Deep Water

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

Data Science London
London Olympia Conference Centre
3rd November, 2016

About Me: Civil Engineer → Data Scientist

- 2005 - 2015
- Water Engineer
  - Consultant for Utilities
  - Industrial PhD
    - Water Engineering + Machine Learning
    - *Discovered H2O in 2014!*

- 2015 - Present
- Data Scientist
  - Virgin Media (UK)
  - Domino Data Lab (US)
  - H2O.ai (US)

Why? Long story – see bit.ly/joe_h2o_talk2

Agenda

- Conference Talk
  - About H2O.ai
  - Demo
  - Why H2O?
  - What's Next?
    - New H2O Features

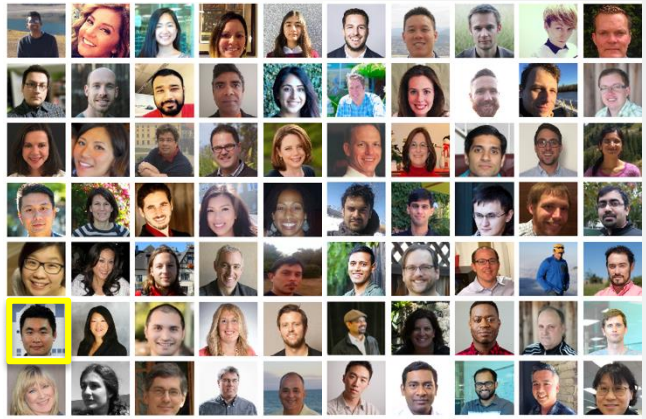
- This Talk
  - About H2O.ai
  - Deep Water
    - Motivation
    - Roadmap
    - Live Demo

About H2O.ai

About H2O.ai

- H2O.ai, the Company
  - Team: 80 (70 shown)
  - Founded in 2012
  - HQ: Mountain View, California

- H2O, the Platform
  - Open Source (Apache 2.0)
  - Algorithms written in Java
    - Fast, distributed and scalable
  - Multiple interfaces to suit different users
    - Web, R, Python, Java, Scala, REST/JSON
  - Works with desktop/laptop, cloud, Spark and Hadoop



Scientific Advisory Council

Dr. Trevor Hastie

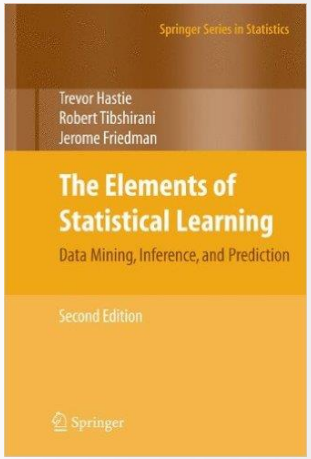
- John A. Overdeck Professor of Mathematics, Stanford University
- PhD in Statistics, Stanford University
- Co-author, *The Elements of Statistical Learning: Prediction, Inference and Data Mining*
- Co-author with John Chambers, *Statistical Models in S*
- Co-author, *Generalized Additive Models*

Dr. Robert Tibshirani

- Professor of Statistics and Health Research and Policy, Stanford University
- PhD in Statistics, Stanford University
- Co-author, *The Elements of Statistical Learning: Prediction, Inference and Data Mining*
- Author, *Regression Shrinkage and Selection via the Lasso*
- Co-author, *An Introduction to the Bootstrap*

Dr. Steven Boyd

- Professor of Electrical Engineering and Computer Science, Stanford University
- PhD in Electrical Engineering and Computer Science, UC Berkeley
- Co-author, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*
- Co-author, *Linear Matrix Inequalities in System and Control Theory*
- Co-author, *Convex Optimization*



Current Algorithm Overview

Statistical Analysis

- Linear Models (GLM)
- Naïve Bayes

Ensembles

- Random Forest
- Distributed Trees
- Gradient Boosting Machine
- R Package - Stacking / Super Learner

Deep Neural Networks

- Multi-layer Feed-Forward Neural Network
- Auto-encoder
- Anomaly Detection
- Deep Features

Clustering

- K-Means

Dimension Reduction

- Principal Component Analysis
- Generalized Low Rank Models

Solvers & Optimization

- Generalized ADMM Solver
- L-BFGS (Quasi Newton Method)
- Ordinary Least-Square Solver
- Stochastic Gradient Descent

Data Munging

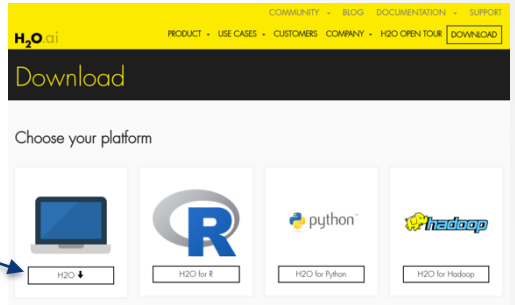
- Scalable Data Frames
- Sort, Slice, Log Transform

Joe's Strata Hadoop London Talk
bit.ly/joe_h2o_talk4

Joe's LondonR Talk
bit.ly/joe_h2o_talk3

H2O Flow (Web Interface) Demo

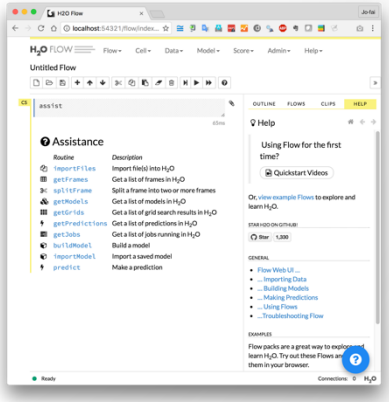
- Download and unzip **jar** from www.h2o.ai



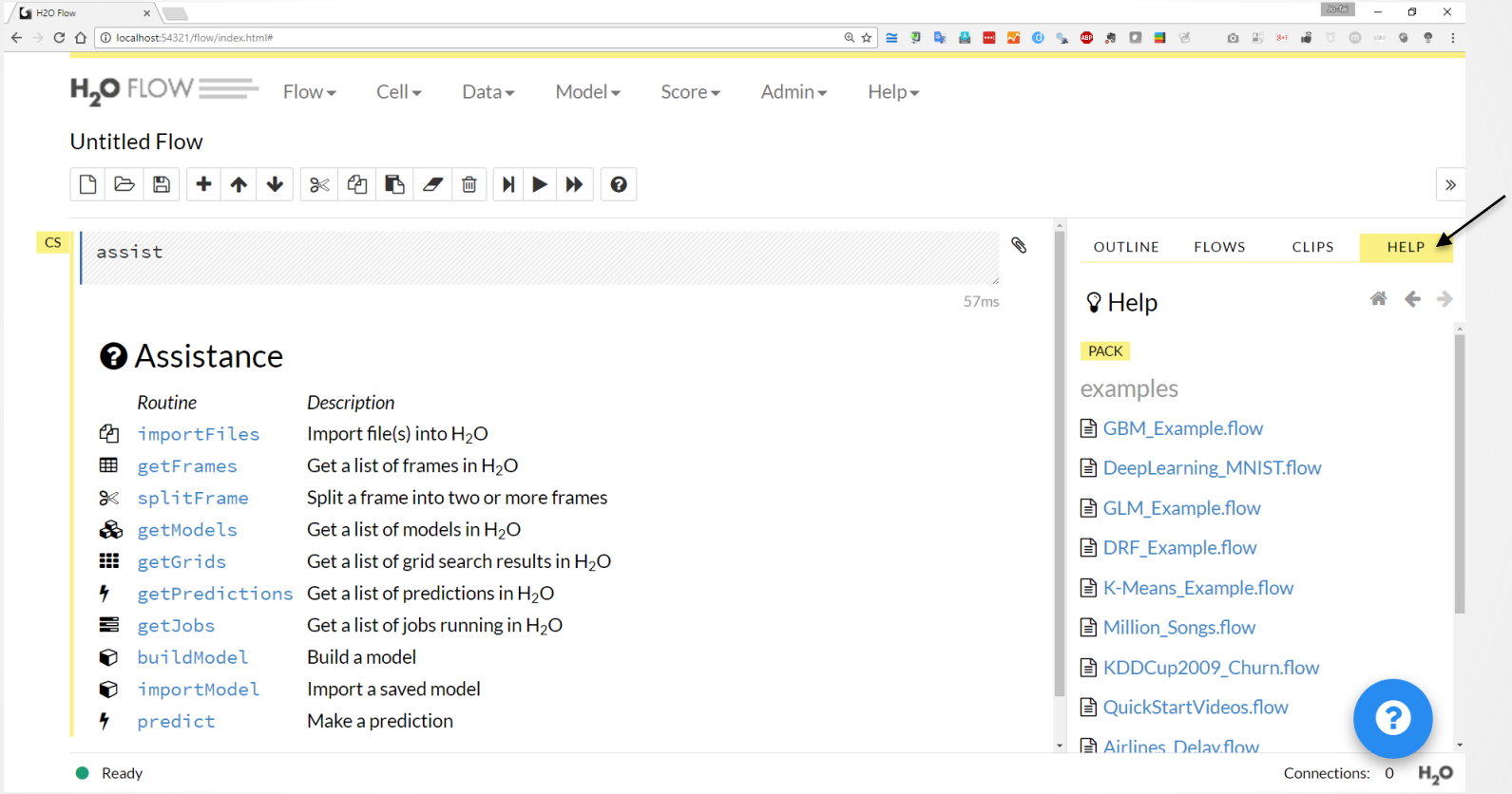
- In terminal:
  - java -jar h2o.jar

```
h2o-3.10.0.6 — java -jar h2o.jar — 123×24
Jo-fais-MacBook-Pro-2:~ jofaichow$ cd h2o-3.10.0.6
Jo-fais-MacBook-Pro-2:h2o-3.10.0.6 jofaichow$ java -jar h2o.jar
09-18 13:16:13.620 192.168.0.6:54321     8620   main      INFO: ----- H2O started  -----
09-18 13:16:13.636 192.168.0.6:54321     8620   main      INFO: Build git branch: rel-turing
09-18 13:16:13.636 192.168.0.6:54321     8620   main      INFO: Build git hash: 3b286dea7b719b6ef2c2f5f7728648f2440a1502
09-18 13:16:13.636 192.168.0.6:54321     8620   main      INFO: Build git describe: jenkins-rel-turing-6
09-18 13:16:13.636 192.168.0.6:54321     8620   main      INFO: Build project version: 3.10.0.6 (latest version: 3.10.0.6)
```

- Web browser:
  - localhost:54321



H2O Flow Examples



Other H2O Interfaces

- R

```r
# Load H2O R package
library(h2o)

# Initialize and Connect to H2O
h2o.init()
```

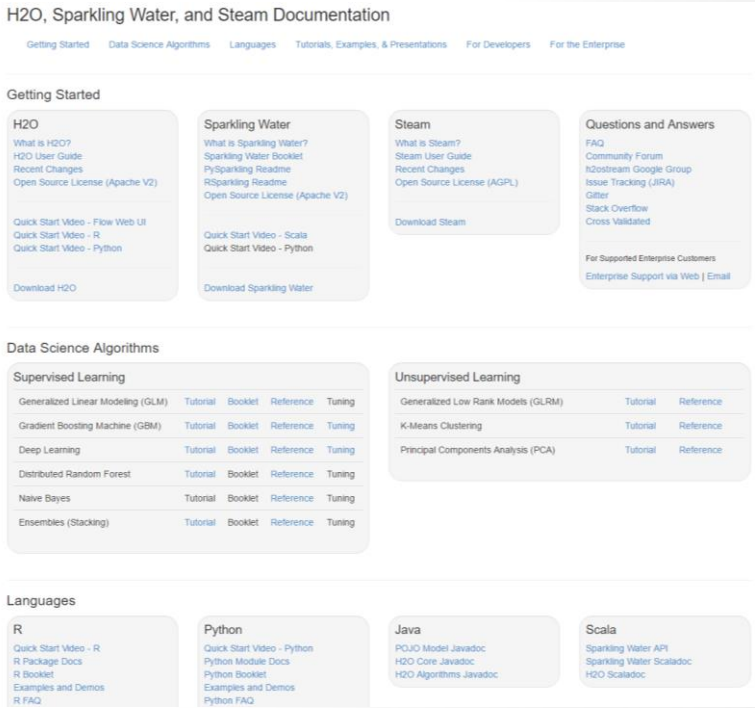
- Python

```python
# Import H2O Python module
import h2o

# Initialize and Connect to H2O
h2o.init()
```

Key Resources

- docs.h2o.ai



Why H2O?

