H2O4GPU

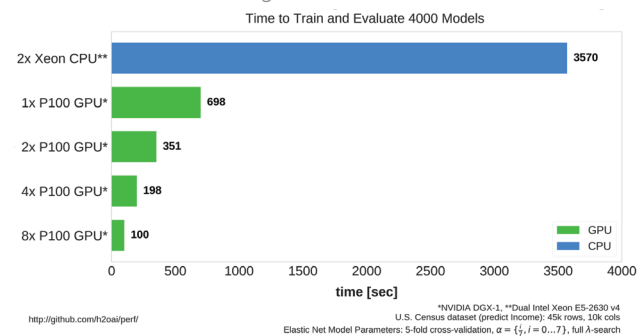H2O4GPU Platform Powered by GPUs for Lightning-Fast Model Building

H2O4GPU is an open-source collection of GPU solvers created by H2O.ai. It builds on the easy-to-use scikit-learn API and its well-tested CPU-based algorithms. It can be used as a drop-in replacement for scikit-learn with support for GPUs on selected (and ever-growing) algorithms. H2O4GPU inherits all the existing scikit-learn algorithms and falls back to CPU algorithms when the GPU algorithm does not support an important existing scikit-learn class option.

Today, select algorithms are GPU-enabed. These include Gradient Boosting Machines (GBM's), Generalized Linear Models (GLM's), and K-Means Clustering.

Gradient Linear Model (GLM)

- Framework utilizes Proximal Graph Solver (POGS)

- Solvers include Lasso, Ridge Regression, Logistic Regression, and Elastic Net Regularization

Generalized Linear Modeling

Time to Train and Evaluate 4000 Models

	time [sec]
2x Xeon CPU**	3570
1x P100 GPU*	698
2x P100 GPU*	351
4x P100 GPU*	198
8x P100 GPU*	100

http://github.com/h2oai/perf/

*NVIDIA DGX-1, **Dual Intel Xeon E5-2630 v4
U.S. Census dataset (predict income): 45k rows, 10k cols
Elastic Net Model Parameters: 5-fold cross-validation, $\alpha = \{\frac{i}{7}, i = 0...7\}$, full λ-search

SPECIFICATIONS

Software

- PC with Ubuntu 16.04+

- Install CUDA with bundled display drivers CUDA 8 or CUDA 9

Hardware

- Nvidia GPU with Compute Capability >= 3.5

H2O4GPU ROAD MAP

Currently available

- GLM (POGS)
- Pyton API for scoring and training
- GBM
- Inference on GPU (GLM)
- Random Forest
- Inference on GPU (GBM)
- k-Means clustering
- Scikit learn API for compatibility
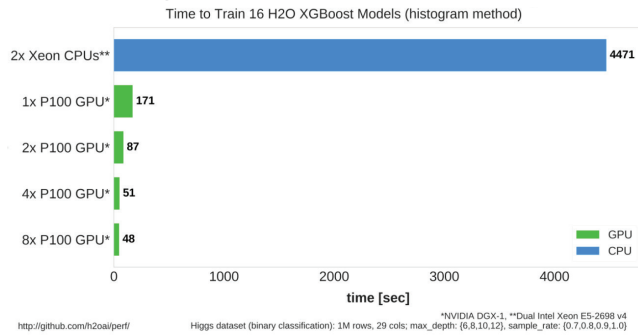- PCA
- R API for training and scoring
- SVD

Coming Q2 2018

- k-Nearest Neighbors
- Matrix Factorization
- Factorization Machines
- Quantiles
- Kalman Filters
- Sort
- Aggregator
- API Support:
  - GOAI API support
  - Data.table
- Performance & Scalability:
  - Multi machine

