

Deep Water



Jo-fai (Joe) Chow

Data Scientist

joe@h2o.ai

Paris Machine Learning Meetup

Murex

21st September, 2016

Agenda

- **This Talk (30 mins)**

- About H2O.ai
- Demos
 - Web Interface
 - R
- Why H2O?
- What's Next?

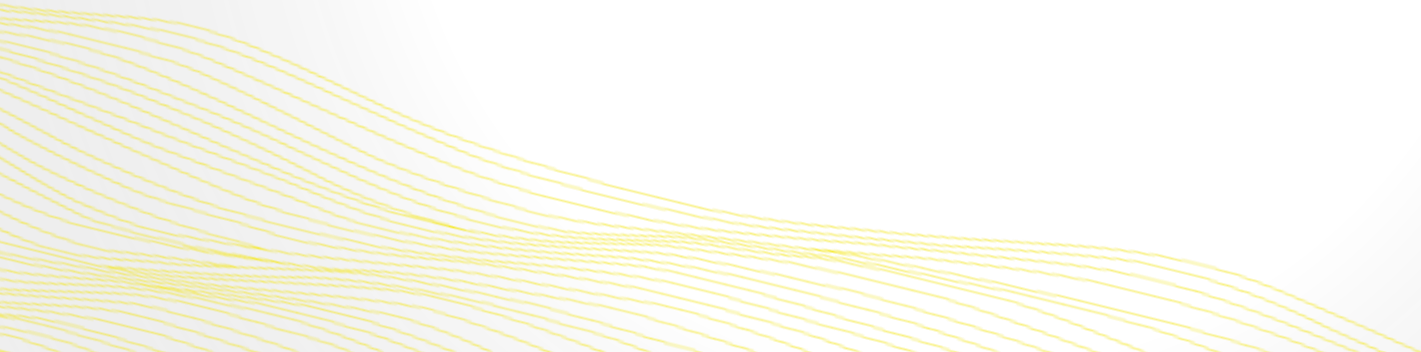
- **Second Talk (30 mins)**

- Deep Water
- Demos
 - H2O + mxnet
 - H2O + TensorFlow

- **Third Talk (45 mins)**

- H2O + Spark = Sparkling Water
- Demos
 - H2O + Spark MLlib

Deep Learning in H2O



H2O Overview

Computer Science (CS)

Artificial Intelligence (A.I.)

Machine Learning (ML)

Deep Learning (DL)

