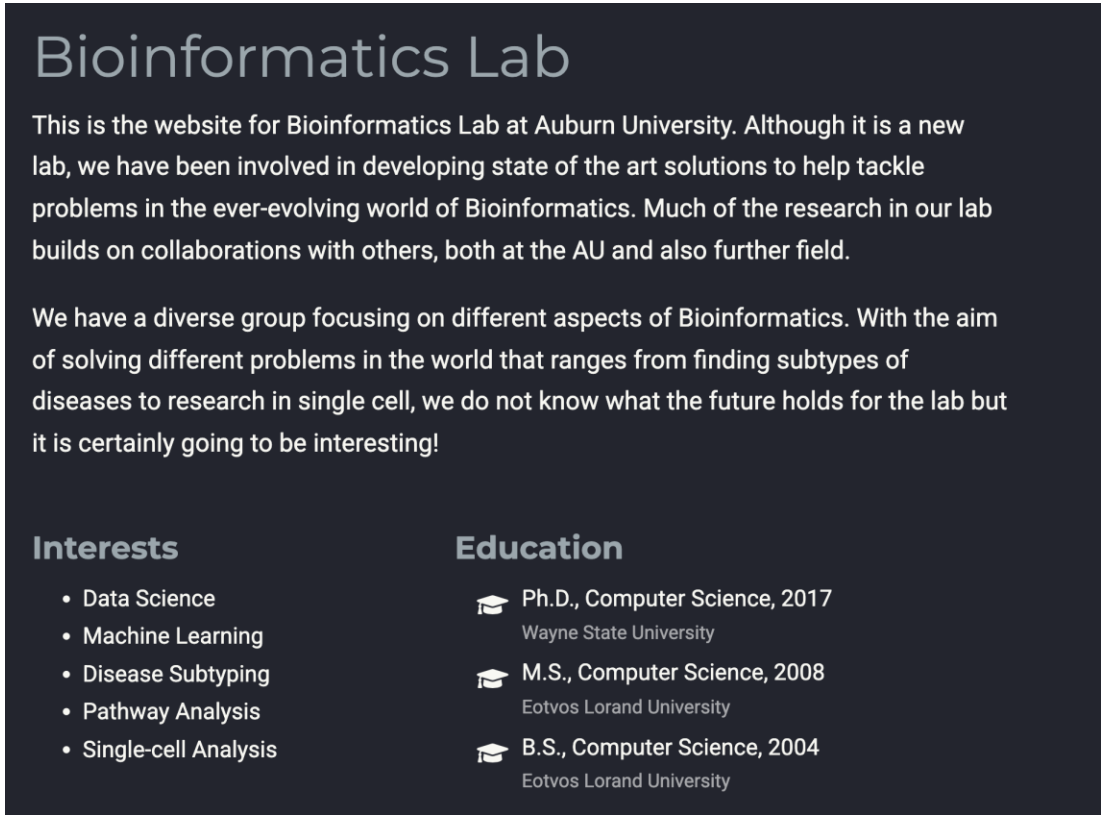
- Teaching Assistants
  - Shubham Trehan, Email: szt0113@auburn.edu
  - Giang Nguyen, Email: vzn0025@auburn.edu

About Me

- BS and MS in CS
  - Eotvos Lorand University, Budapest, Hungary, 2008
- Ph.D. in Computer Science
  - Wayne State University, Detroit, MI, 2011 - 2017
- Assistant Professor in CSSE
  - University of Nevada, Reno, NV, 2017-2023
- Associate Professor in CSSE
  - Auburn University, AL, 2023- Present

My Research

- https://tinnguyen-lab.com/home



Course Administration

- Textbook:
  - No particular book is required for this course. Recommended
  - Textbook 1: Christopher M Bishop, Pattern Recognition and Machine Learning
  - Textbook 2: Introduction to Machine Learning, Third Edition by Ethem Alpaydin.
- Supplementary material will be posted on Canvas as needed.

