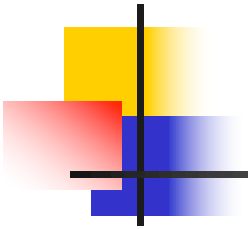
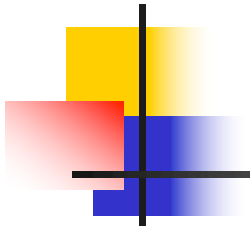


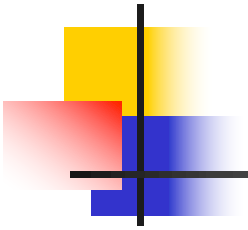
Machine Learning

Introduction



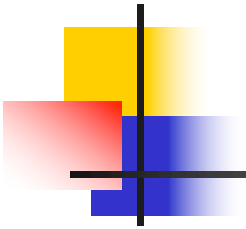
Course Organization

- Course Hours
  - Tuesday, Thursday 3:30pm – 4:50pm Central Time
  - All Classes will be synchronous over Microsoft Teams (001)
    - Conference for each lecture will be initiated about 10 minutes before the start of the class
    - During class I might not be able to see the chat window so you should use your microphone to ask questions
  - Classes will be recorded and available after class (900)
- Course Information
  - All course material will be on Canvas
  - There is also a web page for the course at http://ranger.uta.edu/~huber/cse6363



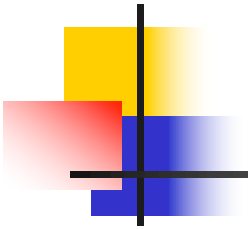
Course Organization

- Office Hours
  - Instructor office hours will be in Microsoft Teams (they are in the Teams Calendar)
    - Tuesday, Thursday 5:00pm – 6:00pm Central Time
    - Wednesday 10:00am – 11:00am Central Time
  - Instructor office hours will be held On Microsoft Teams
    - Permit private and public chat as well as audio/video conversation
  - Instructor email is *huber@cse.uta.edu* (please include the course number in the subject line)
  - Teams Chats can also be used to communicate
  - Canvas also supports a Q&A discussion board where you can post questions



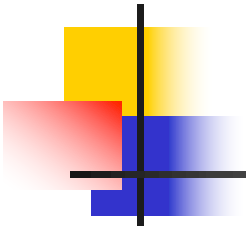
Course Organization

- Course Work
  - The course work includes 3 Homework assignments, 3 Project assignments, 6 Quizzes, and a Final Project
  - Homework and project assignments will consist of theoretical questions as well as programming components
    - Programming can be in any common programming language that has a free Linux compiler (e.g. Python, Java, C++, Matlab, …)
    - Use of specific ML-related libraries is not permitted (except if stated)
      - If you have questions regarding a language or toolbox, ask first
  - The Final Project will be a slightly larger, individual project that will be presented by each student during finals week



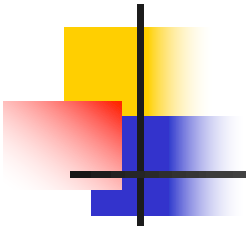
Course Organization

- There will be a Quiz after each assignment submission
  - Part of the quizzes are questions about *your specific* submitted assignment and will require that you annotate and submit an electronic copy of your assignment code as part of the quiz
    - You will need to have an electronic (ideally PDF) version of your assignment at the time of the quiz.
- Grading:
  - 3 Homework - 20%
  - 3 Small Projects - 20%
  - Final Project - 10%
  - 6 Quizzes - 45%
  - Participation - 5%
- For more information look at the Syllabus posted on Canvas



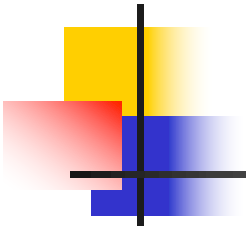
What is Machine Learning ?

- T. Mitchell
  - Improving performance via experience
  - Formally, a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T as measured by P, improves with experience.
- H. Simon
  - Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the task or tasks drawn from the same population more efficiently and more effectively the next time.



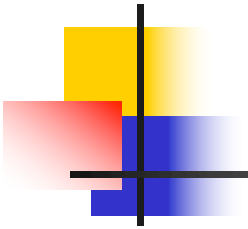
Machine Learning

- Machine learning algorithms vary significantly in terms of the complexity of their representations and the learning mechanism
  - Non-parametric representations
    - Data sets
    - Weighted sample distribution
  - Parametric representations
    - Neural networks
    - Graphical models
  - Ensemble representaitons



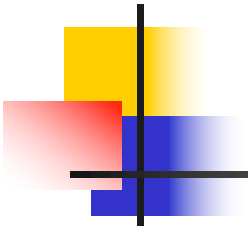
Machine Learning

- Machine learning problems address
  - Generalization
  - Prediction
  - Grouping
  - Structure identification
  - Feature formation
  - Decision
  - …



