H2O Deep Learning

Arno Candel, PhD

Chief Architect, H2O.ai

H2O WORLD 2015

Need a handy conference guide?

Download our app, “H2O World 2015”

H2O WORLD

What do these stickers mean?

I have H2O Installed

I have R installed

I have Python installed

I have the H2O World data sets

Pick up stickers or get install help at the information booth

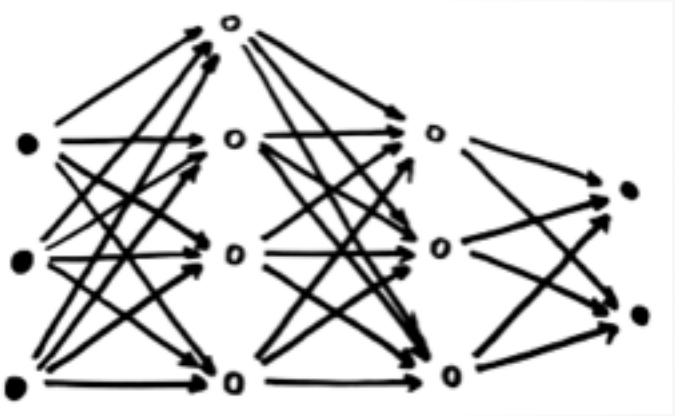
H2O WORLD

Why Deep Learning?

- Deep Learning is trending (so it must be useful)

What is Deep Learning?

- Deep Learning learns a hierarchy of non-linear transformations
- Black-box, brute-force method, really good at pattern recognition
- Deep Learning got a boost in the last decade due to faster hardware and algorithmic advances
- Results are generally not 100% reproducible, but within a small tolerance



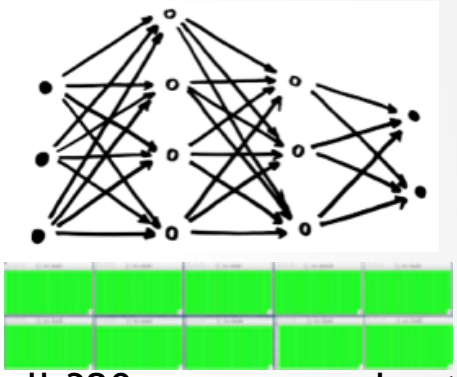
model = set of connecting weights + type of non-linearity in each layer

What is H2O Deep Learning?

Multi-layer Artificial Neural Network trained with back-propagation (stochastic gradient descent)