Grading Information (5600)

Exams (20%)

- One Midterm Exam

Quizzes (25%)

- One quiz at the end of every week on Canvas.
- Open book and open notes

Assignments (35%)

- Four (4) /Five (5) assignments – One every two weeks
- Typically contains two (2) to three (3) problems and one (1) bonus problem

Final Exam (20%)

- A Comprehensive final exam at the end of the semester

- Talk to us about any concerns about projects/exams

Grading Information (6600)

Midterm (20%)

- One Midterm Exam

Quizzes (20%)

- One quiz at the end of every week on Canvas.
- Open book and open notes

Assignments (25%)

- Four (4) assignments – One every two weeks
- Typically contains two (2) to three (3) problems and one (1) bonus problem

Final Project (15%)

- Team effort (2-3 students per team)
- Proposal due at the end of 6th week
- Project report and presentation due at the end of the course
  - Only for graduate section (6600)

Final Exam (20%)

- A Comprehensive final exam at the end of the semester

- Talk to us about any concerns about projects/exams

Grading Policies

Percentage	Grade	GPA Quality Points
90 - 100	A	4.0
80 - 89	B	3.0
70 - 79	C	2.0
60 - 69	D	1.0

Tentative Schedule

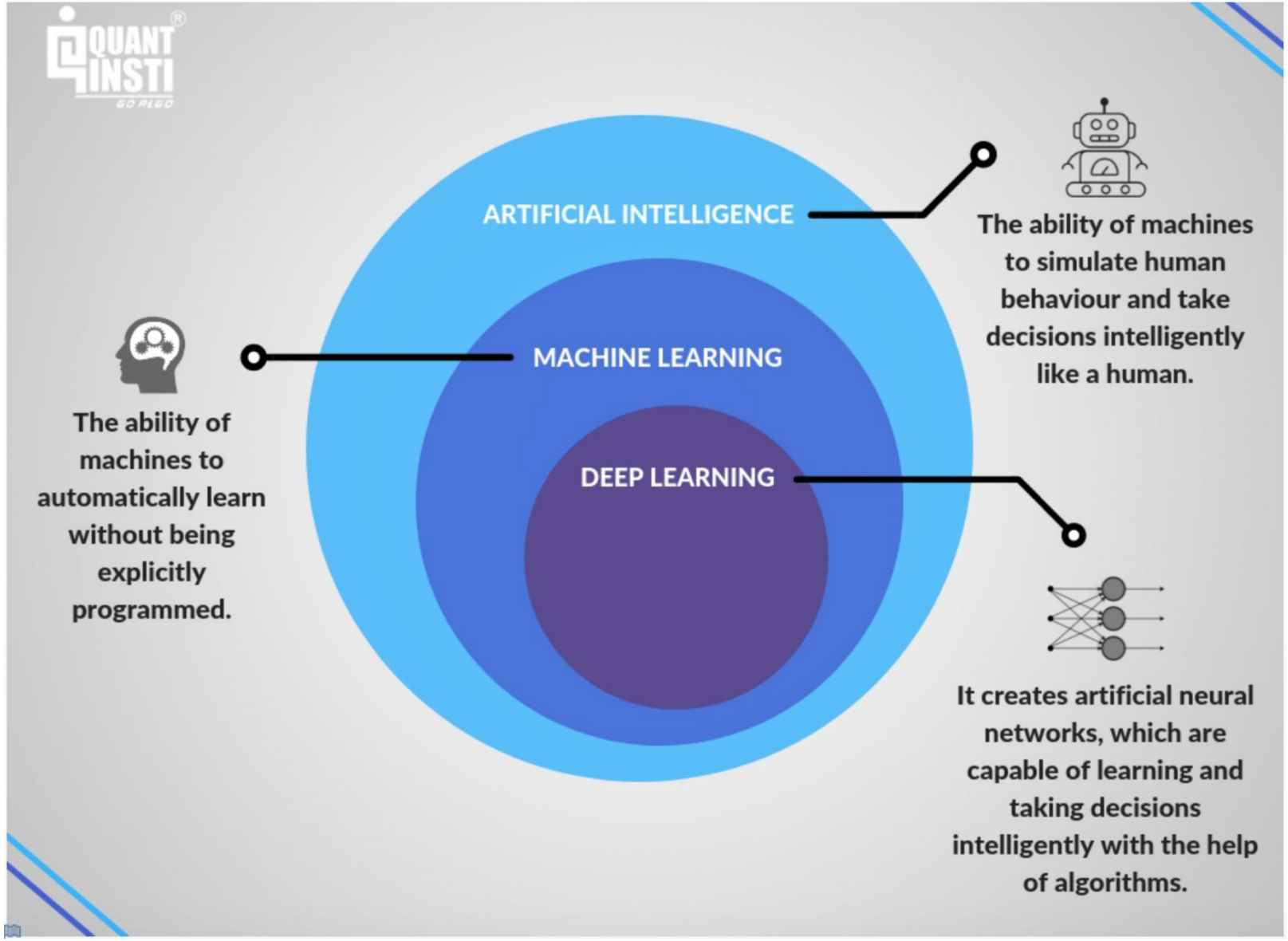
Week 1	Syllabus, Course policies, What is ML?, ML Basics
Week 2	Linear Regression, Logistic Regression,
Week 3	Model Selection, Evaluation Metrics
Week 4	Evaluation Metrics, Neural Networks
Week 5	Deep Learning - Convolutional Neural Networks
Week 6	Deep Learning - Sequence Learning Project Proposal Due: Oct 1 (COMP: 6630)
Week 7	Deep Learning - Recent Advances:NLP
Week 8	Midterm , Naïve bayes

