Machine learning for physics and astronomy (PHYS 555)

Hossen Teimoorinia hossteim@uvic.ca
University of Victoria, Fall 2024, Tuesdays and Thursdays 15:30-17:30

Course Overview:

This course provides students with a comprehensive understanding of popular machine-learning algorithms and their mathematical foundations, equipping them with practical skills to implement these algorithms using Python libraries. The course emphasizes applying machine learning techniques across various domains. Students will explore the critical role of data in computing and its significance in solving real-world challenges through machine learning. They will also learn to align real-world problems with appropriate ML algorithms, formulating these issues as machine-learning tasks. The course contains an in-depth exploration of foundational machine learning concepts and cutting-edge models, including generative models.

Course Content:

Introduction to Machine Learning:

- Overview of machine learning, its role in computer science, and its applications in problem-solving.
- Data representation, including features, linear transformations, and matrix-vector operations.
- Problem formulation, focusing on classification and regression tasks.
- Probability distributions, prior probabilities, and Bayes' Rule.
- Introduction to performance metrics for evaluating machine learning models.
- Learning paradigms: supervised, unsupervised, and a brief overview of others.

Fundamentals of Machine Learning:

- Dimensionality reduction techniques: Principal Component Analysis (PCA) and Singular Value Decomposition (SVD).
- Nearest Neighbors and K-Nearest Neighbors (KNN) algorithms.
- Regression techniques: Linear Regression, Logistic Regression, and Polynomial Regression (with feature extraction).
- Decision Tree Classifiers: theory, training.
- Concepts of generalization and overfitting in machine learning models.
- Importance of training, validation, and testing datasets in ensuring model generalization.
- Ensemble methods: boosting, bagging, and Random Forests (RF).
- Support Vector Machines (SVM) and kernel methods.
- Naive Bayes classifiers: theory and practical applications.

Unsupervised Learning and Visualization Techniques:

- Clustering techniques: K-means and Gaussian Mixtures.
- Visualization of high-dimensional data using t-SNE.

Deep Learning:

- Introduction to Neural Networks and deep learning concepts.
- Backpropagation and its role in training deep learning models.
- Regularization techniques to prevent overfitting.
- Fully connected Neural Networks and their applications.
- Deep Computer Vision: Convolutional Neural Networks (CNNs).
- Introduction to Autoencoders and their applications in dimensionality reduction and data generation.

Generative Models:

- Variational Autoencoders (VAEs): theory, implementation, and applications.
- Generative Adversarial Networks (GANs): architecture, training, and real-world use cases.
- Diffusion Models: an introduction to their principles and applications in generative tasks.

Time Series Analysis in Machine Learning:

- Classical time series problems
- Recurrent Neural Networks (RNNs) and their applications in sequential data.
- Introduction to Transformers and their use in time series analysis and other sequential tasks.

Assessment:

- **Assignments (35%)**: Assignments to reinforce theoretical concepts and practical implementation.
- **Paper/Project Presentation (10%)**
- **Final Project (35%)**: A comprehensive project focused on solving a real-world problem using machine learning.
- **Class activities and participation (20%)**

Software: Python, with libraries such as NumPy, Scikit-learn, and TensorFlow.