AutoML + AI

### ...GPU accelerated Neural Networks & Bayesian Optimization

# Preamble

* Live Demos
  + Python
  + Notebooks
  + Docker
    - Postgres
    - “Lambda”
* AWS
* Neural Nets (Keras, TF)
* GPU
  + Mac - Radeon
  + Mac - UHD
  + CUDA - Colab
* Apple vs Nvidia

# What is AutoML?

* Machine Learning Process
  1. Algorithm (algo) + Hyperparameters (hp) + Data Features (df) = Model
  2. Tuning:
     1. Choose: algo, hps, dfs
     2. Induce
     3. Test
     4. Adjust based on results
     5. GOTO i.
* AutoML is the automation of this process:
  1. Choose Limits for: algo, hps, dfs
  2. Auto: Induce
  3. Auto: Test
  4. Auto: Adjust algo, hps, dfs (via Bayesian Optimization)
  5. GOTO a.
* Bayes + Loss → Hyperparams
* What about data prep and feature selection?
  1. Some algos blur the lines

# Data Prep

* Continuous [discretization]
  + Impute? What type?
* Bins / Categories (One-Hot Encode)
  + Impute? What type?
* Binary
  + Impute?

# Algorithms

# Commercial Options

* Dataiku
  + my favorite
  + Multi-level
  + GPU support
  + Docker/K8s support
  + Python code export
  + all libraries
  + Deloitte SNET
  + Easily Deployable
  + demo
* DataRobot (pure play)
  + Deloitte SNET
  + Limited deployment
  + “Black Boxed”
  + Demo
* H2O Driverless AI (upstart, pure play product)
  + GPU support
  + Docker support
  + Easily Deployable
  + demo

# AutoML Lib Overview

* Optuna
  + <https://optuna.org/>
* Hyperopt
  + <http://hyperopt.github.io/hyperopt/>
* BayesianOptimization
  + <https://github.com/fmfn/BayesianOptimization>
* Ray Tune
  + <https://ray.readthedocs.io/en/latest/tune.html>
* H2O
  + <http://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html>
* AutoKeras
  + <https://autokeras.com/>

# Sample Data Sets

* NLP
  + <https://www.yelp.com/dataset/>
* Image
* Fraud
* Traditional (Iris, auto mpg, titanic)

# Optuna

Simple Demo

+DB

+cluster

+AWS

# Cyclomatic Complexity

## 