Tentative Schedule

Week 9	Naïve Bayes, Decision Trees
Week 10	Project Interim Report: Status, Issue, and Changes. Due: Oct 31, Decision Trees
Week 11	KNN, Intro to SVM: kernel, Support Vector Machines
Week 12	Unsupervised Learning: clustering: Kmeans, Hierarcheal, Dimensionality reduction
Week 13	Data visualization, Dimensionality Reduction: PCA, Autoencoder
Week 14	Thanksgiving Break
Week 15	Project Final Report Due: Dec 2, 2024 Ethics in ML/ Advanced Topics
Week 16	Final Exam Week

What is Machine Learning

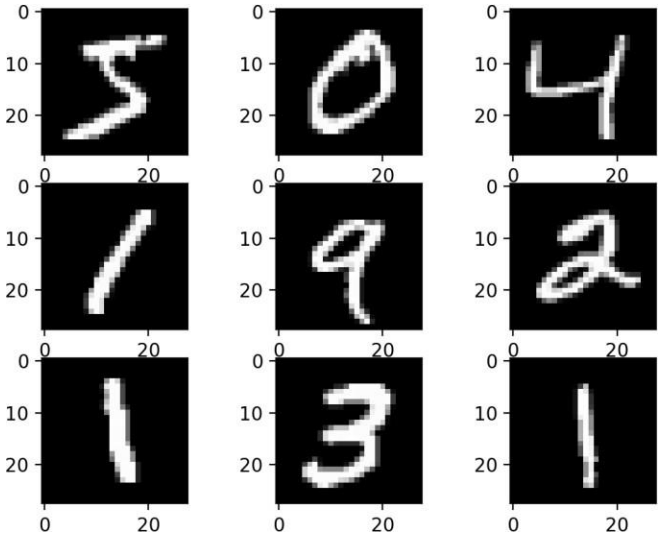
- Artificial Intelligence ?
- Machine Learning ?
- Deep Learning?



Credit: Sathyanarayanan Aakur

What is Machine Learning

- Task: Digit Recognition
- Hard way: Understand each digit
- Easy way: Learn from data!



