**CSE 5095-006, Spring 2019**
**Machine Learning for Physical Sciences**

**Instructor:** Qian Yang, qyang@uconn.edu

**Summary:** This course will cover recent advances in machine learning for materials science, chemistry, and physics, and discuss some of the unique opportunities and challenges at the intersection of machine learning and these fields. The first half of the course will focus on machine learning concepts important in scientific applications; the second half of the course will dive into specific examples of current research. The course will seek to connect students from computer science with students from the physical sciences together to build projects bridging their fields.

Students should either be familiar with basic machine learning, or with applications in the physical sciences. A background in both is preferable but not required.

**Tentative Schedule:**

Week 1: Introduction

- What types of problems is machine learning best suited for? When should we *not* use machine learning? Main themes in current research using machine learning for science: structure-property prediction; speeding up existing computational methods; dimensionality reduction for interpretation of data; etc.
- Quick review of different types of machine learning algorithms. Regression, classification, dimensionality reduction. Supervised, semi-supervised, and unsupervised learning. Reinforcement learning. Generative vs. discriminative models.

Week 2: Data

- How much data do we need? How well does it represent the distribution we are interested in? Interpolation vs. extrapolation. Brief introduction to computational learning theory.
- How can we deal with uncertainty in data? Regularization. Bayesian methods. Multi-fidelity methods.

Week 3: Small Data

- Data Augmentation. Transfer learning.
- Bayesian Optimization. Active Learning.
- Using symmetry and other physical constraints.

Week 4: Clustered, Biased, Imbalanced Data

- Useful strategies: sampling methods; modified cost functions; Bayesian inference; semi-supervised learning; Tree-based methods.
- Anomaly detection.

Week 5: Evaluating Model Performance

- Performance metrics; comparing against random. Key role of benchmark datasets.
- Assumptions in cross-validation. Effect of data distribution on cross-validation.
- Robust machine learning in scientific applications.

Week 6: Feature Engineering

- What considerations are important when manually designing features? Information content of features. Correlations between features. Correlation of features with target values. Categorical features. Standardization of features.

- Feature invariances and symmetries.

Week 7: Deep Learning-based Strategies

- Review neural network designs: feed-forward, convolutional, recurrent. Restricted Boltzmann machines.
- Autoencoders, word2vec.
- Generative Adversarial Networks.

Week 8: Interpretable Machine Learning

- Strategies for designing interpretable models.
- Feature selection; dimensionality reduction; model reduction.
- Model-agnostic strategies for interpretation.

Week 9: Application – Learning Structure-Property Relationships

- Review: molecules, proteins, crystals, QSAR.
- Strategies and Challenges: feature representations that consider symmetries; graph-based models; data collection and validation of results.

Week 10: Application – Prediction of Pathways for Chemical Synthesis

- Review: chemical reactions; chemical reaction databases
- Strategies and Challenges: feature representation of reactions; data distribution of reaction databases and data augmentation strategies; deep learning for chemical reaction prediction

Week 11: Application – Learning Potential Energy Functions for Molecular Dynamics Simulation

- Review: molecular dynamics, empirical potentials, ab initio methods.
- Strategies and Challenges: learning functional forms vs using neural networks; designing features for molecules vs solid state; fitting to energy vs forces.

Week 12: Application – Solving Inverse Problems in Spectroscopy

- Review: different types of spectra; existing strategies and heuristics; defective materials; amorphous materials
- Strategies and Challenges: ill-posedness of the inverse problem; noisy data; featurization of spectra

Week 13: Application – Computer Vision and Scientific Imaging

- Review: machine learning algorithms for image recognition; convolutional neural networks. Examples of scientific imaging: medical imaging, materials imaging via microscopy.
- Strategies and Challenges: biased data and automatic detection of biased data; classifying images; detection and extraction of features in images

Week 14: Conclusion

- Other applications: natural language processing for automatic extraction of data from scientific literature; model reduction for speeding up numerical simulations.
- Future directions.