Deep Water

Jo-fai (Joe) Chow
Data Scientist
joe@h2o.ai
@matlabulous

Agenda

- First Talk (25 mins)
  - About H2O.ai
  - Demo
    - A Simple Classification Task
    - H2O's Web Interface
  - Why H2O?
    - Our Community
    - Our Customers
  - What's Next?
    - New H2O Features

- Second Talk (25 mins)
  - H2O for IoT
    - Predictive Maintenance
    - Anomaly Detection
    - H2O's R Interface

- Third Talk (25 mins)
  - Deep Water
  - Demo
    - H2O + mxnet on GPU
    - H2O's Python Interface

Deep Learning in H2O
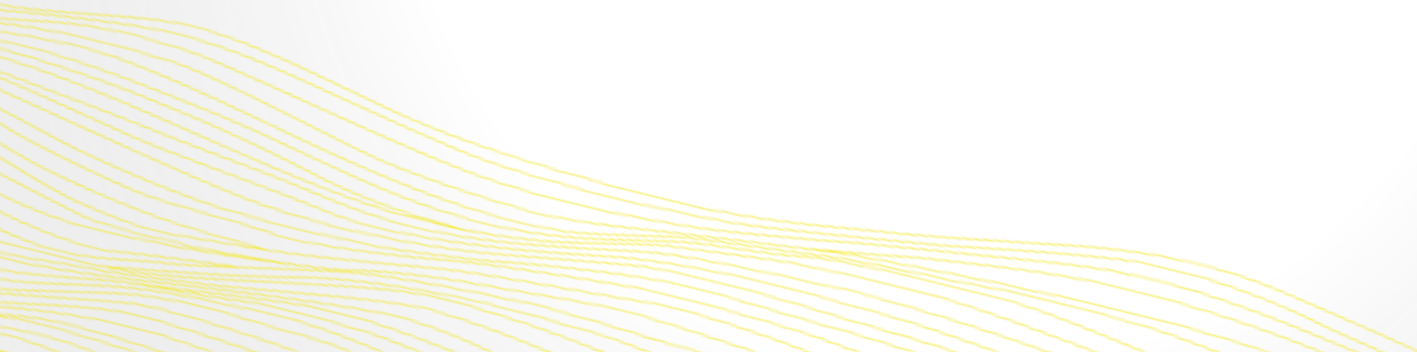
Computer Science (CS)

Artificial Intelligence (A.I.)

Machine Learning (ML)