Q4 2018

- Kernel Methods
- Recommendation Engines - Non-Negative Matrix Factorization Recommendation Engines - Bayesian Neural Nets
- MCMC Solver
- Time Series
- SVM
- Text Analysis-TF-IDF
- Text Analysis - Word2Vec
- Text Analysis -0oc2Vec
- Automatic K for K-means
- H20 GLM - Lasso
- Simulation Techniques
- Sampling Techniques
- Domain Specific Algorithms:
  - Life Sciences
  - Financial Services Underwriting
  - Sampling Techniques

Gradient Boosting Machines

Time to Train 16 H2O XGBoost Models (histogram method)

	time [sec]
2x Xeon CPUs**	4471
1x P100 GPU*	171
2x P100 GPU*	87
4x P100 GPU*	51
8x P100 GPU*	48

Legend: GPU (green), CPU (blue)

http://github.com/h2oai/perf/   *NVIDIA DGX-1, **Dual Intel Xeon E5-2698 v4
Higgs dataset (binary classification): 1M rows, 29 cols; max_depth: {6,8,10,12}, sample_rate: {0.7,0.8,0.9,1.0}

- Improvements to original implementation of POGS:
  - Full alpha search
  - Cross Validation
  - Early Stopping
  - Added scikit-learn-like API
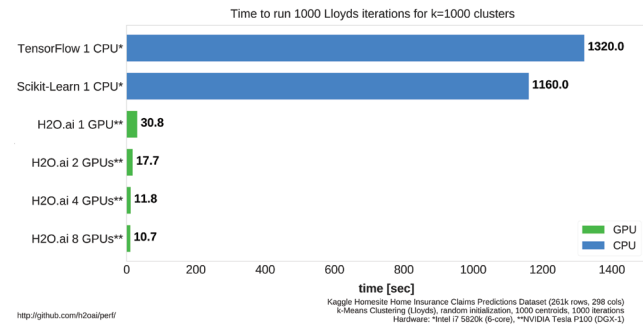 - Supports multiple GPU's

Gradient Boosting Machines

- Based on XGBoost
- Raw floating point data — binned into quantiles
- Quantiles are stored as compressed instead of floats
- Compressed quantiles are efficiently transferred to GPU
- Sparsity is handled directly with high GPU efficiency
- Multi-GPU enabled by sharing rows using NVIDIA NCCL AllReduce

k-Means Clustering

- Based on NVIDIA prototype of k-Means algorithm in CUDA
- Improvements to original implementation:
  - Significantly faster than scikit-learn implementation (50x) and other GPU implementations (5-10x)
  - Supports multiple GPU's

k-Means Clustering

Time to run 1000 Lloyds iterations for k=1000 clusters

	time [sec]
TensorFlow 1 CPU*	1320.0
Scikit-Learn 1 CPU*	1160.0
H2O.ai 1 GPU**	30.8
H2O.ai 2 GPUs**	17.7
H2O.ai 4 GPUs**	11.8
H2O.ai 8 GPUs**	10.7

Legend: GPU (green), CPU (blue)

http://github.com/h2oai/perf/   Kaggle Homesite Home Insurance Claims Predictions Dataset (261k rows, 298 cols)
k-Means Clustering (Lloyds), random initialization, 1000 centroids, 1000 iterations
Hardware: *Intel i7 5820k (6-core), **NVIDIA Tesla P100 (DGX-1)

RESOURCES

- **Github:** https://github.com/h2oai/h2o4gpu
- **FAQ:** https://github.com/h2oai/h2o4gpu/blob/master/FAQ.md

H2O.ai is focused on bringing AI to businesses through software. Its flagship product is H2O, the leading open source platform that makes it easy for financial services, insurance and healthcare companies to deploy machine learning and predictive analytics to solve complex problems. More than 13,000 organizations and 130,000+ data scientists depend on H2O for critical applications like predictive maintenance and operational intelligence. The company accelerates business transformation for 222 Fortune 500 enterprises, 8 of the world's 12 largest banks, 7 of the 10 largest auto insurance companies and all 5 major telecommunications providers.

Follow us on Twitter @h2oai. To learn more about H2O customer use cases, please visit http://www.h2o.ai/customers/. Join the Movement.