Deep Water

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

Data Science Warsaw
Centrum Nowych Technologii UW
9th November, 2016

Agenda

- This Talk
  - Deep Water
    - Deep Learning in H2O
    - Motivation
  - Demo
    - Cat/Dog/Mouse
    - Iris – Grid Search
  - What's Next?

Deep Learning in H2O

A Simple Neural Network

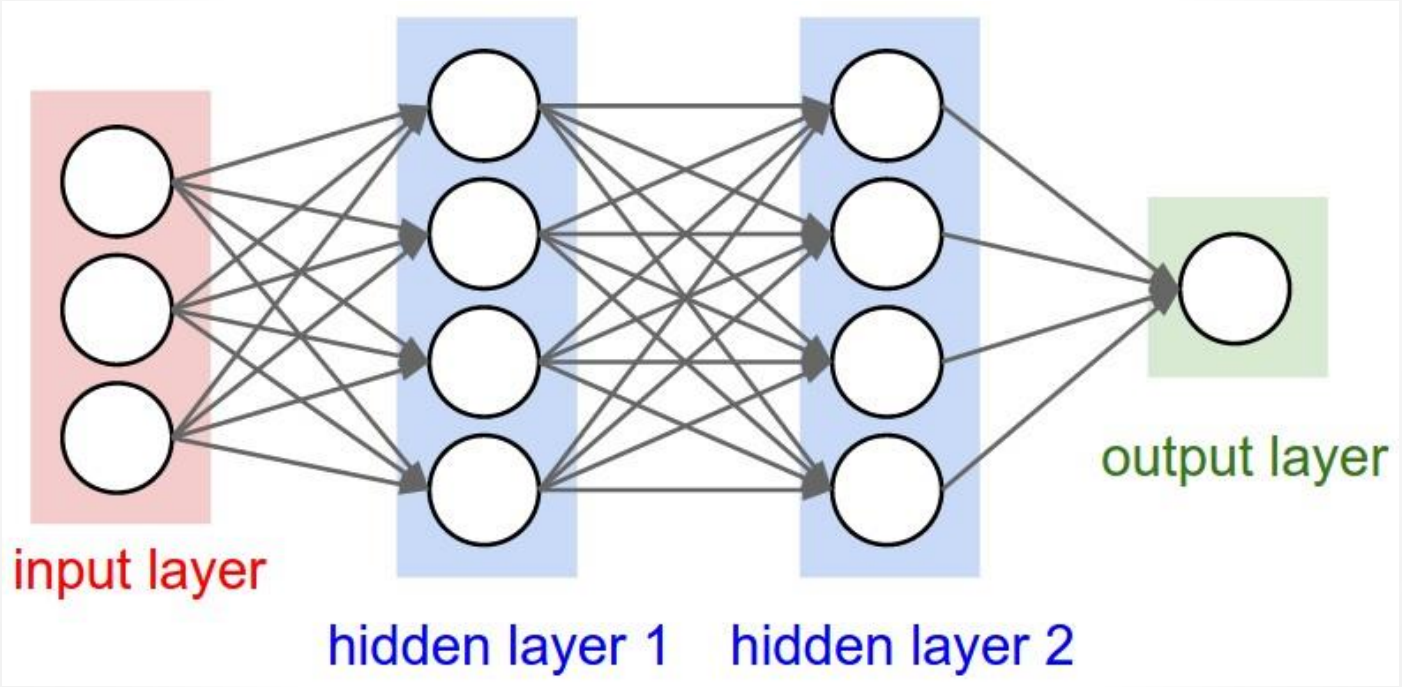


Image credit: http://cs231n.github.io/

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%

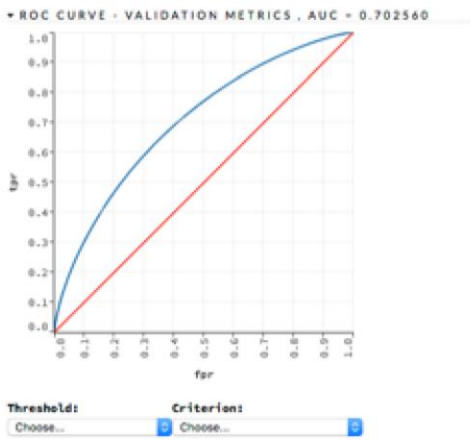
Iterations: 12. Epochs: 0.628821. Speed: 2,243,735 samples/sec. Estimated time left: 21.849 sec

