Greater machine learning-based prediction & decision-support capabilities are
Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.

- Arthur Samuel, 1959

Machine Learning: A set of rules that allows systems to learn directly from examples, data and experience.

- Royal Society, 2017

"Learning" is the process of transforming information into expertise or knowledge; "Machine learning" is automated learning.

- Paraphrased from Jordan et al., 2015

Working Definitions of Machine Learning
Examples of Popular Machine Learning Methods

Deep Learning
- Convolutional Neural Network
- Deep Boltzmann Machine & Belief Networks
- Stacked Auto-Encoders

Bayesian
- Naive Bayes
- Averaged One-Dependence Estimators
- Bayesian Belief Networks
- Gaussian Naive Bayes
- Multinomial Naive Bayes
- Bayesian Network

Ensemble
- Random Forests
- Gradient Boosting Machines
- Boosting
- Boostrapped Aggregation
- AdaBoost
- Stacked Generalization
- Gradient Boosted Regression Trees

Decision Tree
- Classification and Regression Tree
- Iterative Dichotomizer 3
- C4.5
- C5.0
- Chi-squared Automatic Interaction Detection
- Decision Stump
- Conditional Decision Trees
- M5

Neural Networks
- Radial Basis Function Network
- Perceptron
- Back-Propagation
- Hopfield Network

Dimensionality Reduction
- Principal Component Analysis & Regression
- Partial Least Squares Regression
- Multidimensional Scaling
- Projection Pursuit
- Partial Least Squares
- Mixture
- Quadratic
- Linear Discriminants

Regularization
- Least Absolute Shrinkage & Selection Operator (LASSO)
- Elastic Net
- Least Angle Regression

Instance-Based
- k-Nearest Neighbor
- Learning Vector Quantization
- Self-Organizing Map
- Locally Weighted Learning

Clustering
- k-Means
- k-Medians
- Expectation Maximization
- Hierarchical Clustering

Regression
- Linear-
- Ordinary Least Squares-
- Stepwise-
- Logistic Regression
- Multivariate Adaptive Regression Splines
- Locally Estimated Scatterplot Smoothing (LOESS)
Why:
What:
What is a “Priority Research Direction”?

• High-priority area of research for scientific machine learning

• Has following components (ala Heilmeier):

  – Clear statement of key challenge
  – Context in the current scientific landscape to establish timeliness and competition
  – Plausible research pathway(s)
  – Clear scientific impact

• It is not

  – A proposal for a specific project
  – Your favorite area of research without connection to SciML themes
How: Workshop Components

• A Basic Research Needs-inspired agenda
• Plenary talks
  – Highlight status of machine learning, challenges, open questions
• Panel discussions
  – Summarize pre-workshop report
  – Provide perspectives across DOE ASCR facilities, ECP, and other organizations
• Breakout sessions
  – Organized around ~140 submitted Position Papers presented as flash talks
  – The "work" in workshop: Crucible for new Priority Research Directions
  – Need high levels of interaction and input (long days…)
  – Brainstorming (Day 1), Refining (Day 2) & Presenting (Day 3) Priority Research Directions
Agenda Overview - Tuesday

- Note: Observers will be able to watch plenaries, panels, and breakout summaries via Zoom webinar
- Tuesday – Welcome
  - Scientific Machine Learning: ASCR Facilities Perspective
  - Three Principles of Data Science: Predictability, Stability, and Computability: Bin Yu
  - Scientific Machine Learning across Federal Agencies
  - Summary of Pre-Workshop Report & Themes
  - Physics, Structure, and Uncertainty: Probabilistic Learning for Risk Mitigation: Roger Ghanem
- Parallel breakout sessions
Agenda Overview – Wednesday & Thursday

• Wednesday
– Machine Learning in the Wild: Jacob Shapiro
– Preliminary breakout reports and discussion
– Challenges & Scope of Empirical Modeling: Ronald Coifman
– Parallel breakout sessions

• Thursday
– Final breakout reports and discussion
– Summary of priority research directions
Breakout sessions

- Numerical Analysis for Machine Learning
  Mark Ainsworth, James Sethian

- Machine Learning, Multifidelity, & Reduced-order Models
  Karen Willcox, Abani Patra

- Machine Learning, Optimization, & Complexity
  Stefan Wild, Manish Parashar

- Probabilistic Machine Learning
  Habib Najm, Aric Hagberg

- Machine Learning Interpretability
  Timo Bremer, Yannis Kevrekidis
Each Breakout Session developed a list of critical research areas.