H2O for Kaggle Competitions

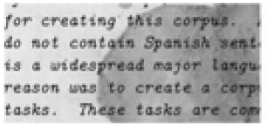
CIFAR-10 Competition Winners: Interviews with Dr. Ben Graham, Phil Culliton, & Zygmunt Zając

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



Completed • Knowledge • 161 teams

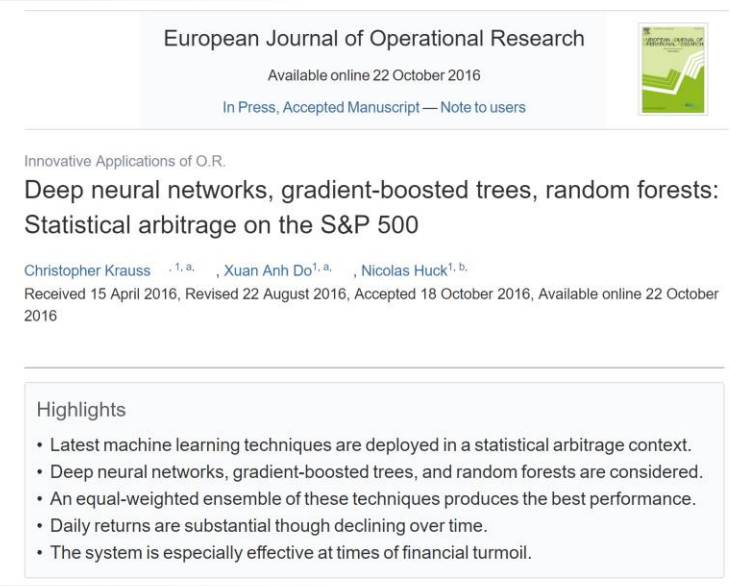
Denoising Dirty Documents

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 for Academic Research



European Journal of Operational Research

Available online 22 October 2016

In Press, Accepted Manuscript — Note to users

Innovative Applications of O.R.

Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500

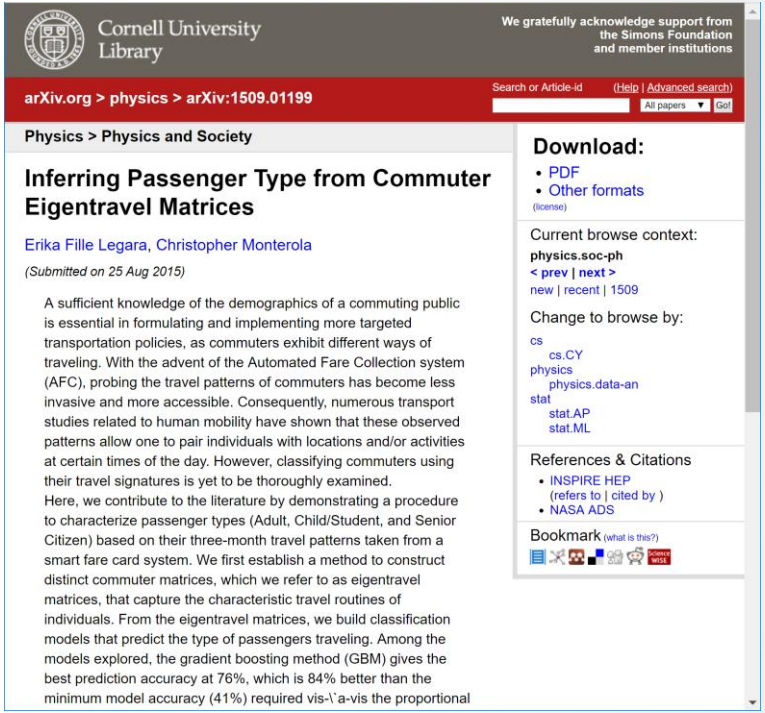
Christopher Krauss [1, a], Xuan Anh Do [1, a], Nicolas Huck [1, b]

Received 15 April 2016, Revised 22 August 2016, Accepted 18 October 2016, Available online 22 October 2016

Highlights

- Latest machine learning techniques are deployed in a statistical arbitrage context.
- Deep neural networks, gradient-boosted trees, and random forests are considered.
- An equal-weighted ensemble of these techniques produces the best performance.
- Daily returns are substantial though declining over time.
- The system is especially effective at times of financial turmoil.

http://www.sciencedirect.com/science/article/pii/S0377221716308657



Cornell University Library

We gratefully acknowledge support from the Simons Foundation and member institutions

arXiv.org > physics > arXiv:1509.01199

Search or Article-id (Help | Advanced search) All papers Go!

Physics > Physics and Society

Inferring Passenger Type from Commuter Eigentravel Matrices

Erika Fille Legara, Christopher Monterola

(Submitted on 25 Aug 2015)

A sufficient knowledge of the demographics of a commuting public is essential in formulating and implementing more targeted transportation policies, as commuters exhibit different ways of traveling. With the advent of the Automated Fare Collection system (AFC), probing the travel patterns of commuters has become less invasive and more accessible. Consequently, numerous transport studies related to human mobility have shown that these observed patterns allow one to pair individuals with locations and/or activities at certain times of the day. However, classifying commuters using their travel signatures is yet to be thoroughly examined. Here, we contribute to the literature by demonstrating a procedure to characterize passenger types (Adult, Child/Student, and Senior Citizen) based on their three-month travel patterns taken from a smart fare card system. We first establish a method to construct distinct commuter matrices, which we refer to as eigentravel matrices, that capture the characteristic travel routines of individuals. From the eigentravel matrices, we build classification models that predict the type of passengers traveling. Among the models explored, the gradient boosting method (GBM) gives the best prediction accuracy at 76%, which is 84% better than the minimum model accuracy (41%) required vis-\`a-vis the proportional

Download:

- PDF
- Other formats

(license)

Current browse context: physics.soc-ph

< prev | next >

new | recent | 1509

Change to browse by:

cs

cs.CY

physics

physics.data-an

stat

stat.AP

stat.ML

References & Citations

- INSPIRE HEP (refers to | cited by)
- NASA ADS

Bookmark (what is this?)

https://arxiv.org/abs/1509.01199

H2O In Action

www.h2o.ai/customers

Capital One

Capital One uses H2O open source machine learning for various use cases.

MarketShare

H2O predictive analytics helps boost the impact and results of digital marketing.

Kaiser

Kaiser uses H2O machine learning to save lives.

Zurich Insurance

Zurich turned to H2O as a strategic differentiator for commercial insurance.

Progressive

Progressive uses H2O predictive analytics for user-based insurance.

Comcast

Comcast uses H2O to improve customer experience.

Hospital Corporation of America

HCA uses H2O to predict patient outcomes in real-time.

McKesson

McKesson discusses the adoption of artificial intelligence in healthcare.

Macy's

Macy's uses H2O for personalized site recommendations.

Transamerica

Transamerica turns to H2O to develop a product recommendation platform for insurance.

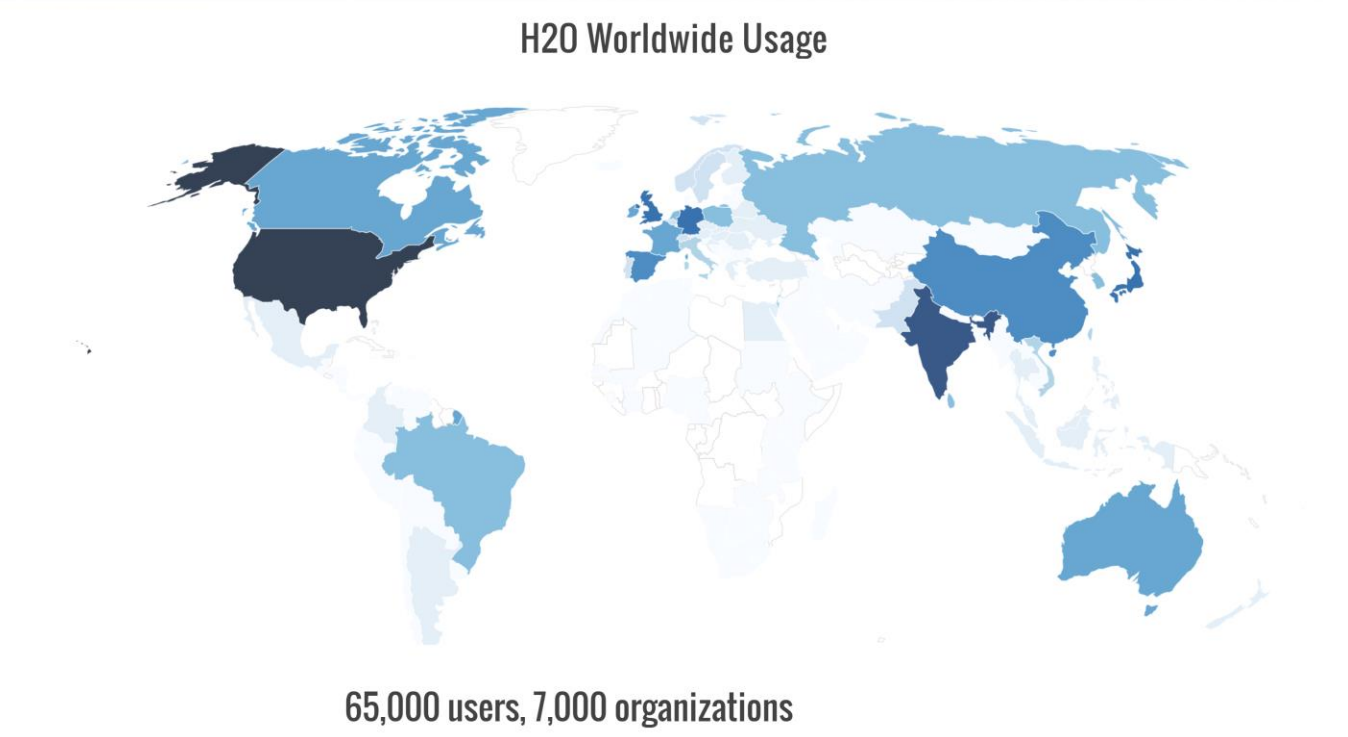
Paypal

Paypal turned to H2O Deep Learning for fraud detection and customer churn.

eBay

eBay chose H2O for open source machine learning.

H2O Worldwide Usage



www.h2o.ai/community

Deep Learning in H2O

A Simple Neural Network

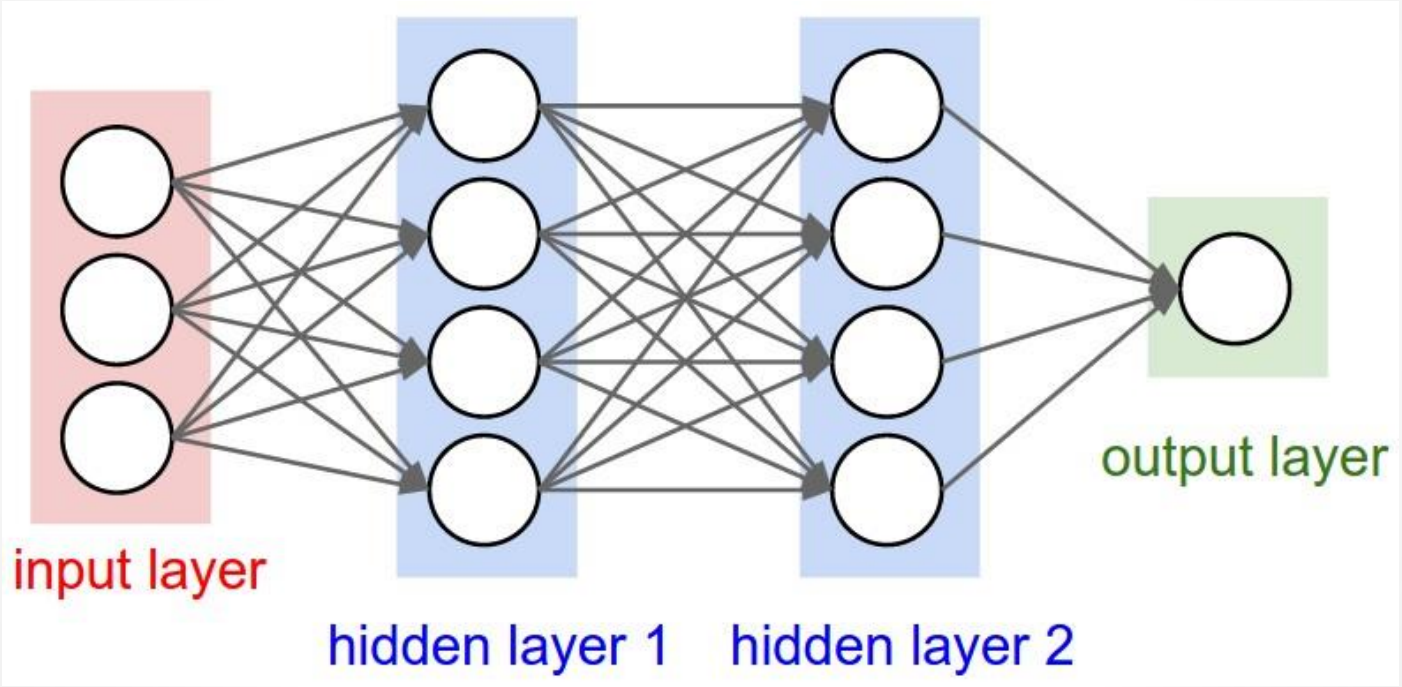


Image credit: http://cs231n.github.io/

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

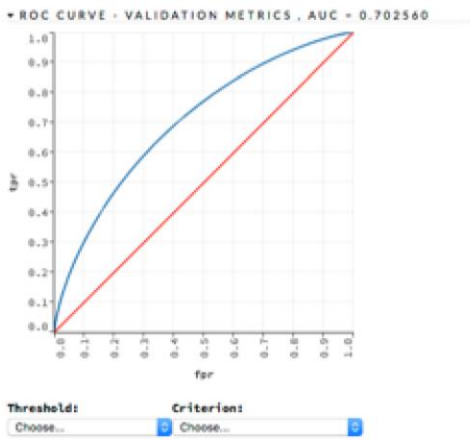
Iterations: 12. Epochs: 0.628821. Speed: 2,243,735 samples/sec. Estimated time left: 21.849 sec