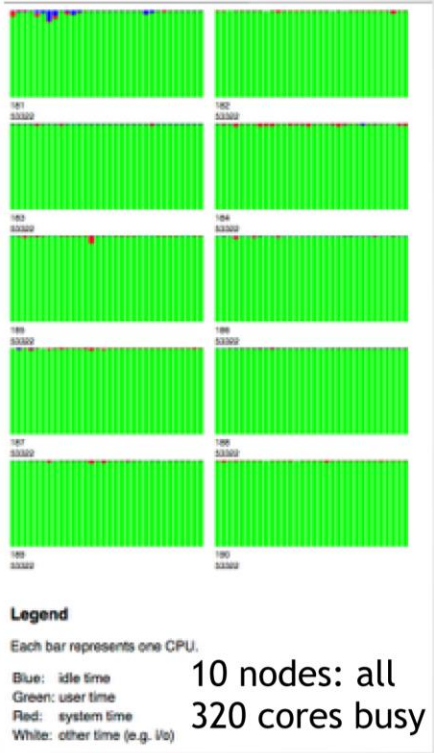
Actions: View | Cancel Job

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

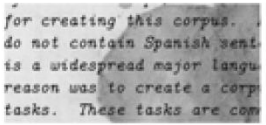
CIFAR-10 Competition Winners: Interviews with Dr. Ben Graham, Phil Culliton, & Zygmunt Zając

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."

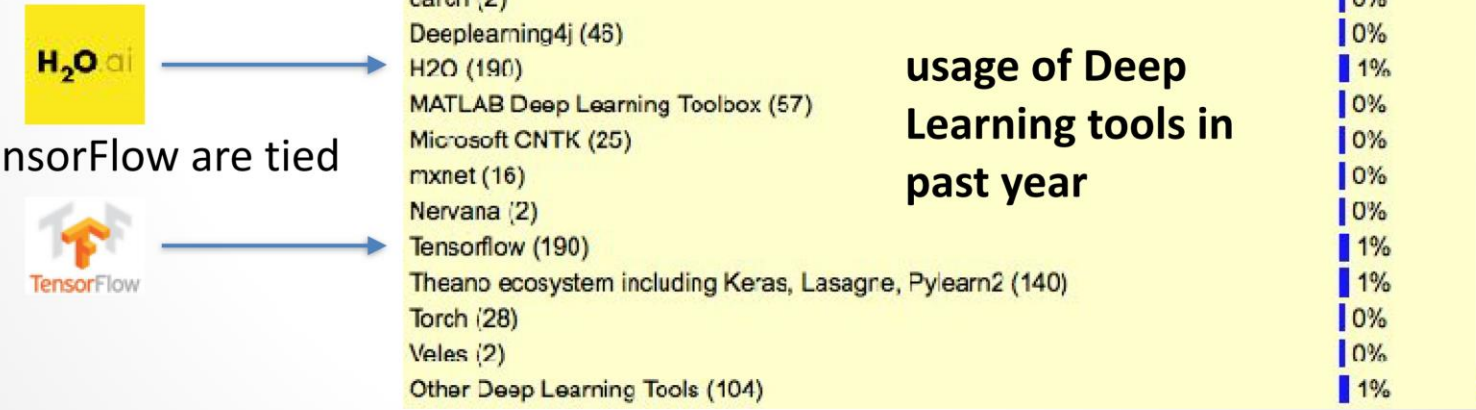
H2O Deep Learning Is Widely Used

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

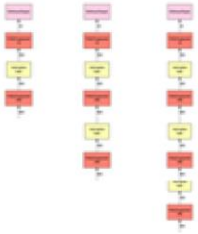


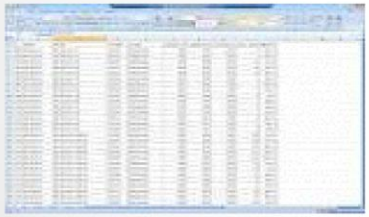
Why Deep Water?