Machine Learning

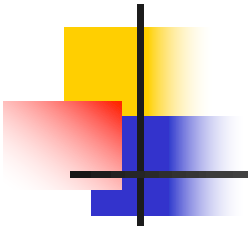
- Machine learning can be divided into different types based on the information they use
  - Supervised Learning
    - Data contains desired output/result
  - Unsupervised Learning
    - Data does not contain desired output/result
  - Semi-Supervised Learning
    - Some data includes desired output/result
  - Reinforcement Learning
    - Data includes some indirect performance measure



Machine Learning

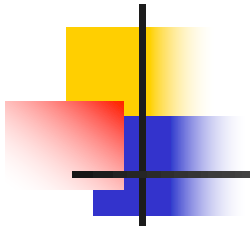
- Different types of machine learning are characterized by the data they learn from, not by the approach they take
  - Different learning areas can potentially use the same representations and techniques
    - E.g. Neural networks can be used in all learning areas

Common Supervised Learning Problems



- Supervised learning usually addresses one of two problems
  - Regression
    - Learn to predict a continuous output value
    - Function approximation
  - Classification
    - Learn to predict a discrete label from a set of pre-defined labels

Common Unsupervised Learning Problems

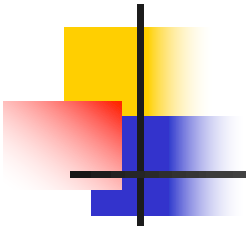


- Unsupervised learning usually addresses
  - Clustering
    - Learn to separate the data into groups
  - Feature learning
    - Learn to compress/transform the representation of the data
  - Structure learning
    - Learn the structure in the data
  - Pattern learning / data mining
    - Learn to find frequent patterns in the data

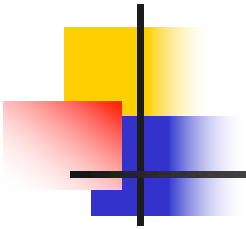
Common Semi-Supervised Learning Problems

- Semi-supervised learning problems address the same issues as supervised learning problems
  - They generally use all instances to generalize structure – an unsupervised problem
  - They generally use the labeled instances to solve the the prediction problem over the generalized structure

Common Reinforcement Learning Problems

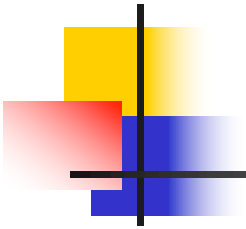


- Reinforcement learning attempts to learn the correct output from quantitative information
  - Decision learning
    - Learn to produce the action that optimizes the quantitative feedback
  - Function/class prediction learning
    - Learn to predict an correct output from possible output values that optimizes the quantitative feedback



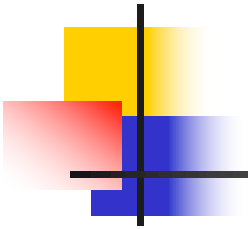
Machine Learning

- Machine learning algorithms generally have three common components
  - Representation
    - How is the learning result represented
  - Evaluation metric
    - What is the criterion used to learn
  - Optimization approach
    - How is the evaluation metric improved with data



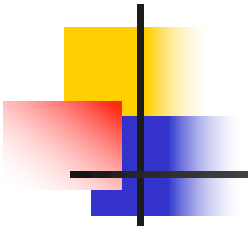
Representations

- Algorithms vary in terms of the way they encode the learned functions
  - Sets of Instances (or distributions)
  - Decision trees
  - Sets of rules
  - Graphical models
  - Neural networks
  - Support vector machines
  - Model ensembles
  - …



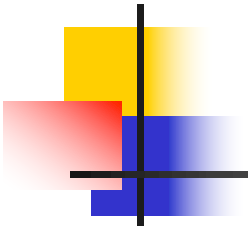
Evaluation Metrics

- Algorithms vary in terms of what they try to improve internally
  - Accuracy /error
  - Precision/recall
  - Likelihood
  - Cost/Utility
  - Margin
  - Entropy
  - Similarity
  - …



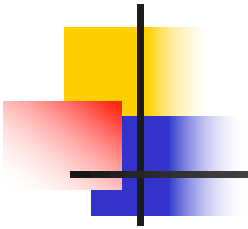
Optimization

- Algorithms vary in terms of how they try to improve the evaluation metric using data
  - Combinatorial optimization
    - Find a global best solution (e.g. combinatorial search)
  - Local optimization
    - Locally optimize (e.g. gradient ascent)
  - Local improvement
    - Locally improve the solution (e.g. hill climbing)
  - Constrained optimization
    - Optimize within given conditions (e.g. linear programming)
  - ...



Applications

- Machine learning has been applied in a wide range of problem domains
  - Security
    - Spam filters, intrusion detection, network profiling, etc.
  - Databases
    - Data mining, personalized web search, re-ranking, etc.
  - Bioinformatics
    - Cancer prediction, toxicity prediction, etc.
  - Operating systems / distributed computing
    - Load balancing, predictive scheduling, job dispatching, etc.
  - User interfaces
    - Interface customization



Applications

- Machine learning has been applied in a wide range of problem domains
  - Artificial Intelligence
    - Game playing, Cognitive reasoning, etc.
  - Vision
    - Object recognition, segmentation, image interpretation, etc.
  - Robotics
    - Learning control, autonomous driving, environment sensing, etc.
  - …