hot hot hot hot hot

H₂O.ai

A Simple Neural Network

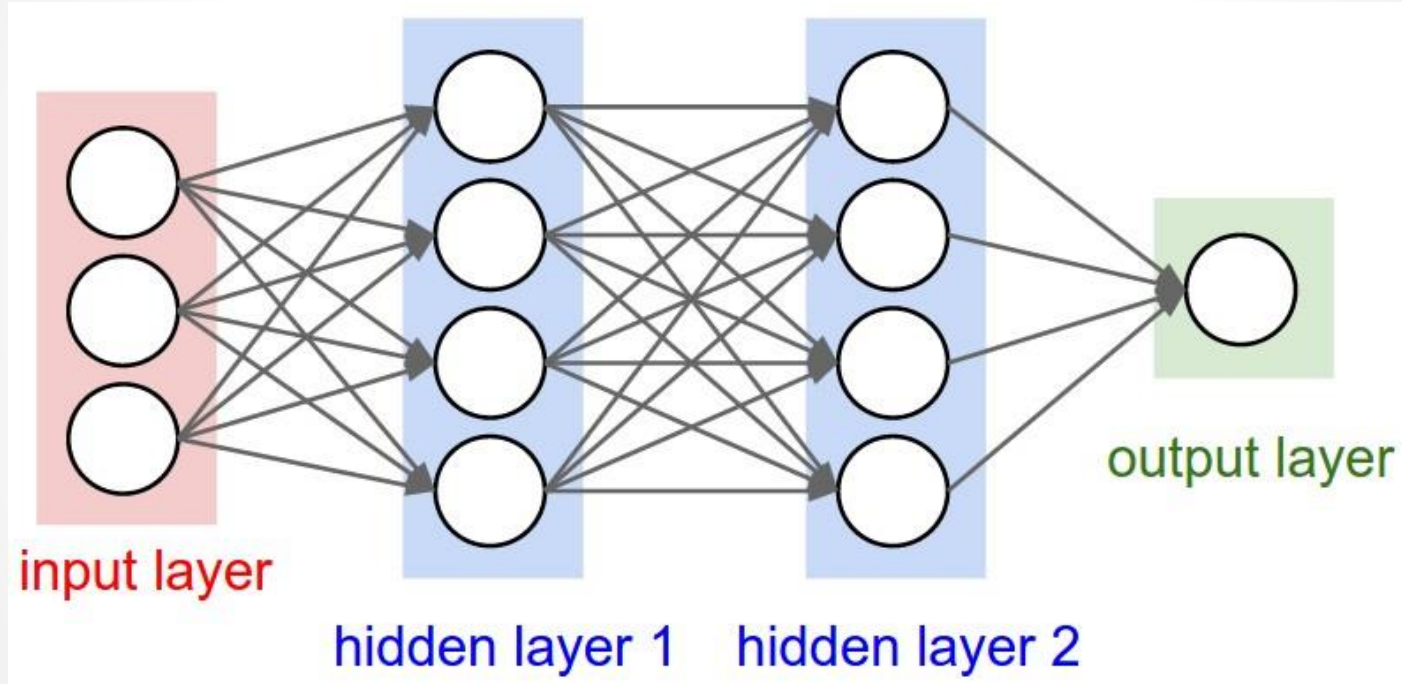


Image credit: <http://cs231n.github.io/>

H2O Deep Learning in Action

116M rows, 6GB CSV file
800+ predictors (numeric + categorical)

airlines_all_selected_cols.hex

Actions: View Data Split Build Model Predict Download Export

| Rows | Columns | Compressed Size |
|-----------|---------|-----------------|
| 116695259 | 12 | 2GB |

Job

Run Time 00:00:36.712

Remaining Time 00:00:17.188

Type Model

Key Q deeplearning-dd2f42f7-81f7-42e8-9d98-e34437309828

Description DeepLearning

Status RUNNING

Progress 69%

Iterations: 12. Epochs: 0.628821. Speed: 2,243,735 samples/sec. Estimated time left: 21.849 sec

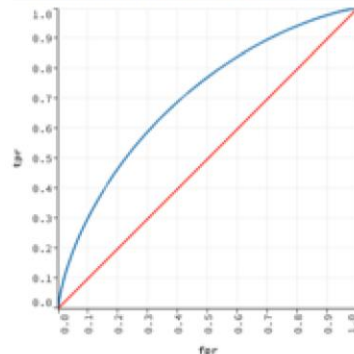
Actions View Cancel Job

model trained in <1 min:
2M+ samples/second

OUTPUT - STATUS OF NEURON LAYERS (PREDICTING ISDEPDELAYED, 2-CLASS CLASSIFICATION, BERNOLLI DISTRIBUTION, CROSSENTROPY LOSS, 17.462 WEIGHTS/BIASES, 221.9 KB, 106.585,365 TRAINING SAMPLES, MINI-BATCH SIZE 1)