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



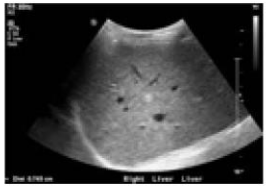


AMD
intel

Limited to business analytics,
statistical models (CSV data)

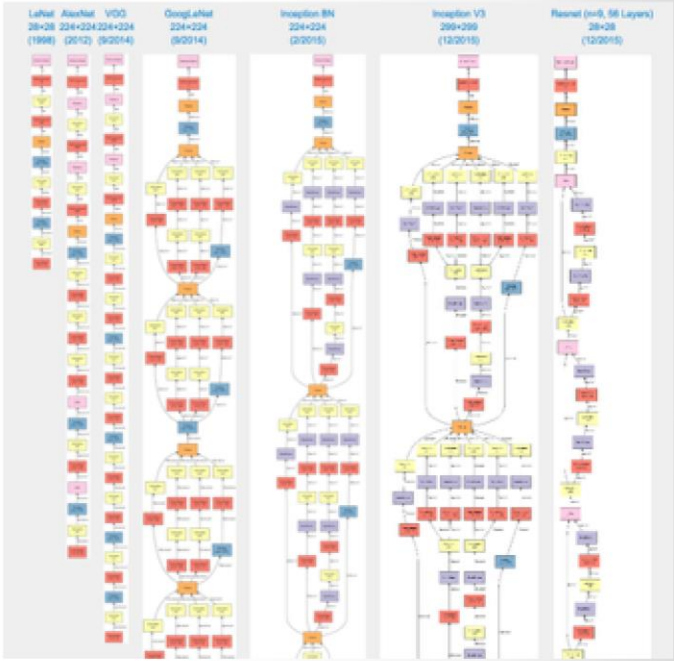
Deep Water: deep complex networks

5-1000 layers
GBs/TBs of data



AMD
intel

NVIDIA
CUDA



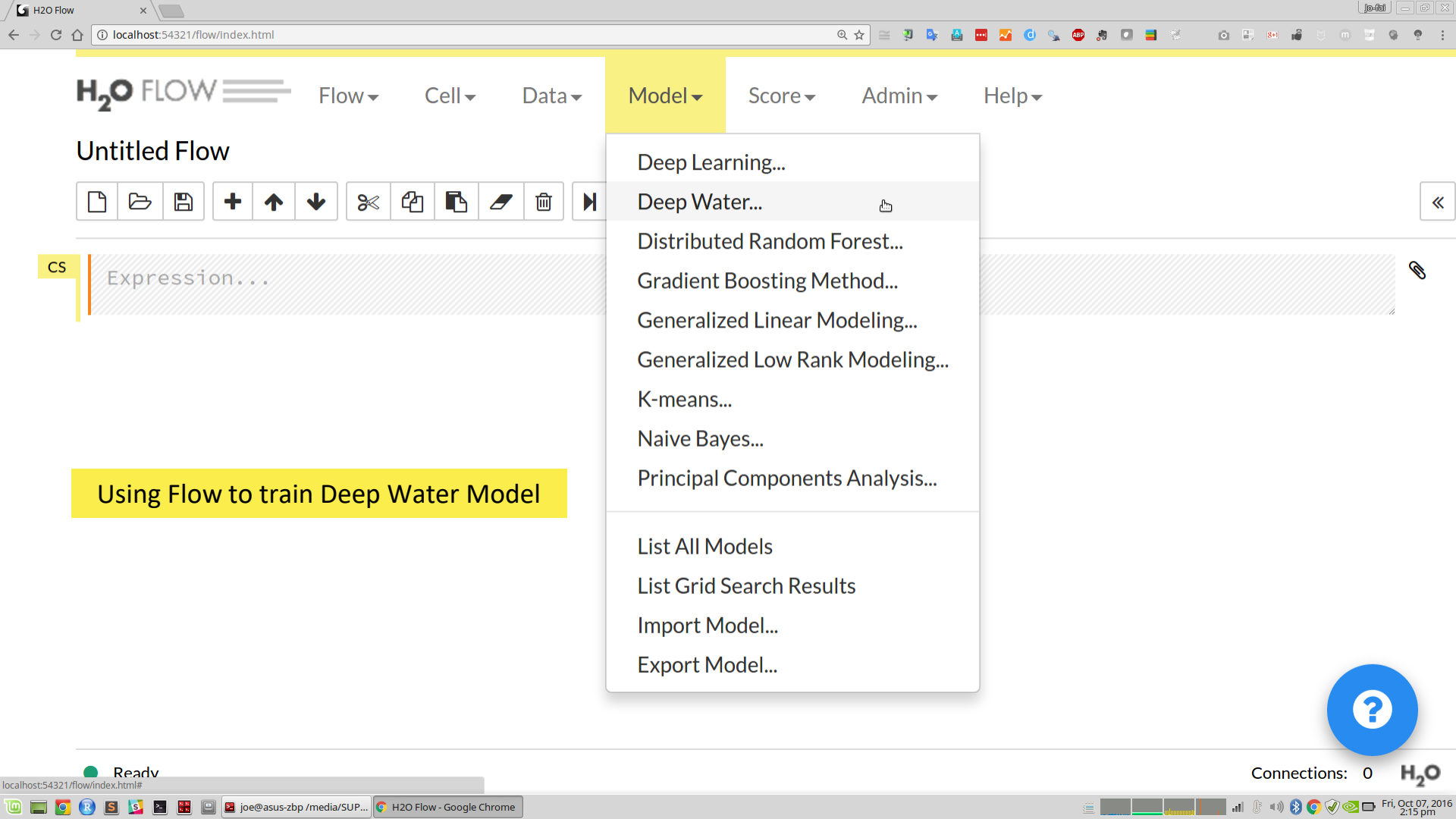
Large networks for big data
(e.g. image 1000x1000x3 -> 3m inputs)

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

Deep Water Roadmap

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
  - Datasets:
    - Cat / Dog / Mouse
    - Iris
  - mxnet GPU backend
  - Training a LeNet (CNN) model
  - Using random grid search for hyper-parameters optimization
- Code and Data
  - bit.ly/h2o_warsaw_1
  - 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.

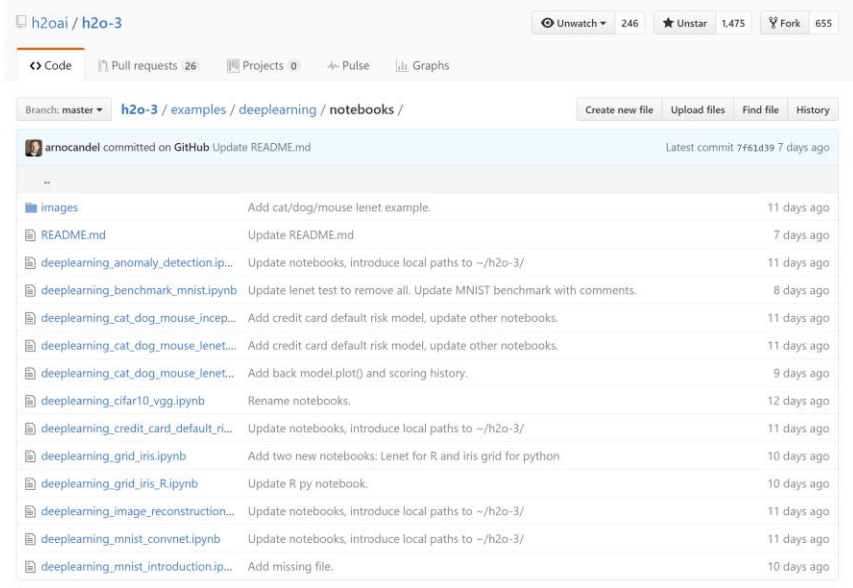
- Required to run Jupyter notebook: H2O Deep Water enabled Python module -- install via `pip install <file>`
- To build custom networks: Matching MXNet Python egg -- install via `easy_install <file>`
- To run from Flow only: H2O Standalone h2o.jar -- launch via `java -jar h2o.jar`