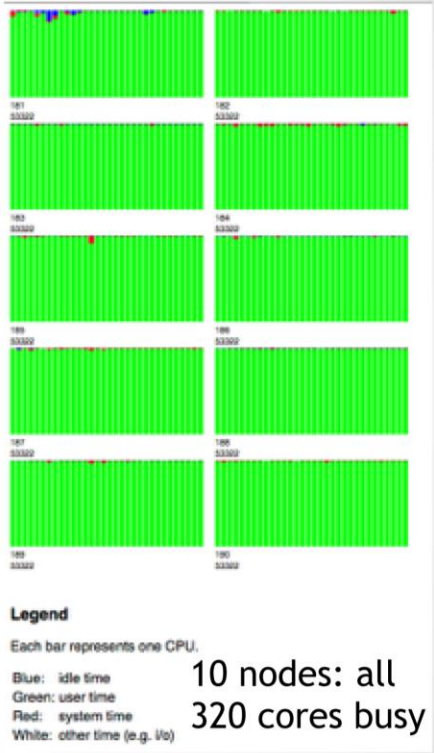
Actions: View | Cancel Job

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

▾ OUTPUT - STATUS OF NEURON LAYERS (PREDICTING ISDEPDELAYED, 2-CLASS CLASSIFICATION, 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	0.0483	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	-1.3908	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	0.0161	0.0083	0	0.0741	0.7260	0.4269	0.2056





10 nodes: all
320 cores busy

real-time, interactive
model inspection in Flow

H2O.ai Deep Learning Model

H2O for Kaggle Competitions

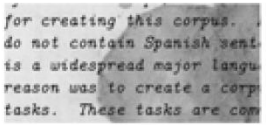
CIFAR-10 Competition Winners: Interviews with Dr. Ben Graham, Phil Culliton, & Zygmunt Zając

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



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."

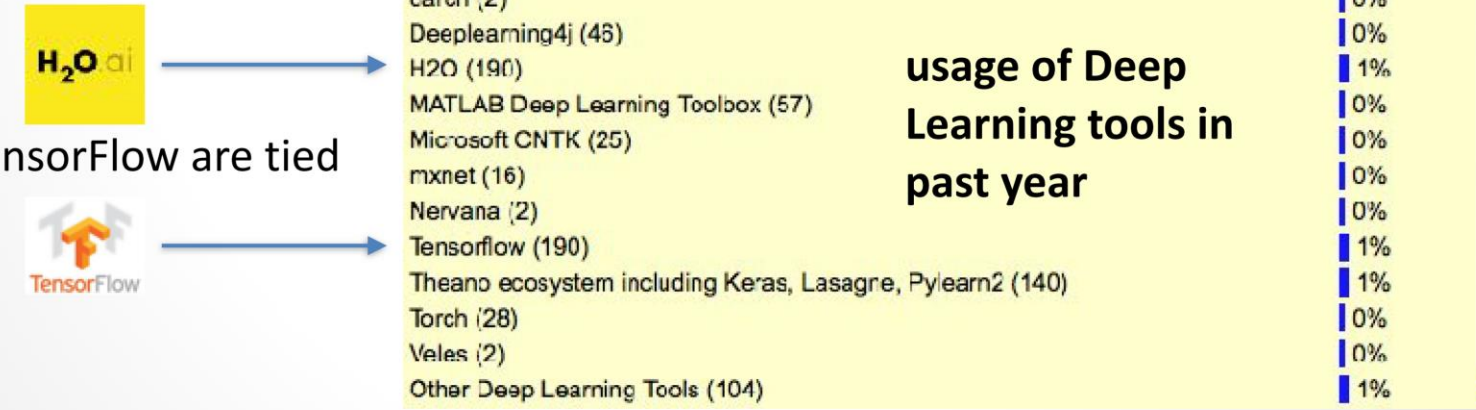
H2O Deep Learning Is Widely Used

The usage of Hadoop/Big Data tools grew to 39%, up from 29% in 2015 (and 17% in 2014), driven by Apache Spark, MLlib (Spark Machine Learning Library) and H2O.
See also

- KDnuggets interview with Spark Creator Matei Zaharia
- KDnuggets interview with Arno Candel, H2O.ai on How to Quick Start Deep Learning with H2O

http://www.kdnuggets.com

H2O and TensorFlow are tied

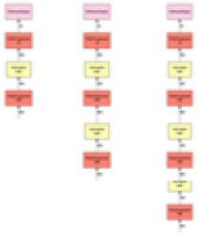


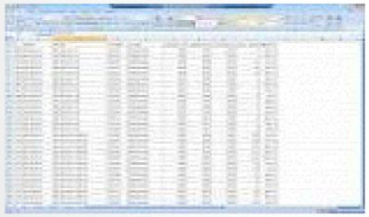
Why Deep Water?

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



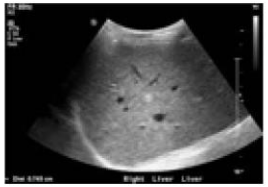


AMD
intel

Limited to business analytics,
statistical models (CSV data)

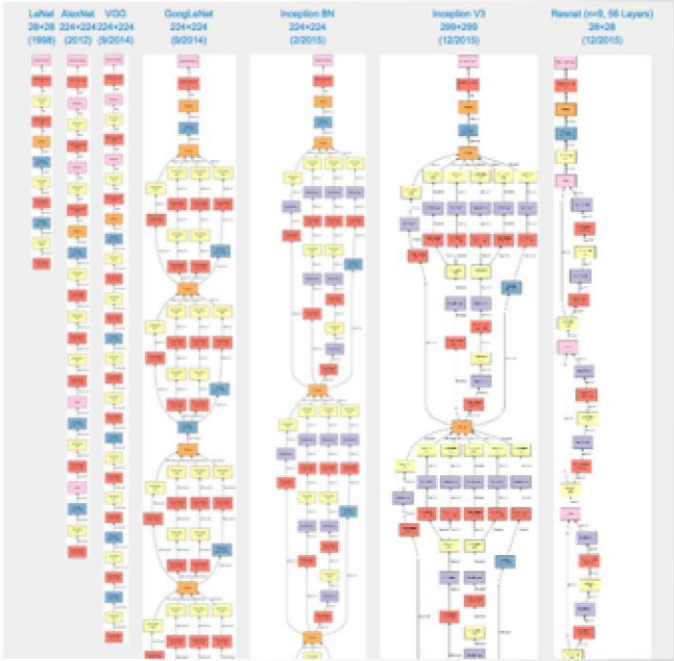
Deep Water: deep complex networks

5-1000 layers
GBs/TBs of data



AMD
intel

NVIDIA
CUDA



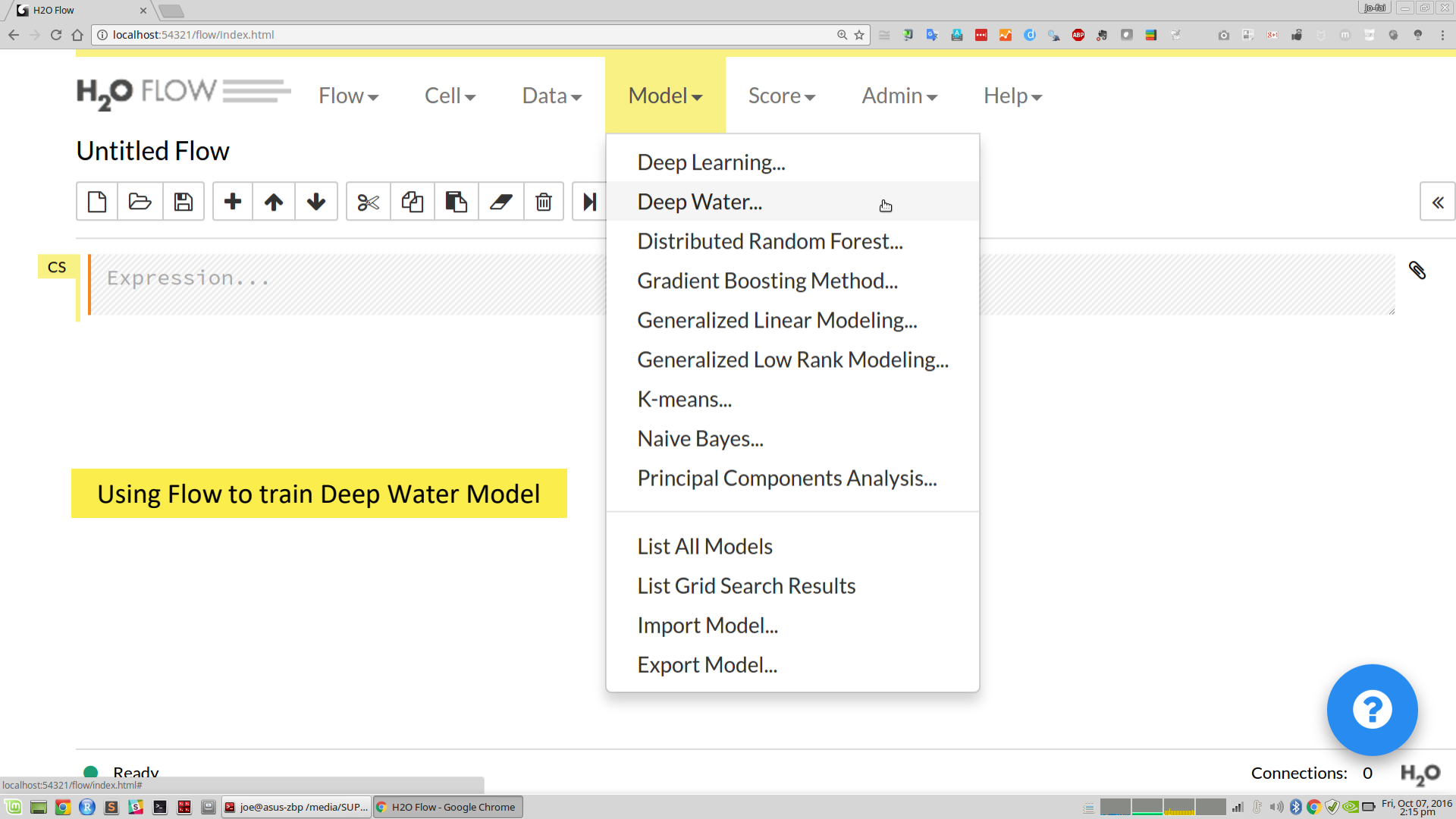
Large networks for big data
(e.g. image 1000x1000x3 -> 3m inputs)

H2O.ai

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 Flow
localhost:54321/flow/index.html
H2O FLOW
Flow
Cell
Data
Model
Score
Admin
Help
Untitled Flow
CS
Expression...
Deep Learning...
Deep Water...
Distributed Random Forest...
Gradient Boosting Method...
Generalized Linear Modeling...
Generalized Low Rank Modeling...
K-means...
Naive Bayes...
Principal Components Analysis...
List All Models
List Grid Search Results
Import Model...
Export Model...
Using Flow to train Deep Water Model
Ready
Connections: 0
H2O
localhost:54321/flow/index.html#
joe@asus-zbp /media/SUP...
H2O Flow - Google Chrome
Fri, Oct 07, 2016
2:15 pm

Roadmap for Deep Water (Q4 2016):

Finish TensorFlow integration (C++/Python/Java):
Package Python on the backend to create trainable graphs

Finish Caffe integration (pure C++/Java):
Optimized Multi-GPU training (NVIDIA NCCL)

Add multi-GPU support for mxnet

Add more capabilities to H2O Deep Water:
Text/NLP, Time Series, LSTM, AutoEncoder,
Feature Extraction, Input/Output shape mapping, etc.