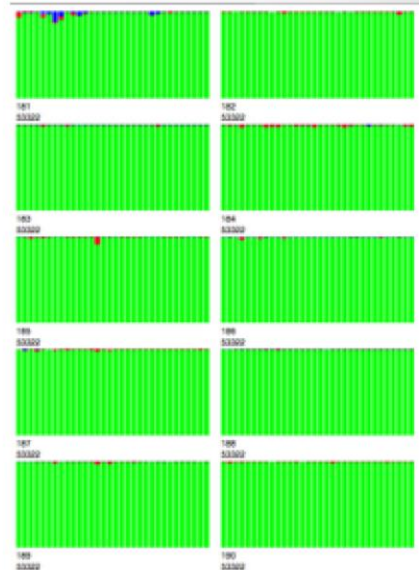
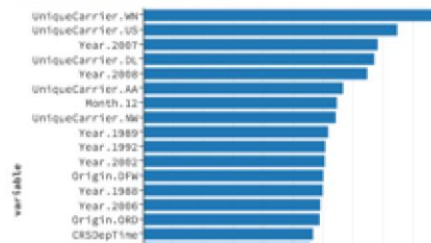
| layer | units | type | dropout | l1 | l2 | mean_rate | rate_RMS | momentum | mean_weight | weight_RMS | mean_bias | bias_RMS |
|-------|-------|-----------|---------|----|----|-----------|----------|----------|-------------|------------|-----------|----------|
| 1 | 887 | Input | 0 | | | | | | | | | |
| 2 | 20 | Rectifier | 0 | 0 | 0 | 0.0493 | 0.2020 | 0 | -0.0021 | 0.2111 | -0.0139 | 1.0036 |
| 3 | 20 | Rectifier | 0 | 0 | 0 | 0.0197 | 0.0227 | 0 | -0.1893 | 0.5362 | -1.3998 | 1.5259 |
| 4 | 20 | Rectifier | 0 | 0 | 0 | 0.0517 | 0.0446 | 0 | -0.1575 | 0.3068 | -0.0846 | 0.6046 |
| 5 | 20 | Rectifier | 0 | 0 | 0 | 0.0761 | 0.0844 | 0 | -0.0374 | 0.2275 | -0.2647 | 0.2481 |
| 6 | 2 | Softmax | 0 | 0 | 0 | 0.0161 | 0.0083 | 0 | 0.0741 | 0.7260 | 0.4269 | 0.2056 |

ROC CURVE - VALIDATION METRICS, AUC = 0.702560



Threshold: Choose... Criterion: Choose...

VARIABLE IMPORTANCES



Legend

Each bar represents one CPU.

Blue: idle time

Green: user time

Red: system time

White: other time (e.g. I/O)

10 nodes: all
320 cores busy

H2O.ai

Deep Learning Model

real-time, interactive
model inspection in Flow



H2O Deep Learning Community Quotes

**CIFAR-10 Competition
Winners: Interviews with Dr.
Ben Graham, Phil Culliton, &
Zygmunt Zajac**

Triskelion | 01.02.2015

“I did really like H2O’s deep learning implementation in R, though - the interface was great, the back end extremely easy to understand, and it was scalable and flexible. Definitely a tool I’ll be going back to.”

[READ MORE](#)

**Kaggle challenge
2nd place winner
Colin Priest**

for creating this corpus. .
do not contain Spanish sent-
is a widespread major langu-
reason was to create a corp-
tasks. These tasks are com

Completed • Knowledge • 161 teams

Denoising Dirty Documents

Mon 1 Jun 2015 – Mon 5 Oct 2015 (3 months ago)

“For my final competition submission I used an ensemble of models, including 3 deep learning models built with R and h2o.”

[READ MORE](#)

H2O Deep Learning Community Quotes



Arno Candel @ArnoCandel · Jun 29

Great use of H2O #DeepLearning for 3rd place at #Kaggle Homesite challenge! github.com/mpearmain/home... (see R scripts)

```
d1.model <- h2o.deeplearning(  
  # data specifications  
  x = xrange, y = max(xrange)+1, training_frame = x0.hex,  
  autoencoder = FALSE,  
  # network structure: activation and geometry  
  activation = "RectifierWithDropout",  
  hidden = c(size1, size2), epochs = 25,  
  input_dropout_ratio = 0.05, hidden_dropout_ratios = c(0.05, 0.02),  
  # parameters of the optimization process  
  rho = 0.99, epsilon = 1e-08, rate = 0.005,  
  rate_annealing = 1e-06, rate_decay = rate_dec, momentum_start = 0.5,  
  l1 = 0, l2 = 0, loss = c("CrossEntropy")  
)
```