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

PreRelease Amazon AWS Image

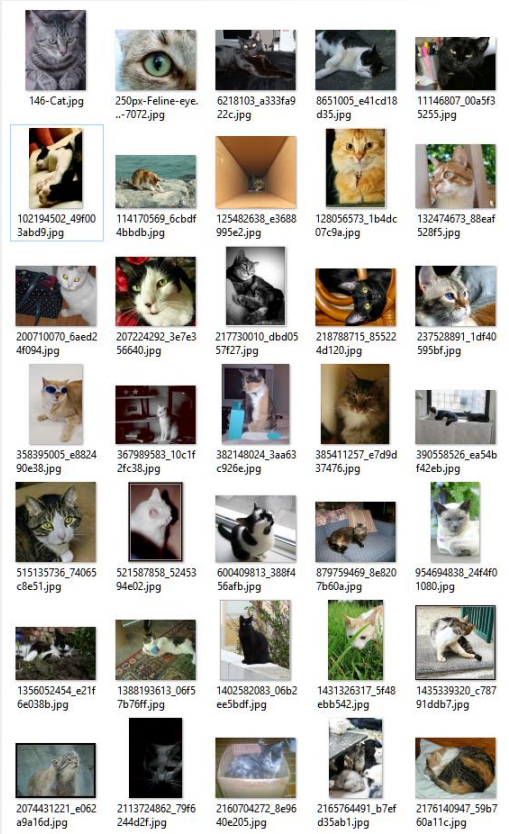
For your convenience, here's a pre-built image for Amazon's EC2 environment, based off our recent H2O Open Tour Hands-On Deep Water workshop (recording coming soon).

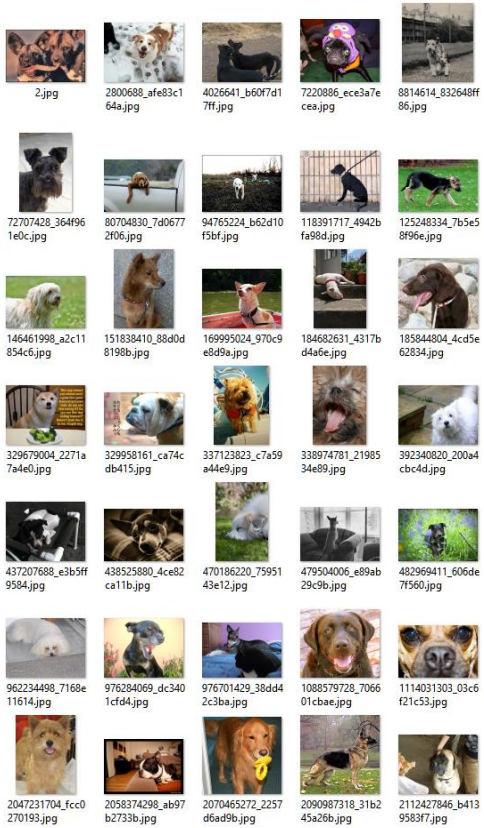
- AMI ID: ami-d32f70c4
- AMI Name: deepwater-dallas-v3
- Recommended instance types: g2.2xlarge or p2.xlarge
- After launching the instance, you can connect to port 8888 (Jupyter Notebook) or port 54321 (H2O Flow).