Deep Learning (DL)
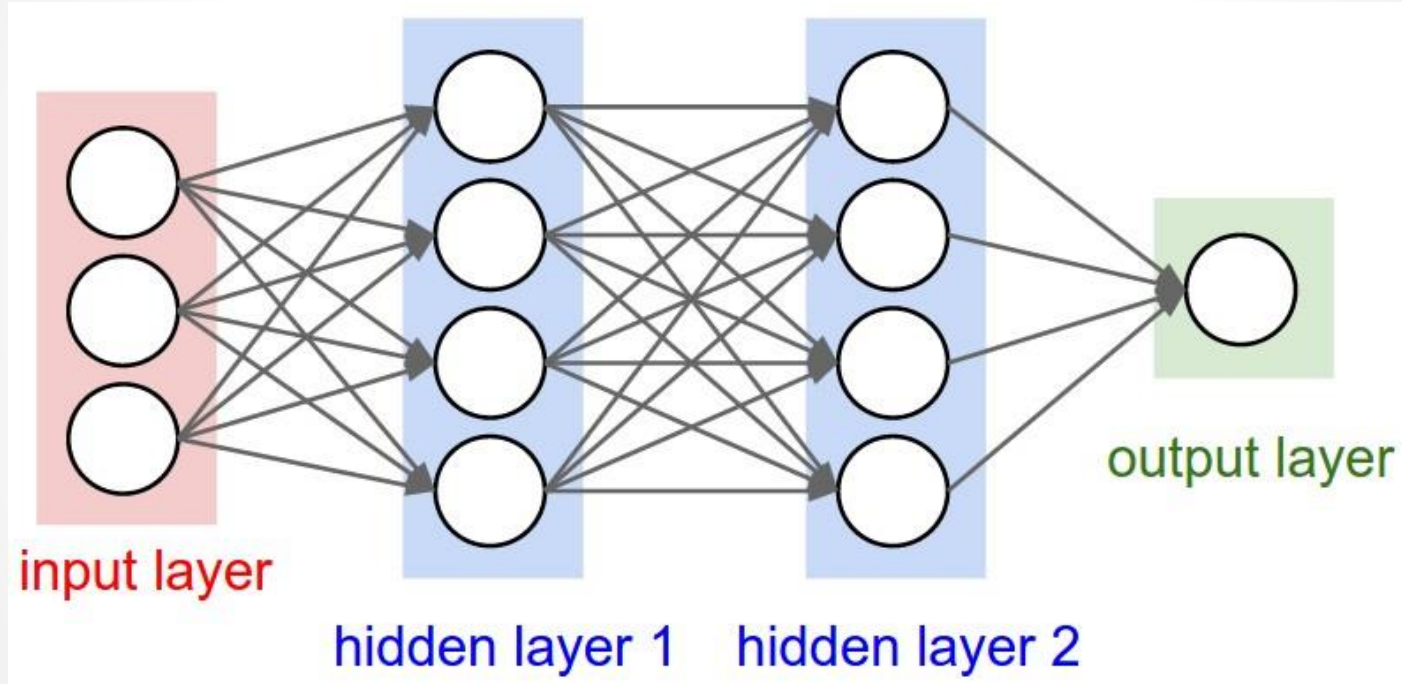
hot hot hot hot hot

H₂O.ai

A Simple Neural Network

input layer

hidden layer 1   hidden layer 2

output layer

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 deeplearning-dd2f42f7-81f7-42e8-9d98-e34437309828
Description DeepLearning
Status RUNNING
Progress 69%

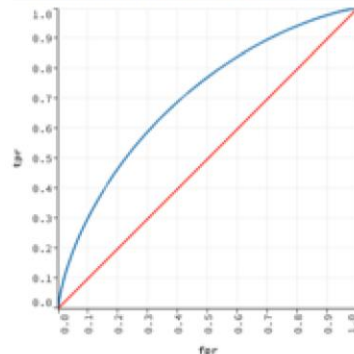Iterations: 12. Epochs: 0.628821. Speed: 2,243,735 samples/sec. Estimated time left: 21.849 sec

Actions Q View | Cancel Job

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

* OUTPUT - STATUS OF NEURON LAYERS (PREDICTING ISDEPDELAYED, 2-CLASS CLASSIFICATION, BERNOULLI DISTRIBUTION, CROSSENTROPY LOSS, 17,462 WEIGHTS/BIASES, 221.3 KB, 106,585,365 TRAINING SAMPLES, MINI-BATCH SIZE 1)
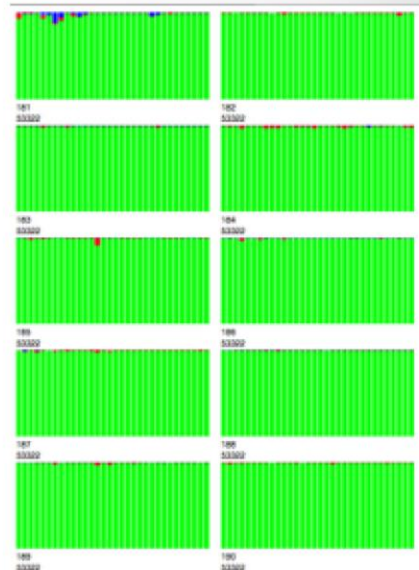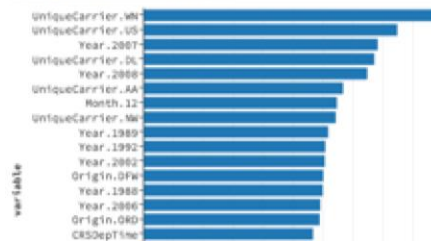
layer	units	type	dropout	l1	l2	mean_rate	rate_RMS	momentum	mean_weight	weight_RMS	mean_bias	bias_RMS
1	807	Input	0									
2	20	Rectifier	0	0	0	0.0493	0.2020	0	-0.0021	0.2111	-0.9139	1.0036
3	20	Rectifier	0	0	0	0.0197	0.0227	0	-0.1033	0.5362	-1.3988	1.5259
4	20	Rectifier	0	0	0	0.0517	0.0446	0	-0.1575	0.3068	-0.8846	0.6046
5	20	Rectifier	0	0	0	0.0761	0.0844	0	-0.8374	0.2275	-0.2647	0.2481
6	2	Softmax		0	0	0.0161	0.0083	0	0.0741	0.7260	0.4269	0.2056