18



33



Completed • \$20,000 • 1,764 teams

Homesite Quote Conversion

Mon 9 Nov 2015 – Mon 8 Feb 2016 (5 months ago)

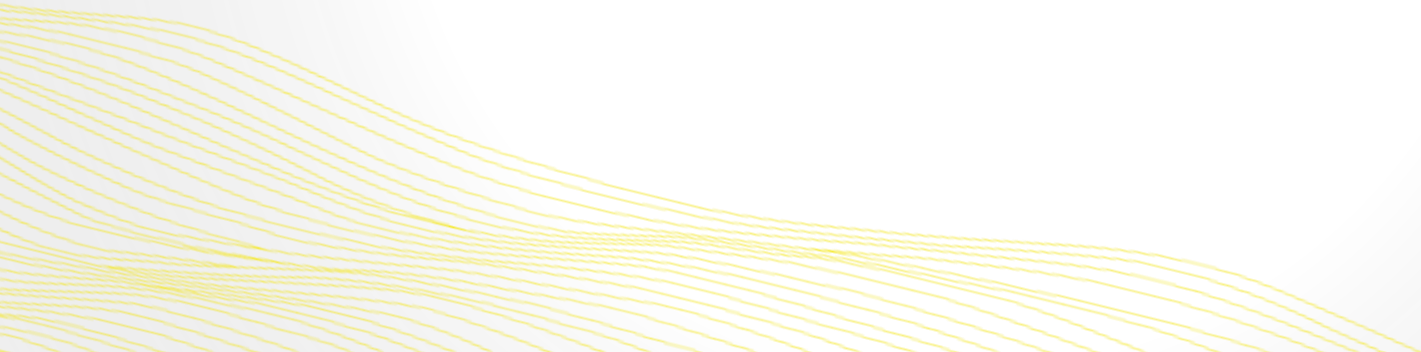
Dashboard

Private Leaderboard - Homesite Quote Conversion

This competition has completed. This leaderboard reflects the final standings.

| # | Rank | Team Name | Model uploaded • In the money | Score | Entries | Last St |
|---|------|----------------------------|-------------------------------|---------|---------|---------|
| 1 | — | KazAnova Faron clobber | 🏆 🌟 | 0.97024 | 350 | Mon, |
| 2 | — | Frenchies | 🏆 🌟 | 0.97019 | 241 | Mon, |
| 3 | 1st | New Model Army CAD & QuY | 🏆 🌟 | 0.97002 | 234 | Mon, |
| <ul style="list-style-type: none">Konrad BanachewiczMike PearmainCharles-Abner DADIquentin.y | | | | | | |

Why Deep Water?






What is Deep Water?



Photo credit: en.wikipedia.org/wiki/Deepwater_Horizon_explosion

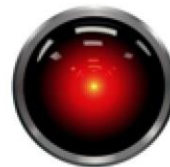
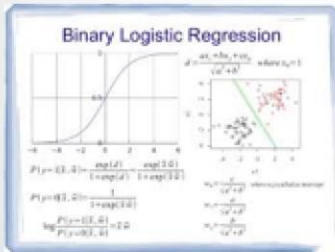
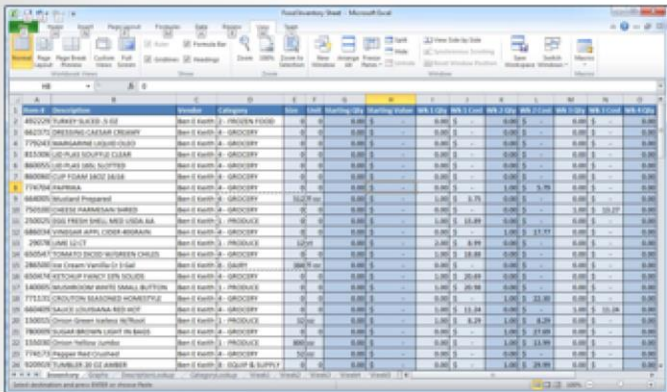
Deep Water: Best Open-Source Deep Learning

