Syllabus: Introduction to Machine Learning for Physics

Course Brief

This course introduces the fundamental principles of Machine Learning through the lens of physics. We move beyond the 'black-box' approach to explore how various Deep Learning algorithms are designed and deployed in research. Students will gain practical experience applying these tools to diverse datasets—including events recorded by particle physics detectors, Gravitational Waves, as well as astronomical, medical, or environmental imagery—culminating in a statistical hypothesis-test project. The goal is to master both the 'how' of the algorithms and the 'why' of their application in the physical sciences.

1. The Physicist's Toolkit & Statistical Foundations

- **From Likelihoods to Loss:** The Neyman-Pearson Lemma as the foundation of ML, proving that Binary Cross-Entropy is the Negative Log-Likelihood of a Bernoulli process.
- **Data Structures:** Moving beyond tables (e.g. `numpy` arrays or `pandas` dataframes). Introduction to **Awkward Arrays** for "ragged" physics data (varying particles per event) and `uproot` for ROOT file integration.
- **Data Preprocessing:** Standardization, handling detector masking/outliers, and the importance of physical units in scaling.

2. Shallow Learning & Decision Trees

- **Boosted Decision Trees (BDTs):** Why they remain the "baseline" in several scientific applications. Introduction to XGBoost/LightGBM.
- **Validation in a Statistically-Limited Regime:** K-folding, overtraining checks, and the Bias-Variance tradeoff.
- **Metrics for Discovery:** Moving from "Accuracy" to **ROC curves**, Purity, Efficiency, and Significance.

3. Deep Learning: Symmetries & Architectures

- **Deep Neural Networks (DNNs):** Optimisation by Gradient descent, using Backpropagation and the Chain Rule (interpreted as differential calculus).
- **Convolutional Neural Networks (CNNs):** Image processing in telescopes, calorimeters, satellites, and medical diagnostic.
- **Graph Neural Networks (GNNs):** solve pattern recognition problems in particle physics.

4. Sequential Data & Generative Models

- **Time-Series (RNNs & Attention):** Analyzing electronic pulses by particle detectors (e.g. LArTPC) or Gravitational Waves strain data (e.g. LIGO). Introduction to the "Attention" mechanism.
- **Generative Models:**
  - **VAEs:** Dimensionality reduction and latent space physics.
  - **Diffusion Models & Normalizing Flows:** Conditional density estimation and ther application to fast detector simulation.

5. Statistical Inference & "The Physics Output"

- **Anomaly Detection:** Unsupervised learning (Autoencoders) to search for "New Physics" without a theoretical model.
- **Limit Setting:** Using ML scores as input to **Profile Likelihood Fits**.
- **Wilks' Theorem:** Translating ML performance into "Sigma" (σ) significance.

Prerequisites

- **Laboratory of Calculus and Statistics.** The competencies and knowledge acquired during this course, on both python programming and statistical methods, are considered fundamental and will be assumed.

Acquired Technical Skills

1. **Libraries:** `PyTorch` (primary), `scikit-learn`, `numpy`, `pandas`, `awkward`, `uproot`, `matplotlib`.
2. **Architectures:** Building custom Loss Functions that include physical constraints (e.g., mass conservation).
3. **Environment:** Managing reproducible research via pip packages, Conda (or python virtual) environments and Docker containers.

Organization & Exam

- **Lectures:** 50% Theory / 50% Hands-on Coding Lab.
- **Final Exam:** Students will be provided with a "blind" dataset and must complete a specific analysis task designed by the instructors. The project requires fulfilling a set of technical objectives and documenting the process in a formal written report. Each report will be evaluated based on its clarity, physical insight, and completeness. Upon

successful assessment of the report, students will proceed to an oral examination. This session will begin with a technical discussion of the student's project before expanding into a comprehensive review of any topic covered during the course.

References

- "Hands-On Machine Learning with Scikit-Learn and PyTorch" - Aurelien Genon (2025) - n.b. an older edition covering the same topics is "Hands-On Machine Learning with Scikit-Learn, Keras and Tensorflow" from the same author is available at the library
- "Understanding Deep Learning" - Simon Prince
- Goodfellow, Bengio and Courville, Deep Learning, MIT Press
- Pen and Paper Exercises in Machine Learning - arXiv:2206.13446