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
  - github.com/h2oai/deepwater

Code and References

Python/R Jupyter Notebooks

Check out a sample of cool Deep Learning Jupyter notebooks!

PreRelease Downloads

For the following system dependencies, we provide recent builds for your convenience.

- Ubuntu 16.04 LTS
- Latest NVIDIA Display driver
- CUDA 8 (latest available) in /usr/local/cuda
- CUDNN 5 (inside of lib and include directories in /usr/local/cuda/)

In the future, we'll have more pre-built jars for more OS/CUDA combinations.

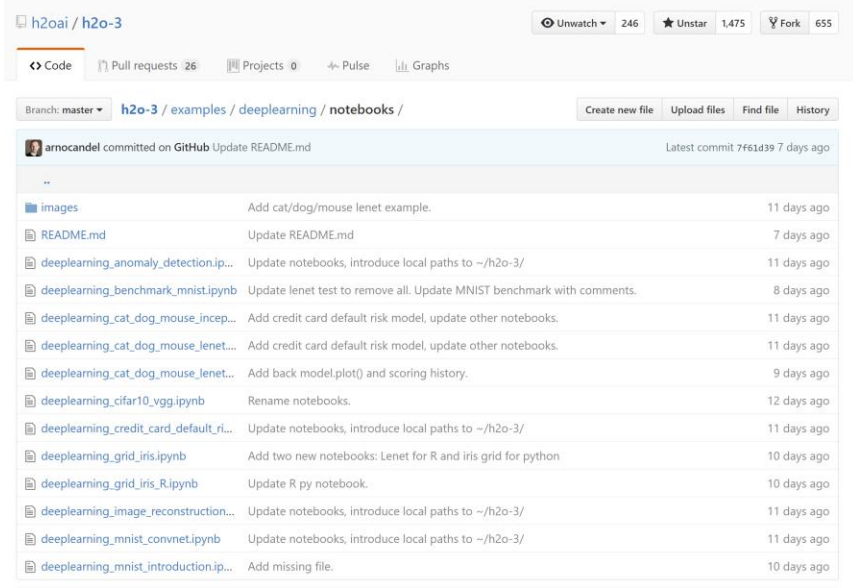
- Required to run Jupyter notebook: H2O Deep Water enabled Python module -- install via `pip install <file>`
- To build custom networks: Matching MXNet Python egg -- install via `easy_install <file>`
- To run from Flow only: H2O Standalone h2o.jar -- launch via `java -jar h2o.jar`

If you are interested in running H2O Deep Water on a different infrastructure, see the DIY build instructions below

PreRelease Amazon AWS Image

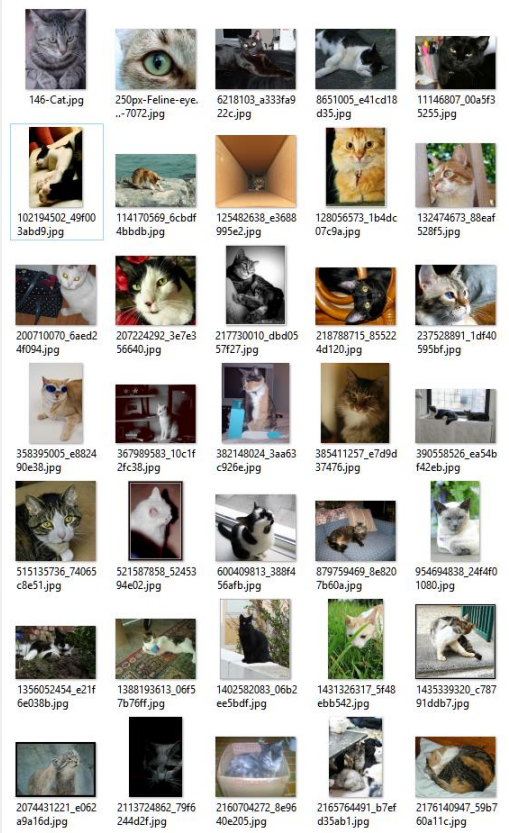
For your convenience, here's a pre-built image for Amazon's EC2 environment, based off our recent H2O Open Tour Hands-On Deep Water workshop (recording coming soon).

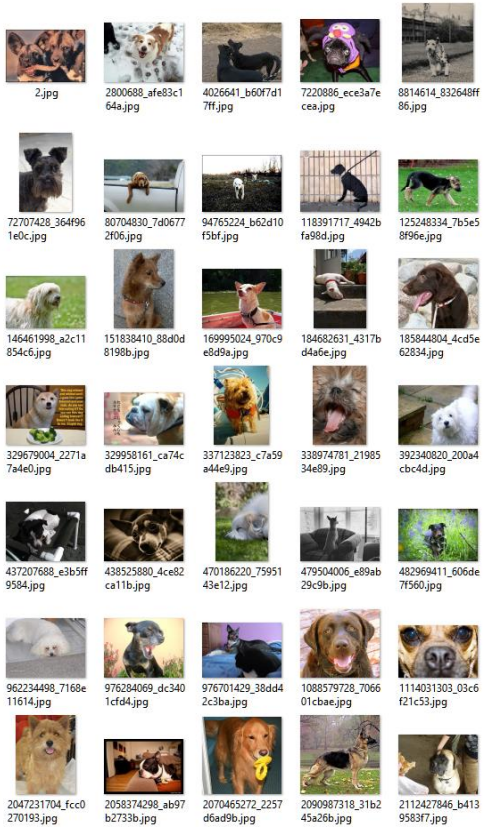
- AMI ID: ami-d32f70c4
- AMI Name: deepwater-dallas-v3
- Recommended instance types: g2.2xlarge or p2.xlarge
- After launching the instance, you can connect to port 8888 (Jupyter Notebook) or port 54321 (H2O Flow).