+ distributed processing for big data

+ multi-threaded speedup



all 320 cores maxed out

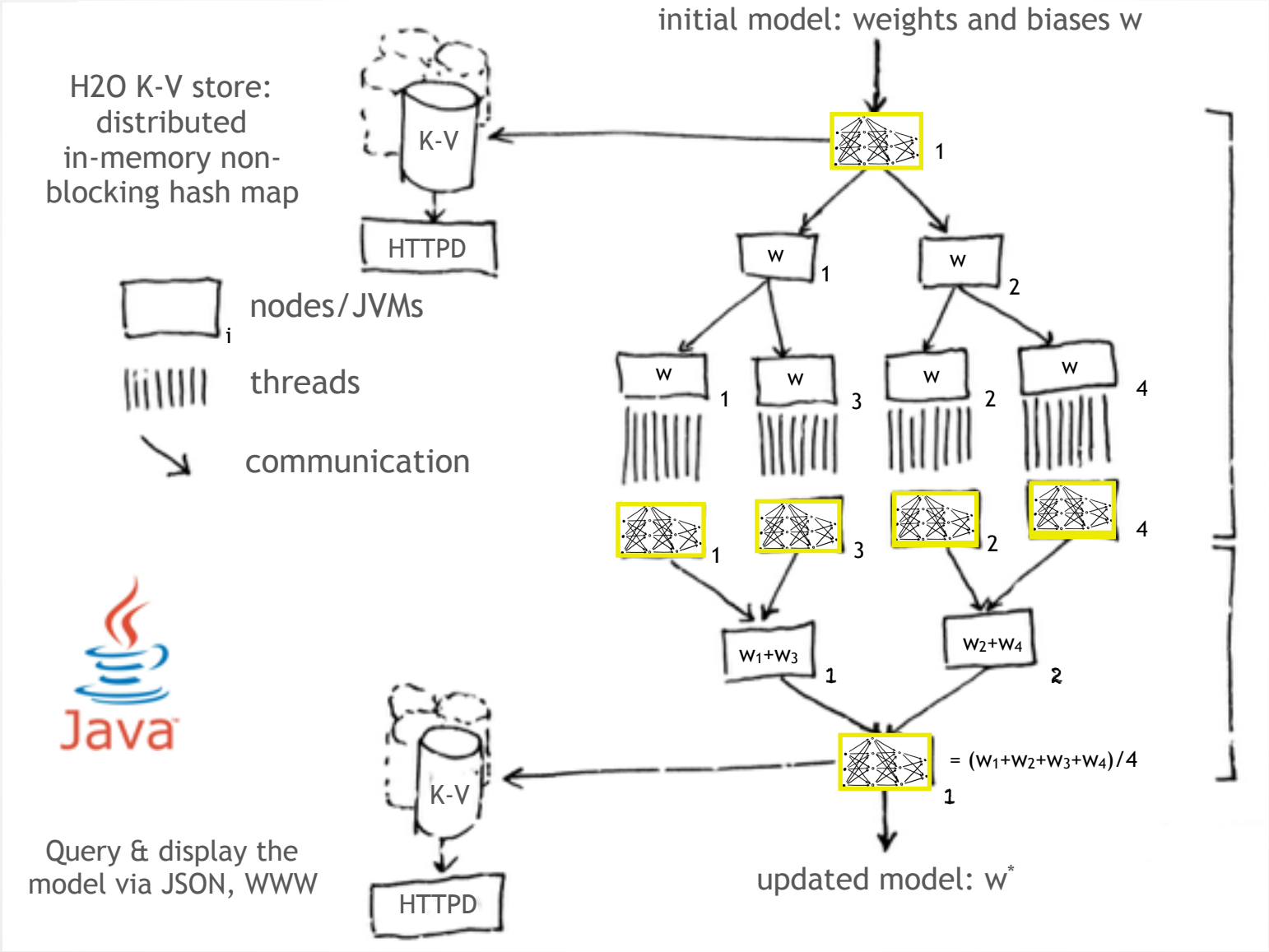
+ smart algorithms for fast & accurate results

(early stopping with convergence checking, adaptive learning rate, automatic standardization, automatic one-hot encoding of categoricals, missing value imputation, automatic model initialization, momentum, dropout/L1/L2 regularization, non-Gaussian distributions and advanced loss functions, offsets, observation weights, grid search, N-fold cross-validation, checkpointing, load balancing, auto-tuning, model averaging, elastic averaging, etc.)

= powerful tool for (un)supervised machine learning on real-world data

H2O WORLD

H2O DL Map/Reduce Iterations



H2O Deep Learning Features

- Regularization techniques: Dropout, L1/L2
- Efficient handling of categorical, missing and sparse data
- Advanced math: Gaussian/Laplace/Poisson/Gamma/Tweedie regression, offsets, observation weights, loss functions
- Unsupervised mode for non-linear dimensionality reduction, anomaly detection, etc.
- H2O Eco-System Benefits:
  - Scalable to massive datasets on large clusters
  - Early convergence-based stopping, N-fold cross-validation, Grid search
  - Low-latency Java ("POJO") scoring code is auto-generated
  - Easy deployment: One .jar for Laptop, Server, Hadoop, Spark
  - APIs include R, Python, Flow UI, Scala, Java, JavaScript, REST
  - Easy to use: like GBM, DRF, GLM, can mix & match, compare

More Tomorrow!

Tuesday 11:00 AM Erdos Stage

Top 10 Deep Learning Tips & Tricks

Hands-On Tutorial

- Introduction
  - Installation and Startup
  - Decision Boundaries
- Cover Type Dataset
  - Exploratory Data Analysis
  - Deep Learning Model
  - Hyper-Parameter Search
  - Checkpointing
  - Cross-Validation
  - Model Save & Load
- Regression and Binary Classification
- Deep Learning Tips & Tricks (more tomorrow!)