Deep Learning Model

* ROC CURVE - VALIDATION METRICS , AUC = 0.702560

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

* VARIABLE IMPORTANCES

UniqueCarrier.WN
UniqueCarrier.US
Year.2007
UniqueCarrier.DL
Year.2008
UniqueCarrier.AA
Month.12
UniqueCarrier.NW
Year.1989
Year.1992
Year.2002
Origin.DFW
Year.1988
Year.2006
Origin.ORD
CRSDepTime

variable

real-time, interactive
model inspection in Flow

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

H2O Deep Learning Community Quotes

CIFAR-10 Competition Winners: Interviews with Dr. Ben Graham, Phil Culliton, & Zygmunt Zając
Triskelion | 01.02.2015

READ MORE

"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."

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
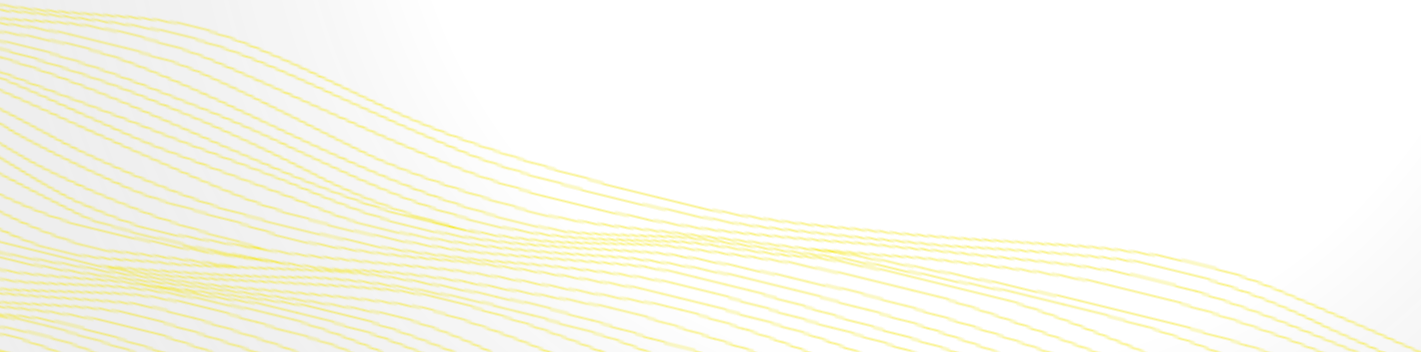tasks. These tasks are com

Completed • Knowledge • 161 teams
Denoising Dirty Documents
Mon 1 Jun 2015 – Mon 5 Oct 2015 (3 months ago)

READ MORE

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

H₂O.ai

Why Deep Water?

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, …	

H₂O.ai

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

H2O Deep Learning:
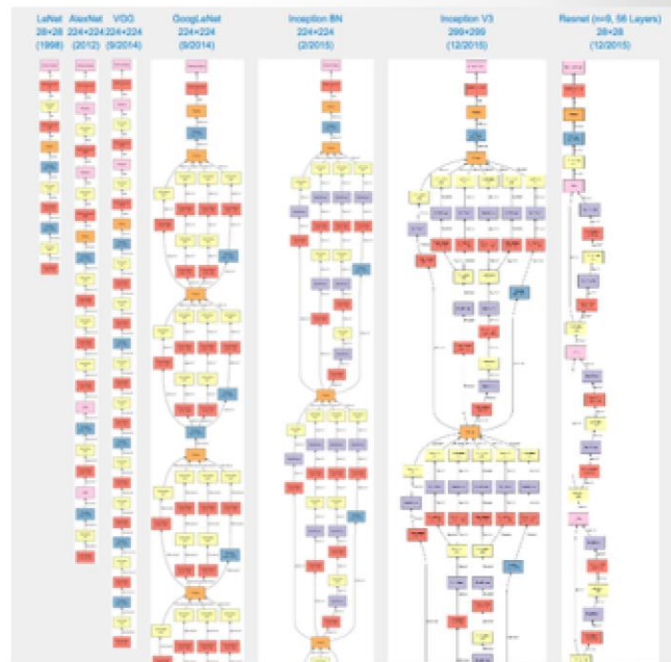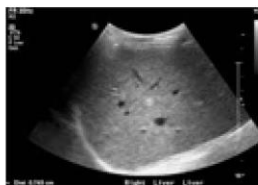### simple multi-layer neural networks