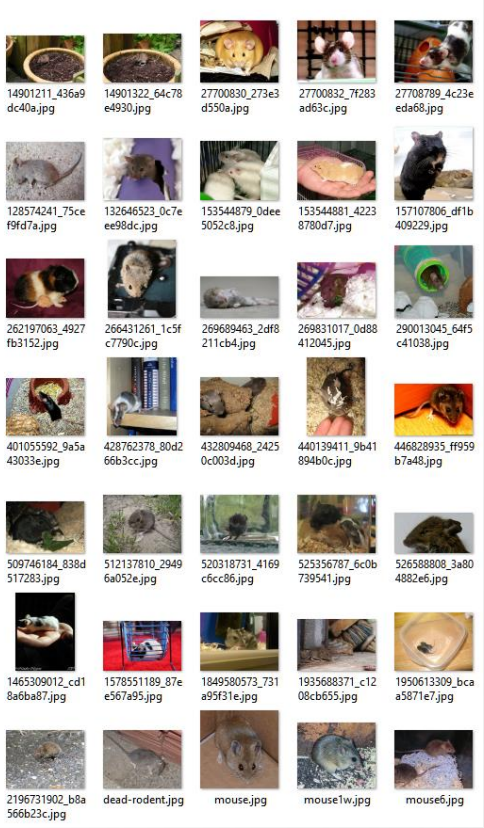


github.com/h2oai/deepwater

Data – Cat/Dog/Mouse Images

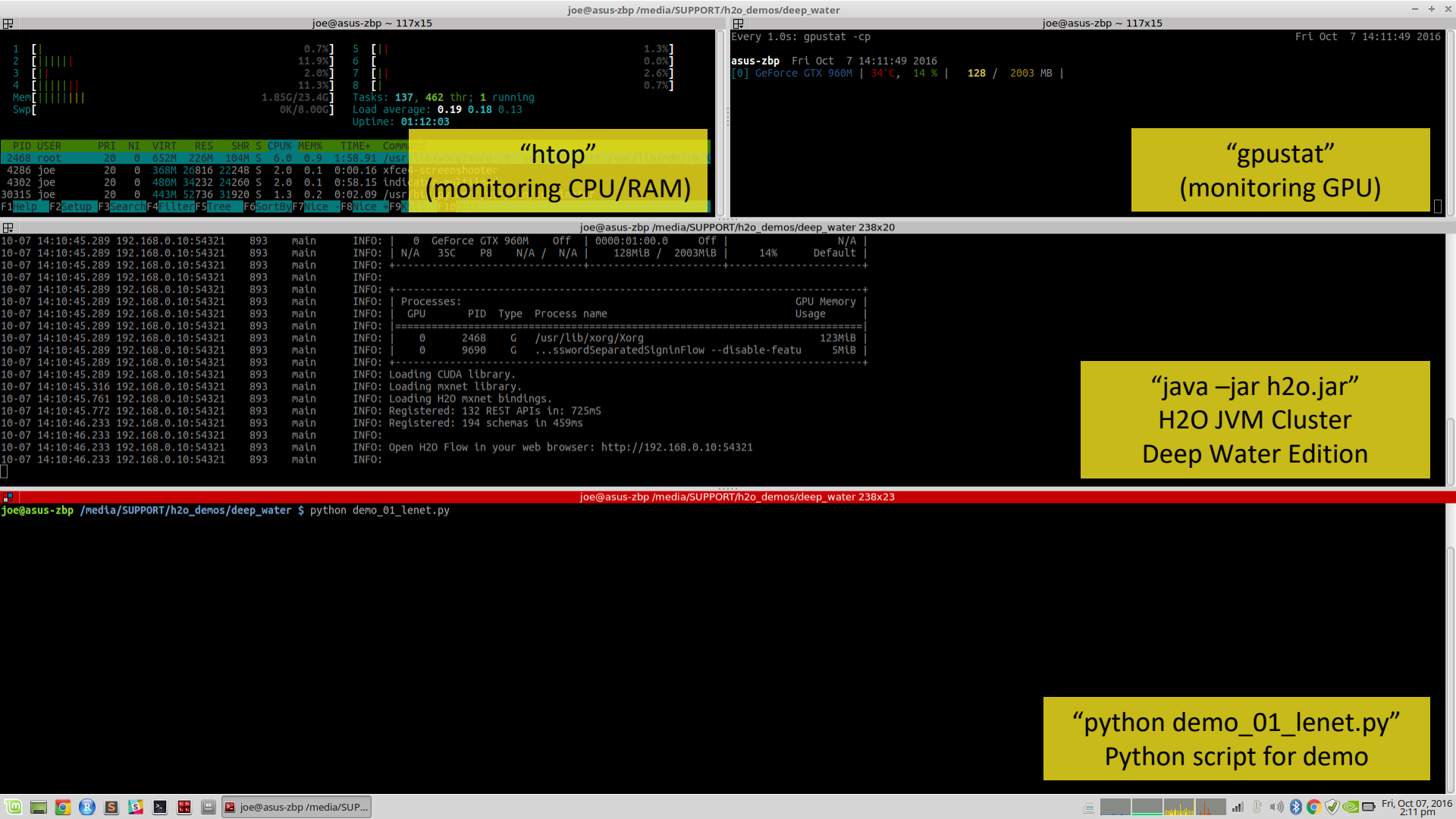






Data - CSV

	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


joe@asus-zbp /media/SUPPORT/h2o_demos/deep_water
joe@asus-zbp ~ 117x15
1 [| 0.7%]
2 [|||||| 11.9%]
3 [|| 2.0%]
4 [||||||| 11.3%]
5 [|| 1.3%]
6 [0.0%]
7 [|| 2.6%]
8 [0.7%]
Mem[|||||||| 1.85G/23.4G]
Swp[0K/8.00G]
Tasks: 137, 462 thr; 1 running
Load average: 0.19 0.18 0.13
Uptime: 01:12:03
PID USER PRI NI VIRT RES SHR S CPU% MEM% TIME+ Command
2468 root 20 0 652M 226M 104M S 6.0 0.9 1:58.91 /usr/lib/xorg/Xorg :0 -seat seat0 -auth /var/lib/gdm/:0.
4286 joe 20 0 368M 26816 22248 S 2.0 0.1 0:00.16 xfce4-screenshooter
4302 joe 20 0 480M 34232 24260 S 2.0 0.1 0:58.15 indicator-multiload
30315 joe 20 0 443M 52736 31920 S 1.3 0.2 0:02.09 /usr/bin/python3 /usr/bin/terminator
F1Help F2Setup F3SearchF4FilterF5Tree F6SortByF7Nice -F8Nice +F9Kill F10Quit
"htop"
(monitoring CPU/RAM)
joe@asus-zbp ~ 117x15
Every 1.0s: gpustat -cp
Fri Oct 7 14:11:49 2016
asus-zbp Fri Oct 7 14:11:49 2016
[0] GeForce GTX 960M | 34'C, 14 % | 128 / 2003 MB |
"gpustat"
(monitoring GPU)
joe@asus-zbp /media/SUPPORT/h2o_demos/deep_water 238x20
10-07 14:10:45.289 192.168.0.10:54321 893 main INFO: | 0 GeForce GTX 960M Off | 0000:01:00.0 Off | N/A |
10-07 14:10:45.289 192.168.0.10:54321 893 main INFO: | N/A 35C P8 N/A / N/A | 128MiB / 2003MiB | 14% Default |
10-07 14:10:45.289 192.168.0.10:54321 893 main INFO: | Processes: GPU Memory |
10-07 14:10:45.289 192.168.0.10:54321 893 main INFO: | GPU PID Type Process name Usage |
10-07 14:10:45.289 192.168.0.10:54321 893 main INFO: | 0 2468 G /usr/lib/xorg/Xorg 123MiB |
10-07 14:10:45.289 192.168.0.10:54321 893 main INFO: | 0 9690 G ...sswordSeparatedSigninFlow --disable-featu 5MiB |
10-07 14:10:45.289 192.168.0.10:54321 893 main INFO: Loading CUDA library.
10-07 14:10:45.316 192.168.0.10:54321 893 main INFO: Loading mxnet library.
10-07 14:10:45.761 192.168.0.10:54321 893 main INFO: Loading H2O mxnet bindings.
10-07 14:10:45.772 192.168.0.10:54321 893 main INFO: Registered: 132 REST APIs in: 725mS
10-07 14:10:46.233 192.168.0.10:54321 893 main INFO: Registered: 194 schemas in 459mS
10-07 14:10:46.233 192.168.0.10:54321 893 main INFO:
10-07 14:10:46.233 192.168.0.10:54321 893 main INFO: Open H2O Flow in your web browser: http://192.168.0.10:54321
10-07 14:10:46.233 192.168.0.10:54321 893 main INFO:
"java –jar h2o.jar"
H2O JVM Cluster
Deep Water Edition
joe@asus-zbp /media/SUPPORT/h2o_demos/deep_water 238x23
joe@asus-zbp /media/SUPPORT/h2o_demos/deep_water $ python demo_01_lenet.py
"python demo_01_lenet.py"
Python script for demo
joe@asus-zbp /media/SUP...
Fri, Oct 07, 2016
2:11 pm

/media/SUPPORT/h2o_demos/deep_water/demo_01_lenet.py - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences Help

demo_01_lenet.py

```
1  from __future__ import print_function
2  import sys, os
3  sys.path.insert(1, os.path.join("..","..","..")) 
4
5  # H2O
6  import h2o
7  from h2o.estimators.deepwater import H2ODeepWaterEstimator
8
9  # Start and connect to H2O local cluster
10 h2o.init()
11
12 # Import CSV
13 frame = h2o.import_file("bigdata/laptop/deepwater/imagenet/cat_dog_mouse.csv")
14 print(frame.head(5))
15
16 # Define LeNet model
17 model = H2ODeepWaterEstimator(epochs=300, rate=1e-3, network='lenet', score_interval=0, train_samples_per_iteration=1000)
18
19 # Train LeNet model on GPU
20 model.train(x=[0],y=1, training_frame=frame)
21 model.show()
22
```

H2O's Python Module

Deep Water module

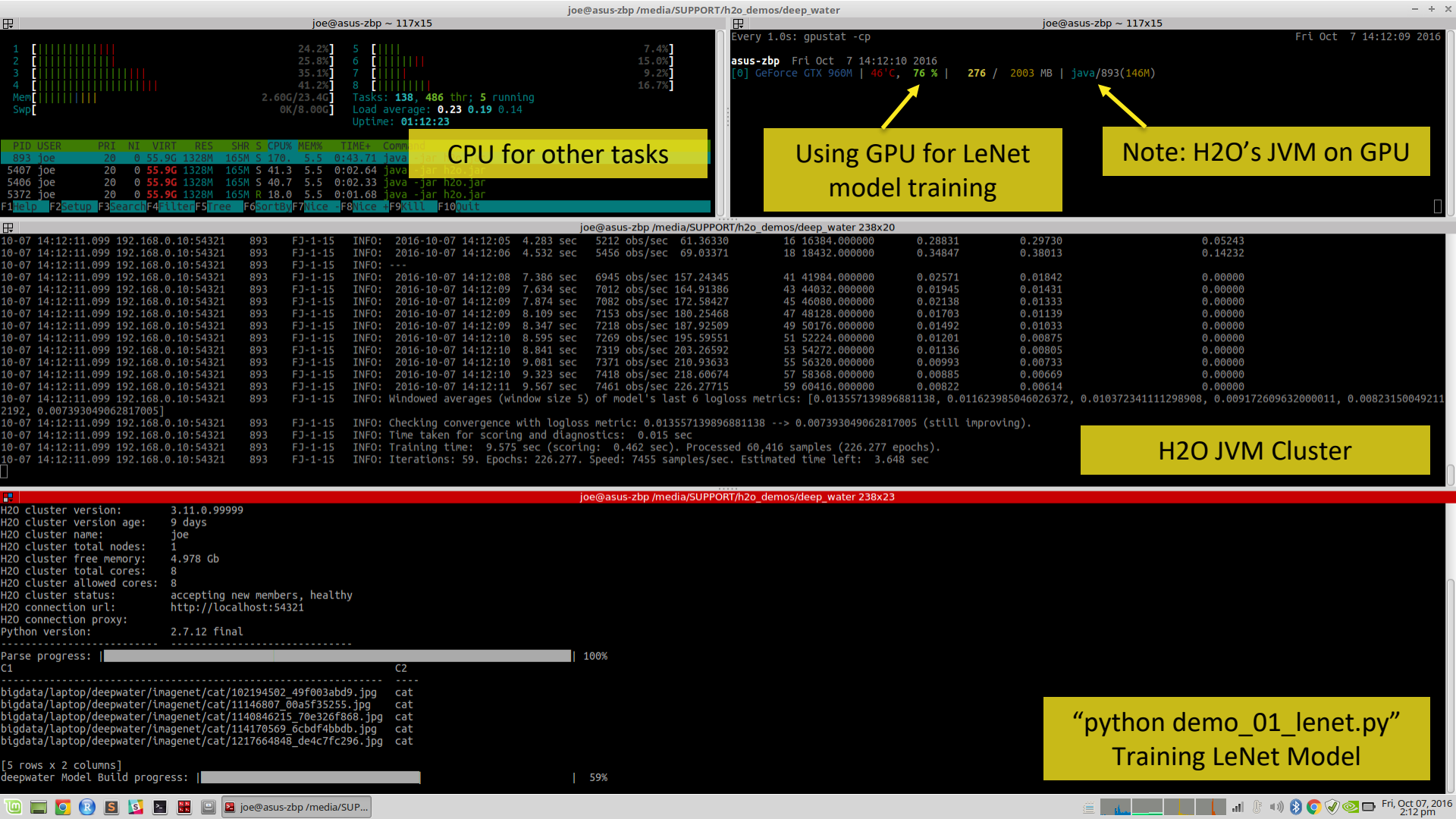
Connect to H2O Cluster

Import CSV

Define LeNet model in Deep Water

Train and show model

Line 1, Column 38 Tab Size: 4 Python


joe@asus-zbp /media/SUPPORT/h2o_demos/deep_water
joe@asus-zbp ~ 117x15
Tasks: 138, 486 thr; 5 running
Load average: 0.23 0.19 0.14
Uptime: 01:12:23
CPU for other tasks
joe@asus-zbp ~ 117x15
Every 1.0s: gpustat -cp
Fri Oct 7 14:12:09 2016
asus-zbp Fri Oct 7 14:12:10 2016
[0] GeForce GTX 960M | 46'C, 76 % | 276 / 2003 MB | java/893(146M)
Using GPU for LeNet model training
Note: H2O's JVM on GPU
joe@asus-zbp /media/SUPPORT/h2o_demos/deep_water 238x20
H2O JVM Cluster
joe@asus-zbp /media/SUPPORT/h2o_demos/deep_water 238x23
H2O cluster version: 3.11.0.99999
H2O cluster version age: 9 days
H2O cluster name: joe
H2O cluster total nodes: 1
H2O cluster free memory: 4.978 Gb
H2O cluster total cores: 8
H2O cluster allowed cores: 8
H2O cluster status: accepting new members, healthy
H2O connection url: http://localhost:54321
H2O connection proxy:
Python version: 2.7.12 final
Parse progress: 100%
C1 C2
bigdata/laptop/deepwater/imagenet/cat/102194502_49f003abd9.jpg cat
bigdata/laptop/deepwater/imagenet/cat/11146807_00a5f35255.jpg cat
bigdata/laptop/deepwater/imagenet/cat/1140846215_70e326f868.jpg cat
bigdata/laptop/deepwater/imagenet/cat/114170569_6cbdf4bbdb.jpg cat
bigdata/laptop/deepwater/imagenet/cat/1217664848_de4c7fc296.jpg cat
[5 rows x 2 columns]
deepwater Model Build progress: 59%
"python demo_01_lenet.py" Training LeNet Model

joe@asus-zbp /media/SUPPORT/h2o_demos/deep_water

joe@asus-zbp ~ 117x15

```
  1  [||                                        2.6%]   5  [                                        0.0%]
  2  [||                                        2.0%]   6  [|                                       0.7%]
  3  [||||||                                    9.2%]   7  [|                                       1.3%]
  4  [||                                        2.7%]   8  [|                                       0.7%]
  Mem[||||||||||                         2.73G/23.4G]   Tasks: 137, 478 thr; 1 running
  Swp[                                     0K/8.00G]   Load average: 0.33 0.22 0.15
                                                       Uptime: 01:12:44

  PID USER      PRI  NI  VIRT   RES   SHR S CPU% MEM%   TIME+  Command
 2468 root       20   0  652M  226M  104M S  3.3  0.9  2:01.04 /usr/lib/xorg/Xorg :0 -audit 0 -auth /var/lib/mdm/:0.X
 4491 joe        20   0  475M 39460 26412 S  2.0  0.2  1:06.28 /usr/lib/x86_64-linux-gnu/xfce4/panel/wrapper-2.0 /usr
 4302 joe        20   0  480M 34232 24260 S  2.0  0.1  0:58.59 indicator-multiload
 6576 joe        20   0  368M 26776 22208 S  1.3  0.1  0:00.16 xfce4-screenshooter
F1Help  F2Setup F3SearchF4FilterF5Tree  F6SortByF7Nice -F8Nice +F9Kill  F10Quit
```

joe@asus-zbp ~ 117x15

```
Every 1.0s: gpustat -cp                                                    Fri Oct  7 14:12:30 2016

asus-zbp  Fri Oct  7 14:12:30 2016
[0] GeForce GTX 960M | 42'C,   4 % |   276 /  2003 MB | java/893(146M)
```