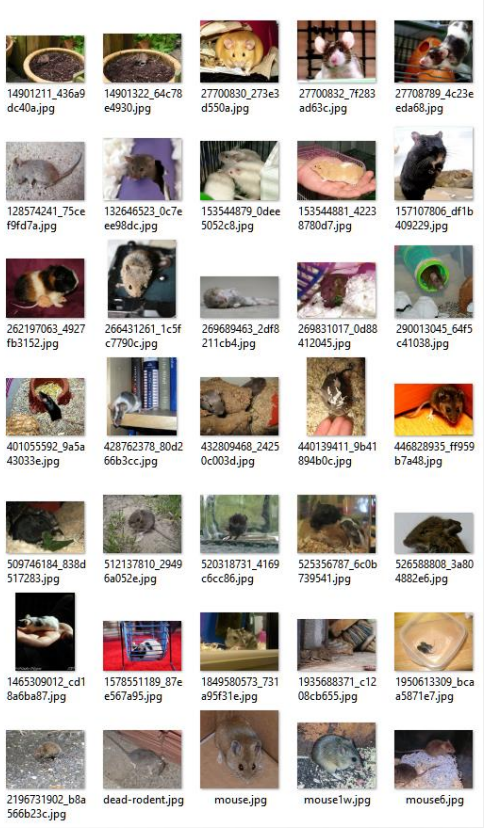


github.com/h2oai/deepwater

Data – Cat/Dog/Mouse Images







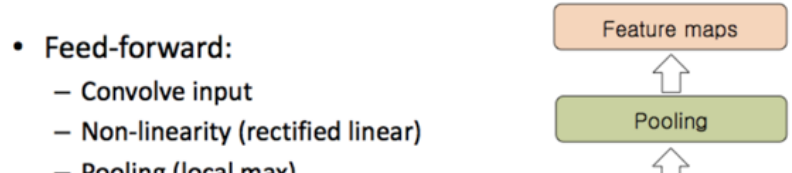
Data - CSV

	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

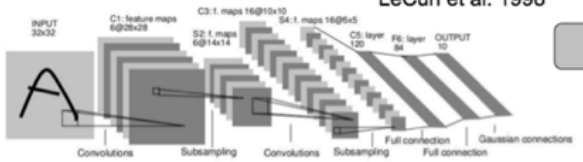
LeNet – Convolutional Neural Network

Convolutional Network

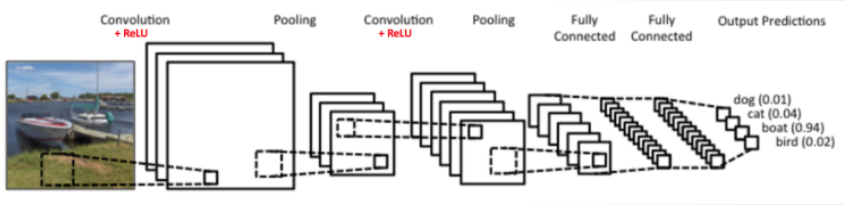
- Feed-forward:
  - Convolve input
  - Non-linearity (rectified linear)
  - Pooling (local max)
- Supervised
- Train convolutional filters by back-propagating classification error



LeCun et al. 1998



Slide: R. Fergus



Deep Water Demo

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

H2O's Mission

Making Machine Learning Accessible to Everyone

Photo credit: Virgin Media

Dziękuję bardzo!

- Data Science Warsaw
  - Dominik Batorski
  - Wit Jakuczun
- Slides & Code
  - bit.ly/h2o_warsaw_1
- Resources
  - github.com/h2oai/h2o-meetups
  - www.h2o.ai
  - docs.h2o.ai
- Contact
  - joe@h2o.ai
  - @matlabulous