1-5 layers
MBs/GBs of data
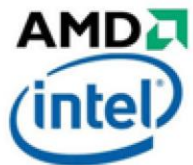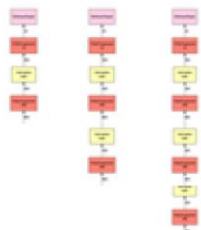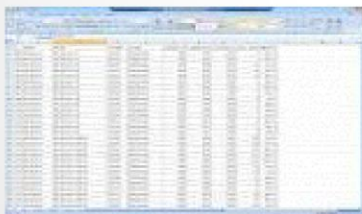
Limited to business analytics,
statistical models (CSV data)

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)

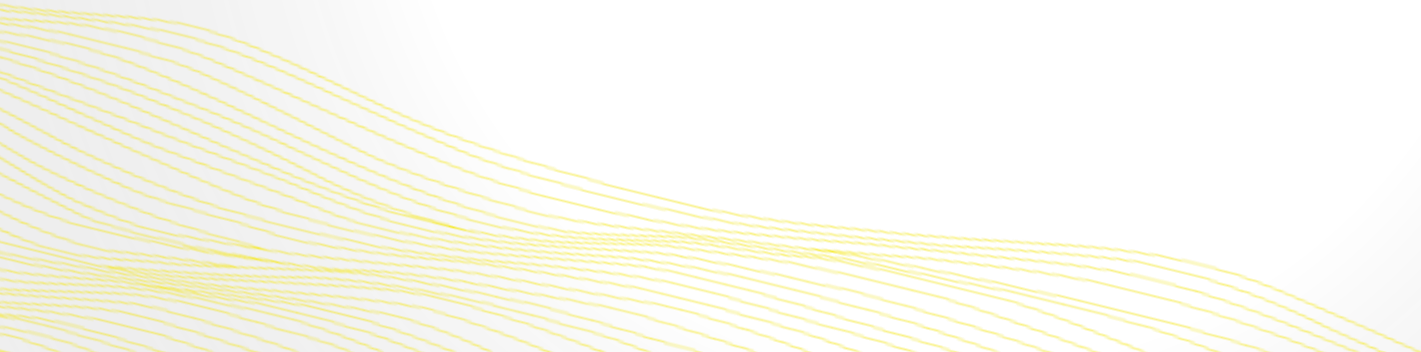Fabrizio Milo   Cyprien Noel   Qiang Kou   Arno Candel
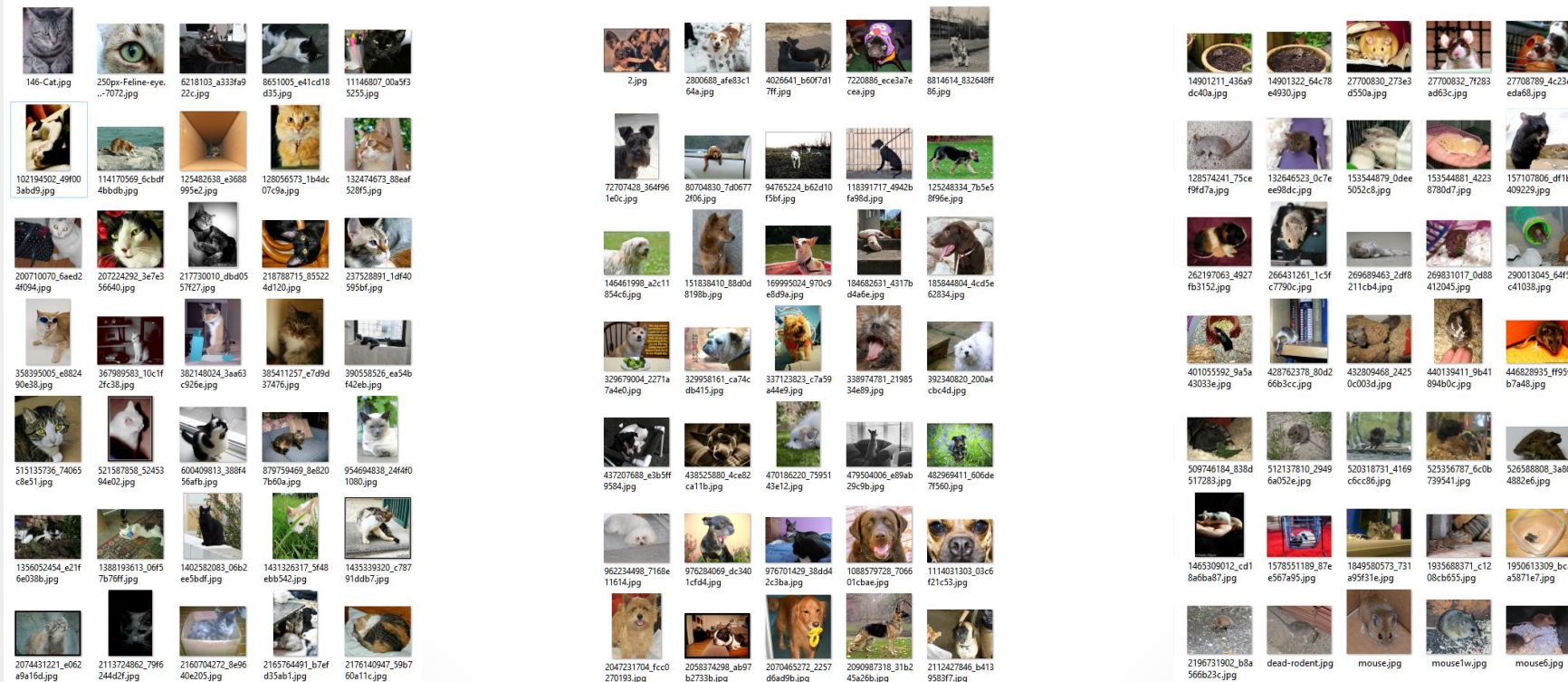
Deep Water Demo