joe@asus-zbp /media/SUPPORT/h2o_demos/deep_water 238x12

```
10-07 14:12:13.564 192.168.0.10:54321    893    FJ-1-15   INFO:  2016-10-07 14:12:11 10.428 sec    7592 obs/sec 253.12360       66 67584.000000     0.00672      0.00506      0.00000
10-07 14:12:13.564 192.168.0.10:54321    893    FJ-1-15   INFO:  2016-10-07 14:12:12 10.676 sec    7623 obs/sec 260.79401       68 69632.000000     0.00621      0.00476      0.00000
10-07 14:12:13.564 192.168.0.10:54321    893    FJ-1-15   INFO:  2016-10-07 14:12:12 10.921 sec    7653 obs/sec 268.46442       70 71680.000000     0.00604      0.00460      0.00000
10-07 14:12:13.564 192.168.0.10:54321    893    FJ-1-15   INFO:  2016-10-07 14:12:12 11.168 sec    7680 obs/sec 276.13483       72 73728.000000     0.00591      0.00446      0.00000
10-07 14:12:13.564 192.168.0.10:54321    893    FJ-1-15   INFO:  2016-10-07 14:12:12 11.311 sec    7684 obs/sec 279.97004       73 74752.000000     0.00580      0.00447      0.00000
10-07 14:12:13.564 192.168.0.10:54321    893    FJ-1-15   INFO:  2016-10-07 14:12:13 11.545 sec    7719 obs/sec 287.64045       75 76800.000000     0.00528      0.00410      0.00000
10-07 14:12:13.564 192.168.0.10:54321    893    FJ-1-15   INFO:  2016-10-07 14:12:13 11.777 sec    7754 obs/sec 295.31086       77 78848.000000     0.00516      0.00396      0.00000
10-07 14:12:13.564 192.168.0.10:54321    893    FJ-1-15   INFO:  2016-10-07 14:12:13 12.021 sec    7778 obs/sec 302.98127       79 80896.000000     0.00495      0.00381      0.00000
10-07 14:12:13.564 192.168.0.10:54321    893    FJ-1-15   INFO:  ====================================================================================================
10-07 14:12:14.102 192.168.0.10:54321    893    #51044-21 INFO: GET /3/Models/DeepWater_model_python_1475845836843_1, parms: {}
10-07 14:12:14.139 192.168.0.10:54321    893    #51044-16 INFO: DELETE /4/sessions/_sid_8b27, parms: {}
```

joe@asus-zbp /media/SUPPORT/h2o_demos/deep_water 238x31

```
cat    dog    mouse    Error    Rate
-----  -----  -------  -------  -------
90     0      0        0        0 / 90
0      85     0        0        0 / 85
0      0      92       0        0 / 92
90     85     92       0        0 / 267
Top-3 Hit Ratios:
k    hit_ratio
---  -----------
1    1
2    1
3    1
Scoring History:
     timestamp            duration    training_speed    epochs         iterations    samples    training_rmse      training_logloss    training_classification_error
---  -------------------  ----------  ----------------  -------------  ------------  ---------  -----------------  ------------------  -------------------------------
     2016-10-07 14:12:02  0.000 sec                     0.0            0             0.0        nan                nan                 nan
     2016-10-07 14:12:03  2.468 sec   715 obs/sec       3.83520599251  1             1024.0     0.625444676946     1.01735967452       0.535580524345
     2016-10-07 14:12:04  2.839 sec   2297 obs/sec      15.34082397    4             4096.0     0.547358282125     0.809359616147      0.348314606742
     2016-10-07 14:12:04  3.086 sec   3047 obs/sec      23.0112359551  6             6144.0     0.515823924554     0.738246051464      0.284644194757
     2016-10-07 14:12:04  3.322 sec   3660 obs/sec      30.6816479401  8             8192.0     0.453695218064     0.600937775265      0.228464419476
---  ---                  ---         ---               ---            ---           ---        ---                ---                 ---
     2016-10-07 14:12:12  11.168 sec  7680 obs/sec      276.134831461  72            73728.0    0.00591118567047   0.00446146351249    0.0
     2016-10-07 14:12:12  11.311 sec  7684 obs/sec      279.970037453  73            74752.0    0.00579576066578   0.00447401006606    0.0
     2016-10-07 14:12:13  11.545 sec  7719 obs/sec      287.640449438  75            76800.0    0.00528307643351   0.00409703365913    0.0
     2016-10-07 14:12:13  11.777 sec  7754 obs/sec      295.310861423  77            78848.0    0.005160950854     0.00395743083299    0.0
     2016-10-07 14:12:13  12.021 sec  7778 obs/sec      302.981273408  79            80896.0    0.00495305059942   0.00381117058655    0.0

See the whole table with table.as_data_frame()
H2O session _sid_8b27 closed.
joe@asus-zbp /media/SUPPORT/h2o_demos/deep_water $
```

"python demo_01_lenet.py"
Showing LeNet Model

joe@asus-zbp /media/SUP...

Fri, Oct 07, 2016
2:12 pm

Using Flow for Deep Water

H2O's Mission

Making Machine Learning Accessible to Everyone

Photo credit: Virgin Media

Current Contributors (more H2O.ai folks joining soon)

Fabrizio Milo

Cyprien Noel

Qiang Kou

Arno Candel

TensorFlow

Caffe

mxnet

H2O.ai

This repository Search

h2oai / deepwater

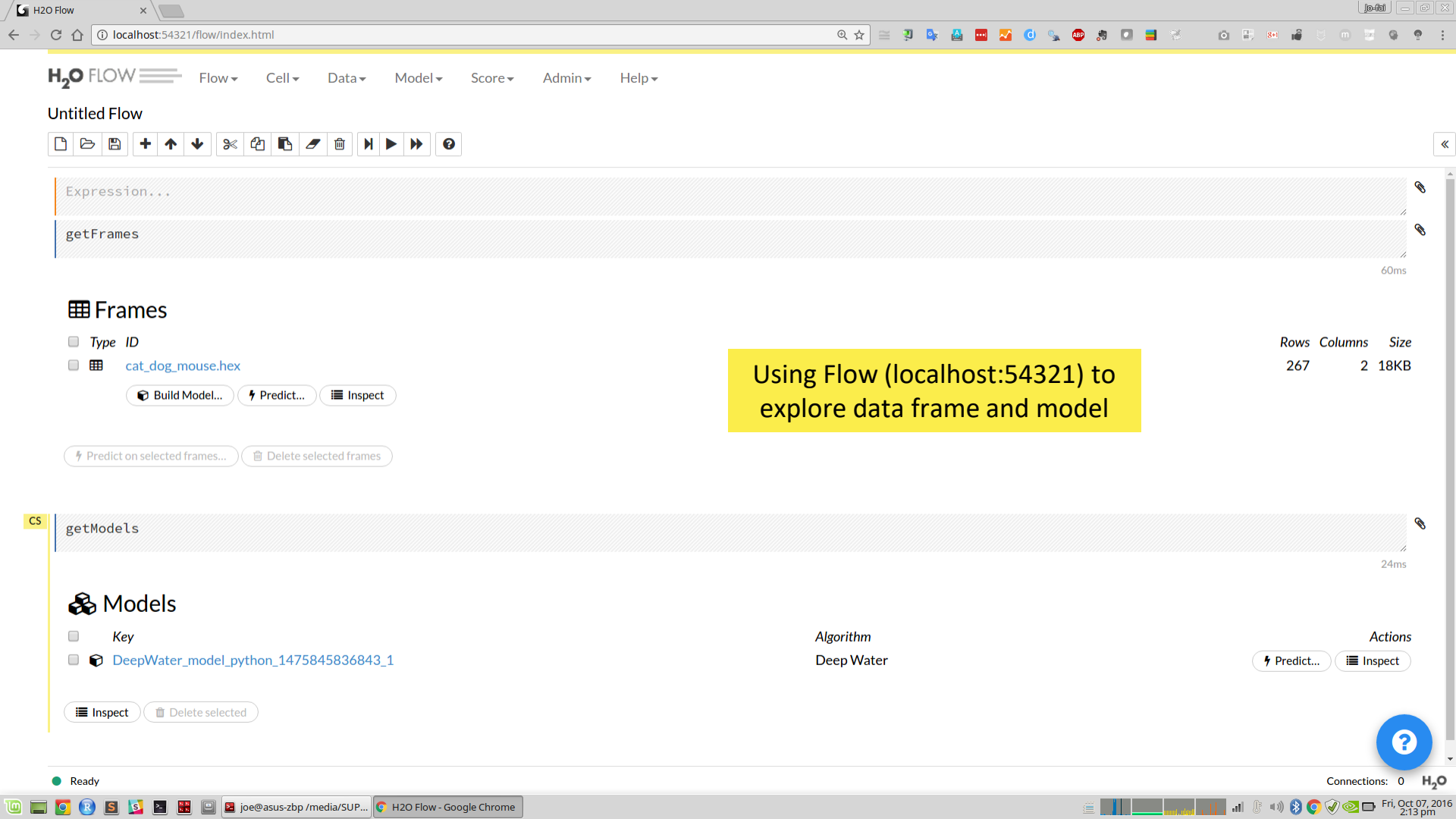
Thanks!

- Big Data London
- Data Science London

- Resources
  - github.com/h2oai/h2o-meetups
  - www.h2o.ai
  - docs.h2o.ai
- Contact
  - joe@h2o.ai
  - @matlabulous

Extra Slides

(Screenshots – using Flow for Deep Water)


H2O Flow
localhost:54321/flow/index.html
H2O FLOW
Flow
Cell
Data
Model
Score
Admin
Help
Untitled Flow
Expression...
getFrames
60ms
Frames
Type ID
cat_dog_mouse.hex
Build Model...
Predict...
Inspect
Rows Columns Size
267 2 18KB
Using Flow (localhost:54321) to explore data frame and model
Predict on selected frames...
Delete selected frames
CS
getModels
24ms
Models
Key
Algorithm
Actions
DeepWater_model_python_1475845836843_1
Deep Water
Predict...
Inspect
Inspect
Delete selected
Ready
Connections: 0
H2O
joe@asus-zbp /media/SUP...
H2O Flow - Google Chrome
Fri, Oct 07, 2016
2:13 pm

H2O Flow

localhost:54321/flow/index.html

H2O FLOW Flow Cell Data Model Score Admin Help

Untitled Flow

Model

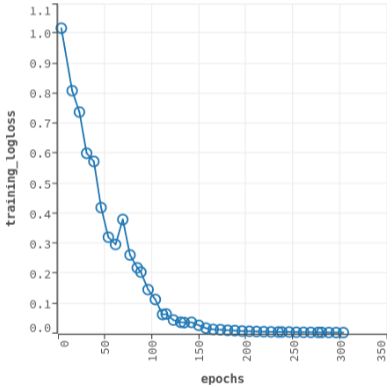
Model ID: DeepWater_model_python_1475845836843_1

Algorithm: Deep Water

Actions: Refresh | Predict... | Download POJO | Export | Inspect | Delete

▸ MODEL PARAMETERS

▾ SCORING HISTORY - LOGLOSS



Using Flow (localhost:54321) to explore data frame and model

▾ TRAINING METRICS - CONFUSION MATRIX VERTICAL: ACTUAL; ACROSS: PREDICTED

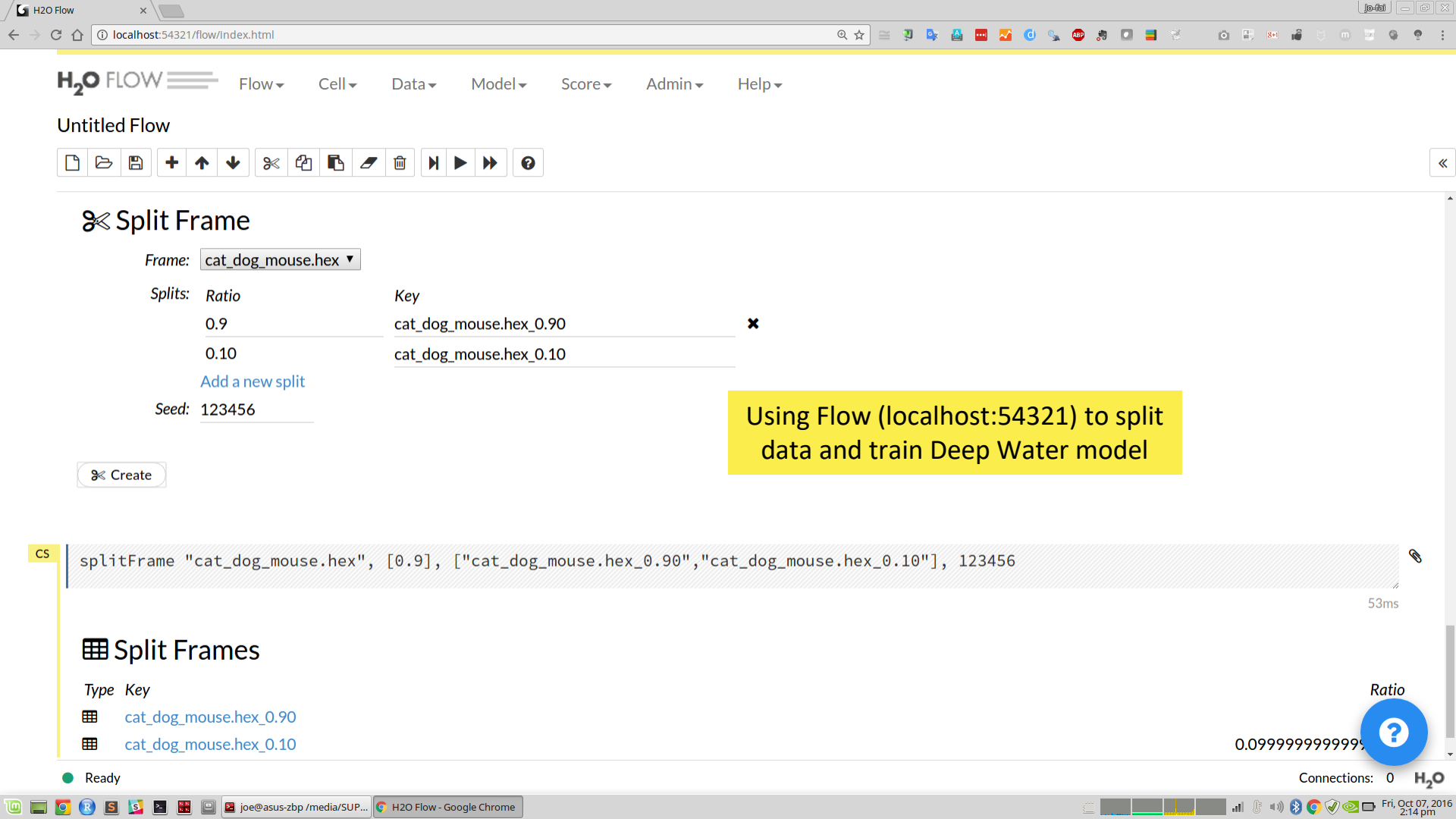
	cat	dog	mouse	Error	Rate
cat	90	0	0	**0**	**0 / 90**
dog	0	85	0	**0**	**0 / 85**
mouse	0	0	92	**0**	**0 / 92**
Total	**90**	**85**	**92**	**0**	**0 / 267**