Example: Netflix Challenge

- Task: Predicting how a viewer will rate a movie

- 10% improvement = 1 million dollars

Slide Credit: Yaser Abu-Mostapha, Dhruv Batra

Example: Netflix Challenge

- Task: Predicting how a viewer will rate a movie
- 10% improvement = 1 million dollars
- Machine Learning Basics:
  - A pattern exists
  - Formulate the problem with example data

Slide Credit: Dhruv Batra

Comparison

- Computer Programming

- Input
  - Data
  - Program

Output

- Machine Learning

- Input
  - Data
  - Output

Program

Slide Credit: Pedro Domingos, Tom Mitchel, Tom Dietterich

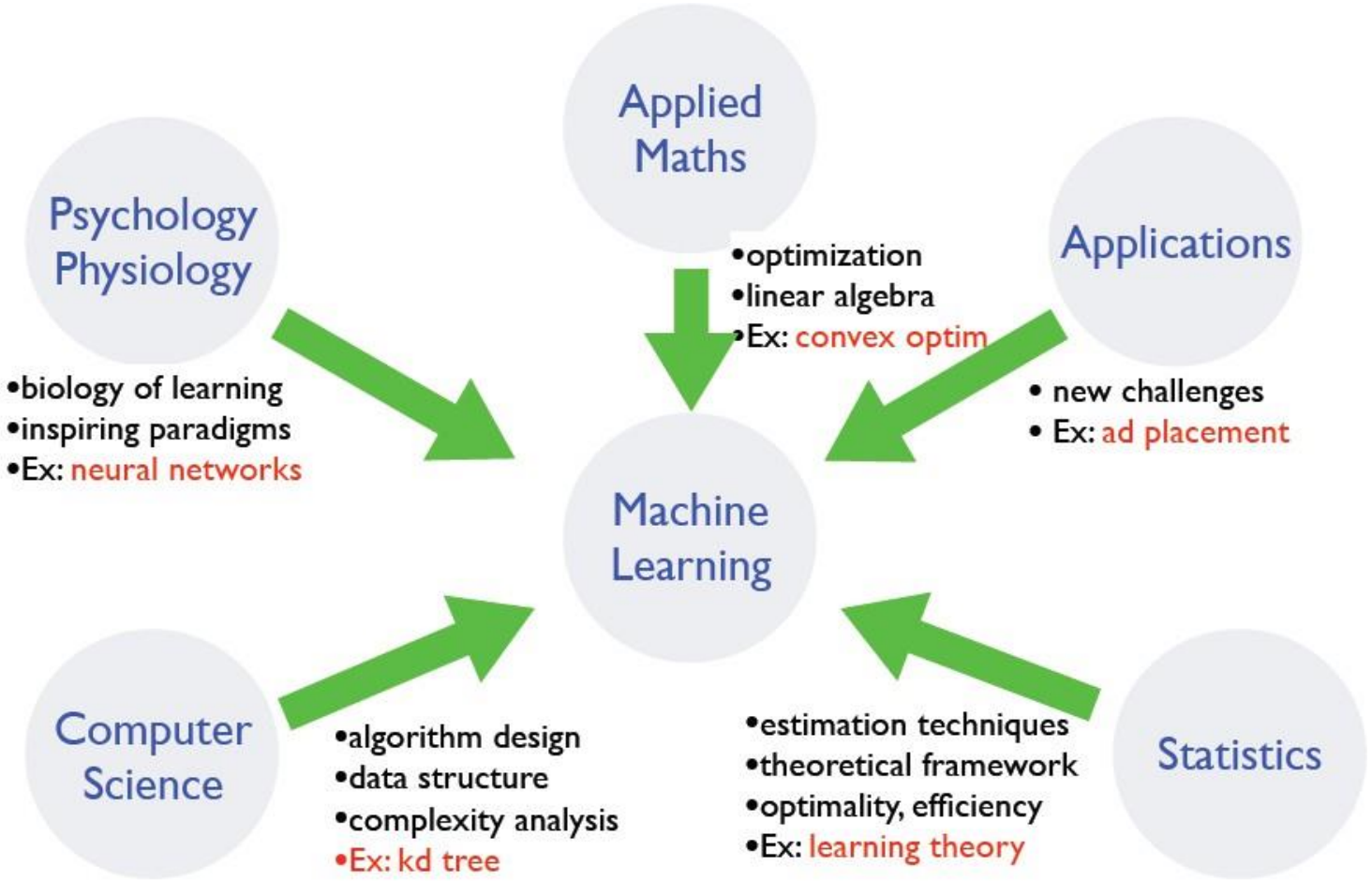
What is Machine Learning?