Deep Water Demo

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

- Code and Data
  - bit.ly/h2o_milan_1
  - subfolder
    - deep_water_demo

Data – Cat/Dog/Mouse Images

Data - CSV

	A	B
1	bigdata/laptop/deepwater/imagenet/cat/102194502_49f003abd9.jpg	cat
2	bigdata/laptop/deepwater/imagenet/cat/11146807_00a5f35255.jpg	cat
3	bigdata/laptop/deepwater/imagenet/cat/1140846215_70e326f868.jpg	cat
4	bigdata/laptop/deepwater/imagenet/cat/114170569_6cbdf4bbdb.jpg	cat
5	bigdata/laptop/deepwater/imagenet/cat/1217664848_de4c7fc296.jpg	cat
6	bigdata/laptop/deepwater/imagenet/cat/1241603780_5e8c8f1ced.jpg	cat
7	bigdata/laptop/deepwater/imagenet/cat/1241612072_27ececbdef.jpg	cat
8	bigdata/laptop/deepwater/imagenet/cat/1241613138_ef1d82973f.jpg	cat
9	bigdata/laptop/deepwater/imagenet/cat/1244562192_35becd66bd.jpg	cat
10	bigdata/laptop/deepwater/imagenet/cat/125482638_e3688995e2.jpg	cat
11	bigdata/laptop/deepwater/imagenet/cat/128056573_1b4dc07c9a.jpg	cat
12	bigdata/laptop/deepwater/imagenet/cat/12945197_75e607e355.jpg	cat
13	bigdata/laptop/deepwater/imagenet/cat/132474673_88eaf528f5.jpg	cat
14	bigdata/laptop/deepwater/imagenet/cat/1350530984_ecf3039cf0.jpg	cat
15	bigdata/laptop/deepwater/imagenet/cat/1351606235_c9fbebf634.jpg	cat
16	bigdata/laptop/deepwater/imagenet/cat/1356052454_e21f6e038b.jpg	cat
17	bigdata/laptop/deepwater/imagenet/cat/1388193613_06f57b76ff.jpg	cat

H2O + mxnet Demo

End of Third Talk – Thanks!

- Data Science Milan
- Gianmario Spacagna
- Politecnico di Milano

- Resources
  - bit.ly/h2o_milan_1
  - www.h2o.ai
  - docs.h2o.ai
- Contact
  - joe@h2o.ai
  - @matlabulous
  - github.com/woobe