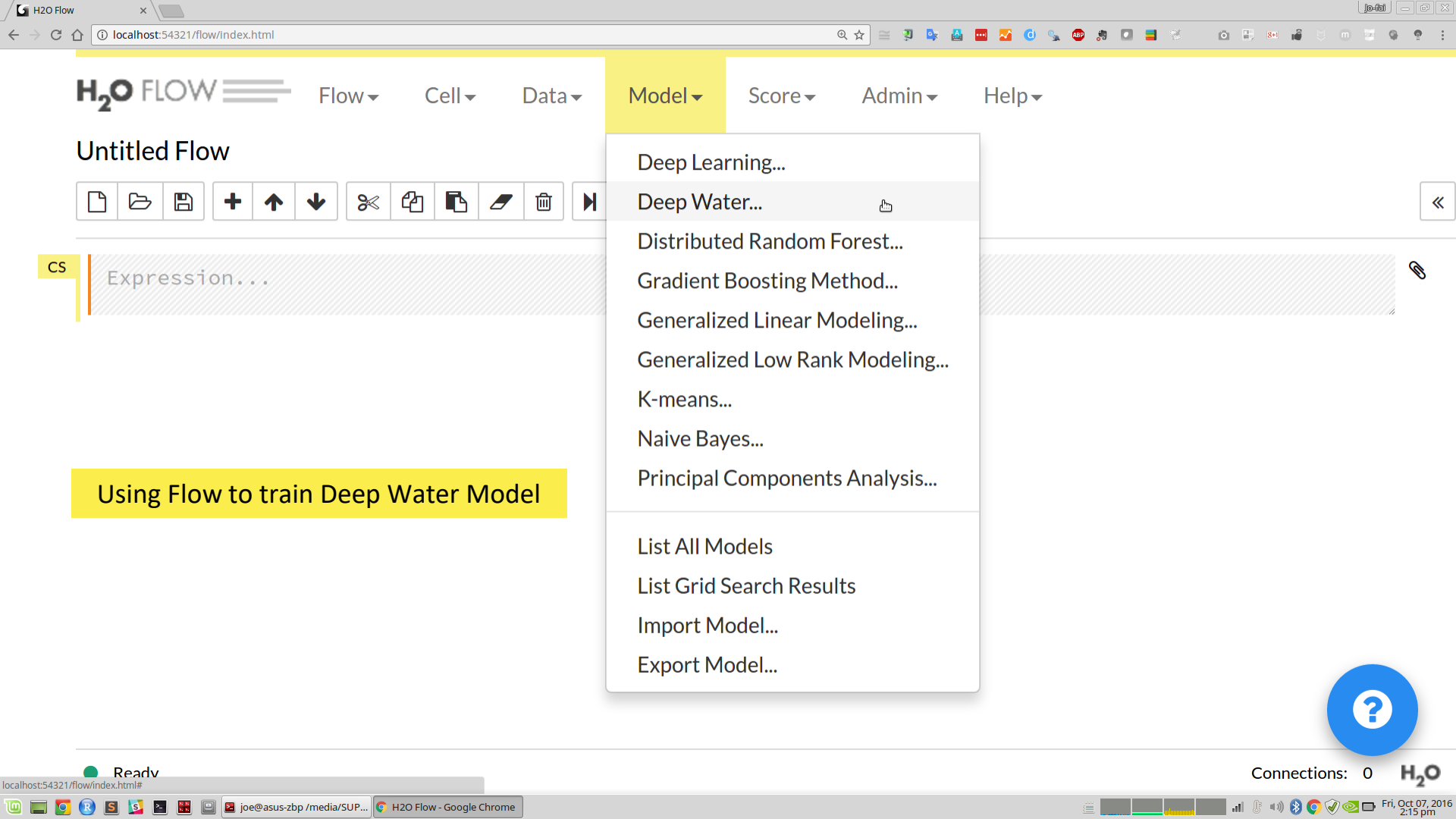
Ready

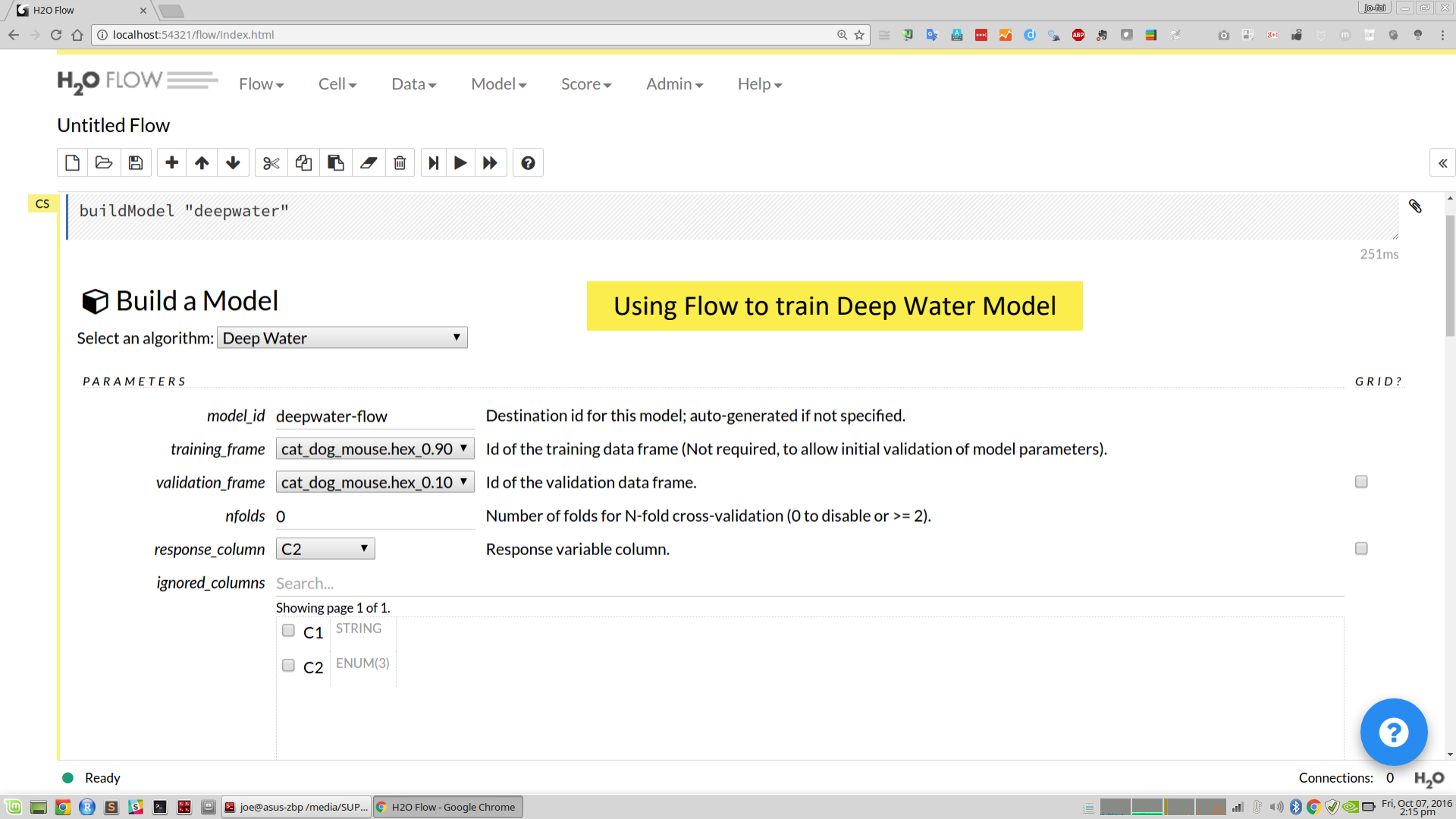
Connections: 0

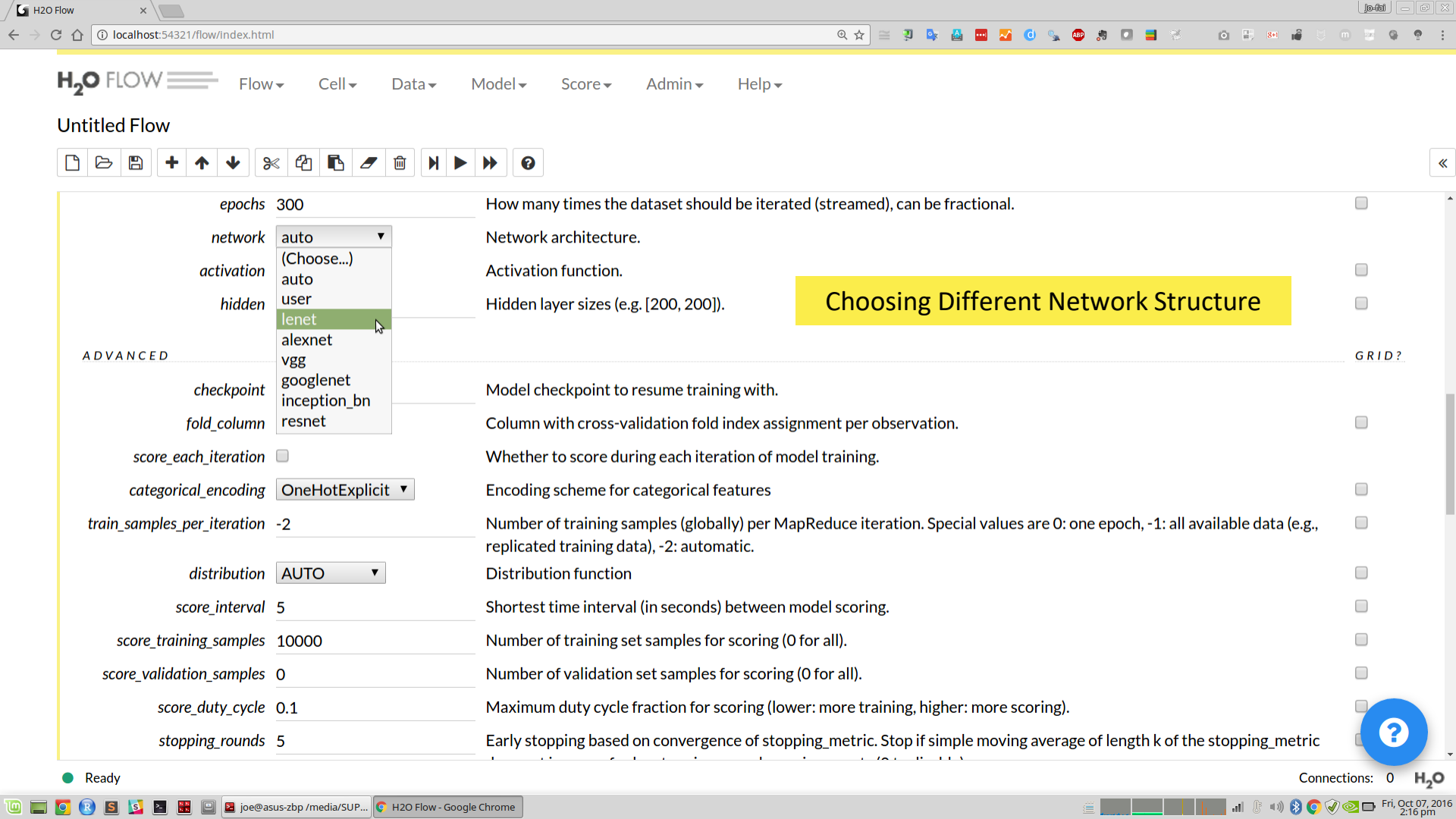
joe@asus-zbp /media/SUP... | H2O Flow - Google Chrome