Day 2: Research areas were evaluated & grouped into topics according to 5 Priority Research Directions.

The Session leads & members joined the relevant Priority Research Direction (PRD) group.

The PRD teams met to formulate the research approaches and thrust areas.

Day 3: Report out and writing of PRDs and Panel reports.
Scientific Machine Learning: Priority Research Directions (PRDs)
PRD1: Domain-Aware
Leveraging Scientific Domain Knowledge
PRD2: Interpretable
Explainable and Understandable Results
PRD3: Robust
Stable, Well-Posed, and Efficient Formulations
PRD4: Data-Intensive
Automated Scientific Inference & Data Analysis
PRD5: Inner-Loop
Hybrid Machine Learning, Models, & Algorithms
PRD6: Outer-Loop
Automated Decision Support, Optimization, Resilience, & Control
Scientific Machine Learning has widespread Science & Energy uses.

Three Capability PRDs (and combinations) seem to cover most examples:

- Data-Intensive SciML
- Inner-Loop SciML
- Outer-Loop SciML

Compelling Big Science use cases include:

- Improved operational capabilities of scientific user facilities
- Better computational models from data-compute convergence
- Automation & adaptivity within scientific method (systems, processes)

...Many more...
Core Research Agenda for Scientific Machine Learning
ASCR Applied Math: Scientific Machine Learning for Transforming Science & Energy Research

Future: Smarter Ecosystems & Versatile Capabilities for Science & Energy Research

Simulation Science

Data Science

Big Science

Scientific Computing Research & Software

Scientific User Facilities, Networks, Testbeds, Workforce & Resources

A Key to the Future: Scientific Machine Learning Research

ASCR Applied Math Core Research Initiatives

Complex, Distributed, Interconnected Systems (2008)
Multiscale Mathematics (2005)
Petascale Scientific Data Analysis (2009)

Data-Intensive SciML: Automated Inference

Uncertain, Noisy
Probabilistic
High-dimensional
Streaming

Inner-Loop SciML: Hybrid Machine Learning, Models & Algorithms

Machine Learning-embedded Models
Machine Learning-embedded Algorithms

Decision-Support
Optimization
Resilience
Control

Outer-Loop SciML: Automated Complex Systems & Processes

Uncertain, Noisy
Probabilistic
High-dimensional
Streaming

Uncertain, Noisy
Probabilistic
High-dimensional
Streaming

Uncertain, Noisy
Probabilistic
High-dimensional
Streaming

Uncertain, Noisy
Probabilistic
High-dimensional
Streaming

Uncertain, Noisy
Probabilistic
High-dimensional
Streaming
Workshop Charge Letter: Scientific Machine Learning (SciML) for transforming the Future of Science & Energy research #3
Machine Learning is a powerful scientific enabling technology

- More than Data. Also for Modeling, Complex Systems, Science
- Basic research in scientific computing & mathematical foundations is essential
- Fast moving area → Need roadmap, blueprint, strategy
- Compelling: Re-visit ML, Re-think scientific computing uses

Pump is Primed for DOE leadership

- Roots from previous decade(s) of Applied Math basic research
- Ready: Researchers & expertise, Professional communities, etc

Future of Science & Energy Research

- Advanced technologies: More complex, more heterogeneous
- Greater Automation & Adaptivity for research breakthroughs

Scientific Machine Learning Priority Research Directions are a basis for a cross-cutting research initiative toward this future