Enterprise Deep Learning for Business Transformation

| | | |
|--|---|---|
| Deep Water = THE Deep Learning Platform | H2O integrates the top open-source DL tools |  |
| Native GPU support |  is up to 100x faster than  | |
| Enterprise Ready | Easy to train and deploy, interactive, scalable, etc. Flow, R, Python, Spark/Scala, Java, REST, POJO, Steam | |
| New Big Data Use Cases (previously impossible or difficult in H2O) | Image - social media, manufacturing, healthcare, ... Video - UX/UI, security, automotive, social media, ... Sound - automotive, security, call centers, healthcare, ... Text - NLP, sentiment, security, finance, fraud, ... Time Series - security, IoT, finance, e-commerce, ... | |

More Data is Better! — Images, Video, Text, Logs, Streams, ...

Example: Fraud Prediction



Today



Tomorrow

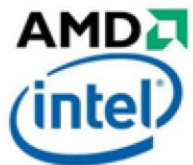
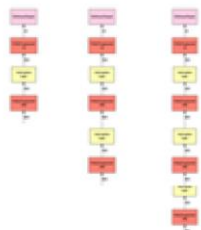
Better Data — Better Models — Better Results

H₂O.ai

Deep Water opens the Floodgates for state-of-the-art Deep Learning

H2O Deep Learning: simple multi-layer neural networks

1-5 layers
MBs/GBs of data

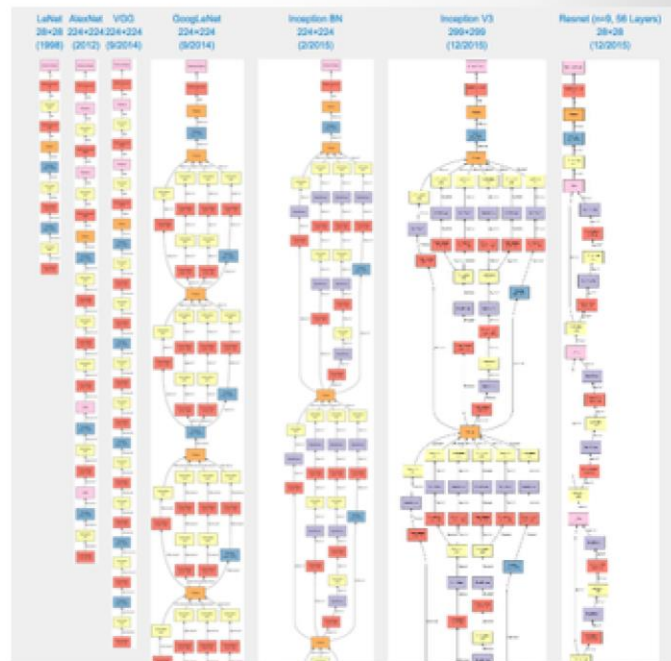
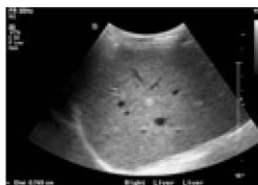


Limited to business analytics,
statistical models (CSV data)

H₂O.ai

Deep Water: deep complex networks

5-1000 layers
GBs/TBs of data



Large networks for big data
(e.g. image 1000x1000x3 -> 3m inputs)

Current Contributors (more H2O.ai folks joining soon)