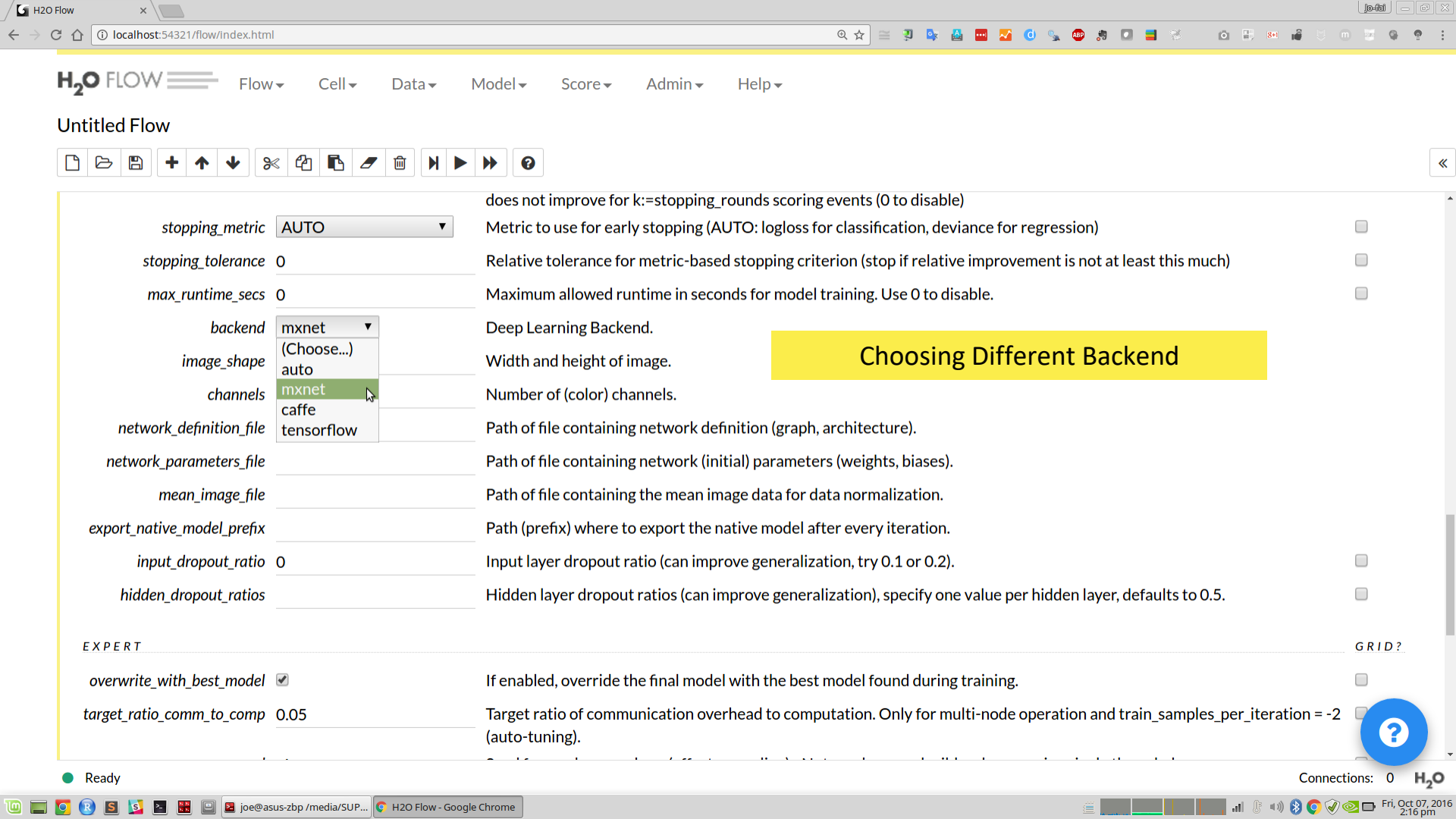
Fri, Oct 07, 2016 2:13 pm


H2O Flow
localhost:54321/flow/index.html
H2O FLOW
Flow
Cell
Data
Model
Score
Admin
Help
Untitled Flow
Split Frame
Frame: cat_dog_mouse.hex
Splits: Ratio Key
0.9 cat_dog_mouse.hex_0.90
0.10 cat_dog_mouse.hex_0.10
Add a new split
Seed: 123456
Create
Using Flow (localhost:54321) to split data and train Deep Water model
CS
splitFrame "cat_dog_mouse.hex", [0.9], ["cat_dog_mouse.hex_0.90","cat_dog_mouse.hex_0.10"], 123456
53ms
Split Frames
Type Key Ratio
cat_dog_mouse.hex_0.90
cat_dog_mouse.hex_0.10
0.0999999999999
Ready
Connections: 0
H2O
joe@asus-zbp /media/SUP...
H2O Flow - Google Chrome
Fri, Oct 07, 2016
2:14 pm


H2O Flow
localhost:54321/flow/index.html
H2O FLOW
Flow
Cell
Data
Model
Score
Admin
Help
Untitled Flow
CS
Expression...
Deep Learning...
Deep Water...
Distributed Random Forest...
Gradient Boosting Method...
Generalized Linear Modeling...
Generalized Low Rank Modeling...
K-means...
Naive Bayes...
Principal Components Analysis...
List All Models
List Grid Search Results
Import Model...
Export Model...
Using Flow to train Deep Water Model
Ready
Connections: 0
H2O
localhost:54321/flow/index.html#
joe@asus-zbp /media/SUP...
H2O Flow - Google Chrome
Fri, Oct 07, 2016
2:15 pm


H2O Flow
localhost:54321/flow/index.html
H2O FLOW
Flow
Cell
Data
Model
Score
Admin
Help
Untitled Flow
CS
buildModel "deepwater"
251ms
Build a Model
Using Flow to train Deep Water Model
Select an algorithm: Deep Water
PARAMETERS
GRID?
model_id
deepwater-flow
Destination id for this model; auto-generated if not specified.
training_frame
cat_dog_mouse.hex_0.90
Id of the training data frame (Not required, to allow initial validation of model parameters).
validation_frame
cat_dog_mouse.hex_0.10
Id of the validation data frame.
nfolds
0
Number of folds for N-fold cross-validation (0 to disable or >= 2).
response_column
C2
Response variable column.
ignored_columns
Search...
Showing page 1 of 1.
C1
STRING
C2
ENUM(3)
Ready
Connections: 0
H2O Flow - Google Chrome
Fri, Oct 07, 2016
2:15 pm


H2O Flow
localhost:54321/flow/index.html
H2O FLOW
Flow
Cell
Data
Model
Score
Admin
Help
Untitled Flow
epochs 300 How many times the dataset should be iterated (streamed), can be fractional.
network auto Network architecture.
activation Activation function.
hidden Hidden layer sizes (e.g. [200, 200]).
(Choose...)
auto
user
lenet
alexnet
vgg
googlenet
inception_bn
resnet
Choosing Different Network Structure
ADVANCED
GRID?
checkpoint Model checkpoint to resume training with.
fold_column Column with cross-validation fold index assignment per observation.
score_each_iteration Whether to score during each iteration of model training.
categorical_encoding OneHotExplicit Encoding scheme for categorical features
train_samples_per_iteration -2 Number of training samples (globally) per MapReduce iteration. Special values are 0: one epoch, -1: all available data (e.g., replicated training data), -2: automatic.
distribution AUTO Distribution function
score_interval 5 Shortest time interval (in seconds) between model scoring.
score_training_samples 10000 Number of training set samples for scoring (0 for all).
score_validation_samples 0 Number of validation set samples for scoring (0 for all).
score_duty_cycle 0.1 Maximum duty cycle fraction for scoring (lower: more training, higher: more scoring).
stopping_rounds 5 Early stopping based on convergence of stopping_metric. Stop if simple moving average of length k of the stopping_metric
Ready
Connections: 0
H2O
joe@asus-zbp /media/SUP...
H2O Flow - Google Chrome
Fri, Oct 07, 2016
2:16 pm


H2O Flow
localhost:54321/flow/index.html
H2O FLOW
Flow
Cell
Data
Model
Score
Admin
Help
Untitled Flow
does not improve for k:=stopping_rounds scoring events (0 to disable)
stopping_metric
AUTO
Metric to use for early stopping (AUTO: logloss for classification, deviance for regression)
stopping_tolerance
0
Relative tolerance for metric-based stopping criterion (stop if relative improvement is not at least this much)
max_runtime_secs
0
Maximum allowed runtime in seconds for model training. Use 0 to disable.
backend
mxnet
Deep Learning Backend.
(Choose...)
auto
mxnet
caffe
tensorflow
image_shape
Width and height of image.
channels
Number of (color) channels.
network_definition_file
Path of file containing network definition (graph, architecture).
network_parameters_file
Path of file containing network (initial) parameters (weights, biases).
mean_image_file
Path of file containing the mean image data for data normalization.
export_native_model_prefix
Path (prefix) where to export the native model after every iteration.
input_dropout_ratio
0
Input layer dropout ratio (can improve generalization, try 0.1 or 0.2).
hidden_dropout_ratios
Hidden layer dropout ratios (can improve generalization), specify one value per hidden layer, defaults to 0.5.
Choosing Different Backend
EXPERT
GRID?
overwrite_with_best_model
If enabled, override the final model with the best model found during training.
target_ratio_comm_to_comp
0.05
Target ratio of communication overhead to computation. Only for multi-node operation and train_samples_per_iteration = -2 (auto-tuning).
Ready
Connections: 0
H2O
joe@asus-zbp /media/SUP...
H2O Flow - Google Chrome
Fri, Oct 07, 2016
2:16 pm

H2O Flow

localhost:54321/flow/index.html

H2O FLOW Flow Cell Data Model Score Admin Help

Untitled Flow

CS

```
buildModel 'deepwater', {"model_id":"deepwater-
flow","training_frame":"cat_dog_mouse.hex_0.90","validation_frame":"cat_dog_mouse.hex_0.10","nfolds":0,"response_column":"C2","ignored_c
olumns":[],"epochs":"300","network":"lenet","hidden":
[],"checkpoint":"","score_each_iteration":true,"categorical_encoding":"OneHotExplicit","train_samples_per_iteration":-2,"distribution":"
AUTO","score_interval":"0","score_training_samples":10000,"score_validation_samples":0,"score_duty_cycle":0.1,"stopping_rounds":5,"stopp
ing_metric":"AUTO","stopping_tolerance":0,"max_runtime_secs":0,"backend":"mxnet","image_shape":
[0,0],"channels":3,"network_definition_file":"","network_parameters_file":"","mean_image_file":"","export_native_model_prefix":"","input
_dropout_ratio":0,"hidden_dropout_ratios":
[],"overwrite_with_best_model":true,"target_ratio_comm_to_comp":0.05,"seed":-1,"rate":0.005,"rate_annealing":0.000001,"momentum_start":0
.9,"momentum_ramp":10000,"momentum_stable":0.99,"single_node_mode":false,"shuffle_training_data":true,"mini_batch_size":32,"clip_gradien
t":10,"gpu":true,"device_id":0}
```

4.1s

Job

Run Time	00:00:03.374
Remaining Time	00:00:00.0
Type	Model
Key	deepwater-flow
Description	DeepWater
Status	DONE
Progress	100%

Training Deep Water Models without Programming

Ready

Connections: 0

H2O

joe@asus-zbp /media/SUP... H2O Flow - Google Chrome

Fri, Oct 07, 2016 2:17 pm