H2O Deep Learning

Arno Candel, PhD

Chief Architect, H2O.ai

H2O WORLD 2015

Need a handy conference guide?

Download our app, “H2O World 2015”

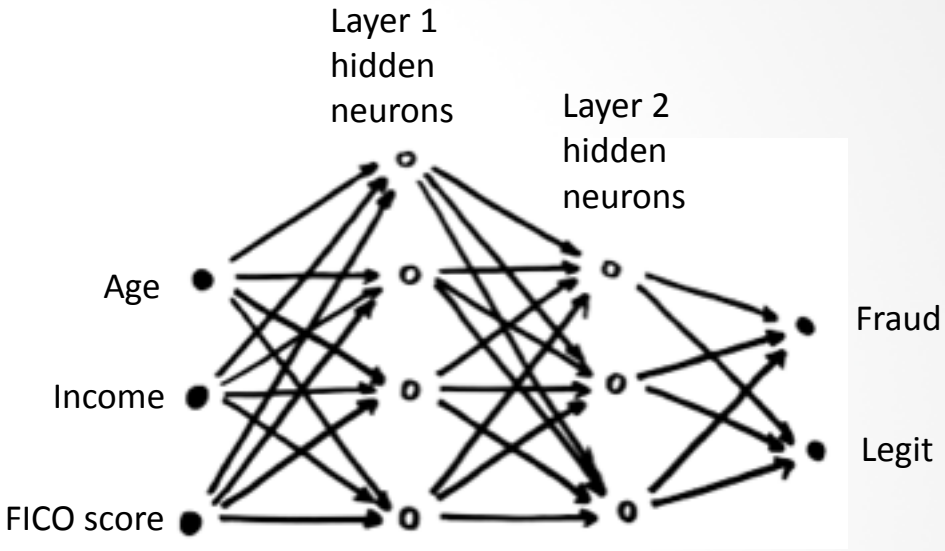
H2O WORLD

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
- Neurons transform their input in nonlinear way
- 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



model

=

set of connecting weights

+

type of non-linearity

Deep Learning: Practical Use

strengths

- non linear
- robust to correlated features
- conceptually simple
- learned features can be extracted
- can stop training at any time
- can be fine-tuned with more data
- great ensemble member
- world-class at pattern recognition

weaknesses

- slow to train
- slow to score
- not interpretable
- results not fully reproducible
- overfits, need regularization
- many hyper-parameters
- expands categorical variables
- must impute missing values

H2O Deep Learning Features

- H2O Eco-System Benefits:
  - Scalable to massive datasets on large clusters, fully parallelized
  - Low-latency Java ("POJO") scoring code is auto-generated
  - Easy to deploy on Laptop, Server, Hadoop cluster, Spark cluster, HPC
  - APIs include R, Python, Flow UI, Scala, Java, JavaScript, REST
- Regularization techniques: Dropout, L1/L2
- Early stopping, N-fold cross-validation, Grid search
- Handling of categorical, missing and sparse data
- Gaussian/Laplace/Poisson/Gamma/Tweedie regression with offsets, observation weights, various loss functions
- Unsupervised mode for non-linear dimensionality reduction, anomaly detection, etc.

Learn More about H2O Deep Learning

Top 10 Deep Learning Tips & Tricks

Tomorrow 11:00 AM Erdos Stage

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

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!)