- [Arthur Samuel, 1959]: popularized the term “machine learning ”

  “Field of study that gives computers the ability to learn without being explicitly programmed”

- [Tom Mitchell] algorithms that
  - improve their performance (P)
  - at some task (T)
  - with experience (E)

Machine Learning (ML) Applications



Slide Credit: Fei Sha, Dhruv Batra

What is Machine Learning (ML)?

- Input: x (email)
- Output: y (spam or non-spam…)
- (Unknown) Target Function
  - f: X → Y (the "true" mapping / reality)
- Data
  - $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
- ML
  - Learn a "mapping" from input to output f: X → Y
  - Learning objective: minimize the loss between "learned mapping" and "true mapping"

ML Algorithm Components

1. Representation

2. Evaluation / Objective Function

3. Optimization

Representation

- Decision Tree
- Support Vector Machines
- Deep Neural Network
- Embedding

Evaluation / Objective Function

- Accuracy
- Precision
- Recall
- Mean Squared Error
- Entropy

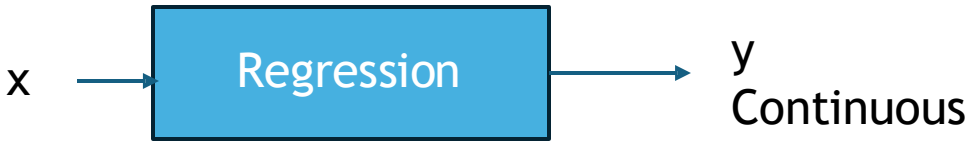
Optimization

- Greedy search
- Beam search
- Convex/non-convex optimization

Types of Learning

- Supervised learning
  - Training data includes labelled outputs
- Unsupervised learning
  - Training data does not include labelled outputs
- Semi-supervised learning
  - Training data includes a few labelled outputs
- Reinforcement learning
  - Rewards from sequence of actions

- Supervised Learning



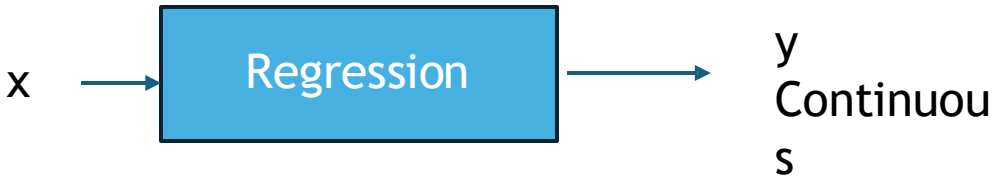
- Unsupervised Learning

Supervised Learning: Classification

Supervised Learning: Classification

- Predicting letter grade of a course
  - Input (x): scores of assignments, quizzes, and exam
  - Output (y): Letter grade

Supervised Learning: Regression



Supervised Learning: Regression

- Predicting temperature
  - Input (x): humidity, speed of the wind, precipitation
  - Output(y): Temperature

Unsupervised Learning: Clustering

Unsupervised Learning: Clustering

- Market Segmentation
  - Input (x): purchasing behavior, demographic
  - Output (y): Groups of customers based on similarity
  - Example:
    - College student discounts

Unsupervised Learning: Clustering

- Data preprocessing stage to select relevant features
- Task: Predict customer demands of room types

Customer	Date of birth	Customer age	Country	Marital status	Room service breakfast	Single room	Double bedroom	Twin room	Suite
Customer A	2/14/1998	23	US	Single	yes	2	-	-	-
Customer B	4/13/1965	56	US	Married	no	-	3	-	1
Customer C	9/28/1984	36	US	Married	yes	-	-	6	-
Customer D	3/8/1972	49	US	Single	no	2	-	-	3

  - Country: same across all inputs
  - Age and Date of Birth: can be merged

Image Credit: https://www.altexsoft.com/blog/unsupervised-machine-learning/