Fabrizio Milo



Cyprien Noel



Qiang Kou



Arno Candel



Caffe



H₂O.ai

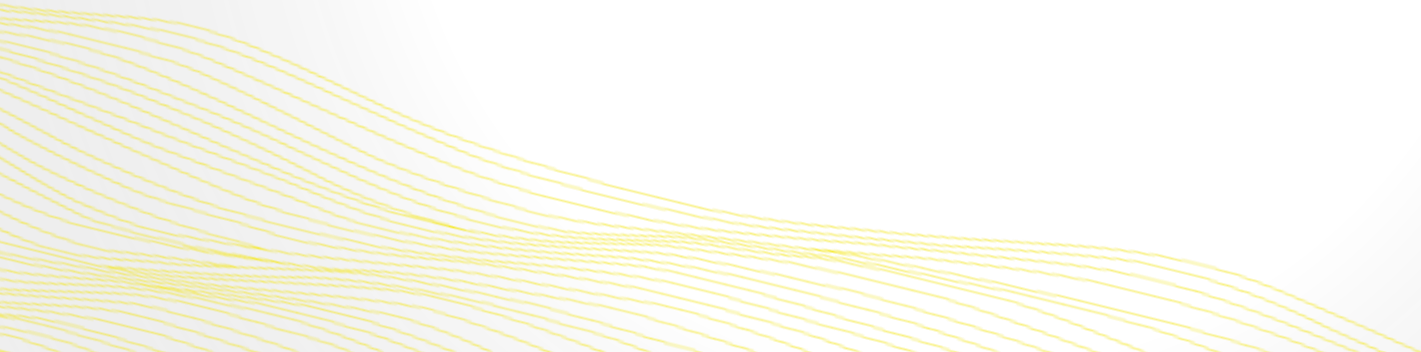


This repository

Search

h2oai / deepwater

Deep Water Demos



Deep Water Demos

- **H2O + mxnet**

- Dataset:
 - Cat / Dog / Mouse
- H2O Python interface
- mxnet GPU backend
- Train a LeNet (CNN) model
- Explore model in Flow

- **H2O + TensorFlow**

- Dataset:
 - MNIST hand-written digits
- Sparkling Water + Jupyter Notebook
- Convert TensorFlow model into H2O
- Explore models in Flow

For Online Audience

- **H2O + mxnet**

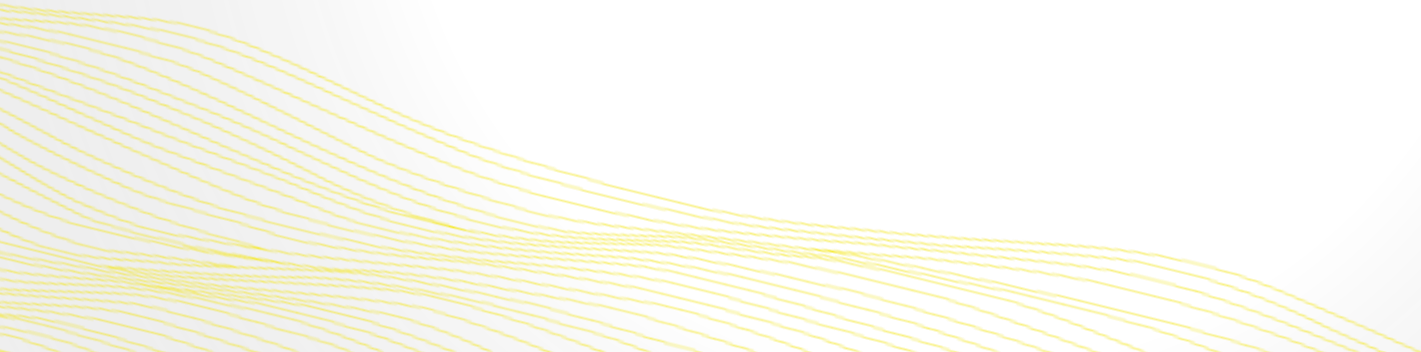
- bit.ly/h2o_paris_1
- **demo_03_mxnet**

- **H2O + TensorFlow**

- bit.ly/h2o_paris_1
- **demo_04_tensorflow**

H2O + mxnet

H2O + TensorFlow



Merci Beaucoup!

- Jiqiong (Ji) Qiu
- Franck Bardol
- Igor Carron
- Slides & Code
 - bit.ly/h2o_paris_1
- Contact
 - joe@h2o.ai
 - [@matlabulous](#)
 - github.com/woobe

