THE BIG BANG IN MACHINE LEARNING

“Google’s AI engine also reflects how the world of computer hardware is changing. (It) depends on machines equipped with GPUs… And it depends on these chips more than the larger tech universe realizes.”
DEEP LEARNING EVERYWHERE
NN playground

Visualising a neural network

http://playground.tensorflow.org
Recurrent neural networks

\( x \) – input; \( h \) – hidden state vector; \( y \) – output

\( f \) – maps input and previous hidden state into new hidden state; \( g \) – maps hidden state into \( y \)

\( f \) – can be a huge feed-forward network

\[
x_t \xrightarrow{f} h_t \xrightarrow{g} y_t
\]

“classical” feed-forward network with shared weights \( f \)

Most commonly trained by back-propagation

To put \( f \)’s weight update together:
sum with much smaller learning rate
averaging
The problem with RNN

Exploding gradient – large increase of the gradient’s norm due to long term dependencies. Results in large increase of the cost function during training.

Address with Rectified-linear (ReLu) activation function

Vanishing gradient – opposite behaviour, long term components go exponentially fast to 0. Results in bad prediction of long term dependencies.

http://www.willamette.edu/
Long short-term memory (LSTM)

Hochreiter (1991) analysed vanishing gradient “LSTM falls out of this almost naturally”

Gates control importance of the corresponding activations

Training via backprop unfolded in time

Long time dependencies are preserved until input gate is closed (-) and forget gate is open (O)

Fig from Vinyals et al, Google April 2015 NIC Generator

Fig from Graves, Schmidhuber et al, Supervised Sequence Labelling with RNNs
Optimising RNNs with cuDNN v5

ParallelForAll


Supports:

- ReLU & tanh activation functions
- Gated Recurrent Units (GRU)
- Long Short-Term Memory (LSTM)
REINFORCEMENT LEARNING

http://www.ausy.tu-darmstadt.de/
Bellman Principle and Q-function

$Q(s; a)$: the maximum expected return achievable by following any strategy after seeing sequence $s$ and taking action $a$

**Bellman principle (dynamic programming)**

if the optimal value $Q(s(t+1); a(t+1))$ of the sequence $s$ at the next time-step $(t+1)$ was known for all possible actions $a'$, then the optimal strategy is to select the action $a'$ which maximizes the expected value of

$$r(t+1) + Q(s(t+1); a(t+1)) \Rightarrow \text{max}$$
DQN: deep Q-learning network

Mastering Breakout

Video: https://www.youtube.com/watch?v=TmPfTpjtdgg
Further resources

- RLPy Framework - Value-Function-Based Reinforcement
- TeachingBox - Java based RL framework
- BeliefBox - Bayesian reinforcement learning library and toolkit
  - [https://code.google.com/p/beliefbox/](https://code.google.com/p/beliefbox/)
- Deep Q-Learning with Tensor Flow
  - [https://github.com/nivwusquorum/tensorflow-deepq](https://github.com/nivwusquorum/tensorflow-deepq)
Sergey Levine

http://videolectures.net/iclr2016_levine_deep_learning/
## END-TO-END PRODUCT FAMILY

<table>
<thead>
<tr>
<th>HYPERSONELE HPC</th>
<th>MIXED-APPS HPC</th>
<th>STRONG-SCALING HPC</th>
<th>FULLY INTEGRATED DL SUPERCOMPUTER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tesla M4, M40</strong></td>
<td><strong>Tesla K80</strong></td>
<td><strong>Tesla P100</strong></td>
<td><strong>DGX-1</strong></td>
</tr>
<tr>
<td>Hyperscale deployment for DL training, inference, video &amp; image processing</td>
<td>HPC data centers running mix of CPU and GPU workloads</td>
<td>Hyperscale &amp; HPC data centers running apps that scale to multiple GPUs</td>
<td>For customers who need to get going now with fully integrated solution</td>
</tr>
</tbody>
</table>
LEARN MORE

nvidia.com/dgx1
nvidia.com/deeplearning
ONLINE CLASSES
Getting started with deep learning

developer.nvidia.com/deep-learning

DEEP LEARNING
Deep learning is the fastest-growing field in machine learning. It uses many-layered Deep Neural Networks (DNNs) to learn levels of representation and abstraction that make sense of data such as images, sound, and text.

Get Started With Deep Learning
Download Deep Learning Software
Deep Learning Institute

NVIDIA GPUs - The Engine of Deep Learning
Traditional machine learning uses handwritten feature extraction and modality-specific machine learning algorithms to label images or recognize voices. However, this method has several drawbacks in both time-to-solution and accuracy.

Today's advanced deep neural networks use algorithms, big data, and the computational power of the GPU to change this dynamic. Machines are now able to learn at a speed, accuracy, and scale that are driving true artificial intelligence.
GTC Europe is a two-day conference designed to expose the innovative ways developers, businesses and academics are using parallel computing to transform our world.

2 Days | 1,000 Attendees | 50+ Exhibitors | 50+ Speakers | 10+ Tracks | 15+ Hands-on Labs | 1-to-1 Meetings
Questions?
alowndes@nvidia.com