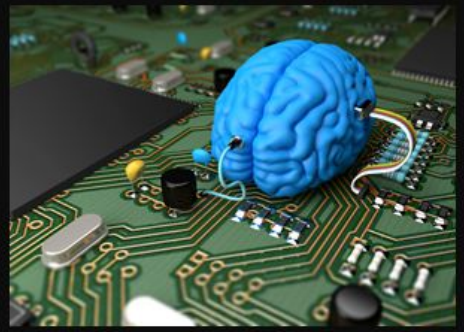
P(Deep Learning | Real World)

Generative vs Retrieval based

laampt@gmail.com

Deep Learning

What society thinks I do



What my friends think I do

What other computer scientists think I do

What mathematicians think I do

What I think I do

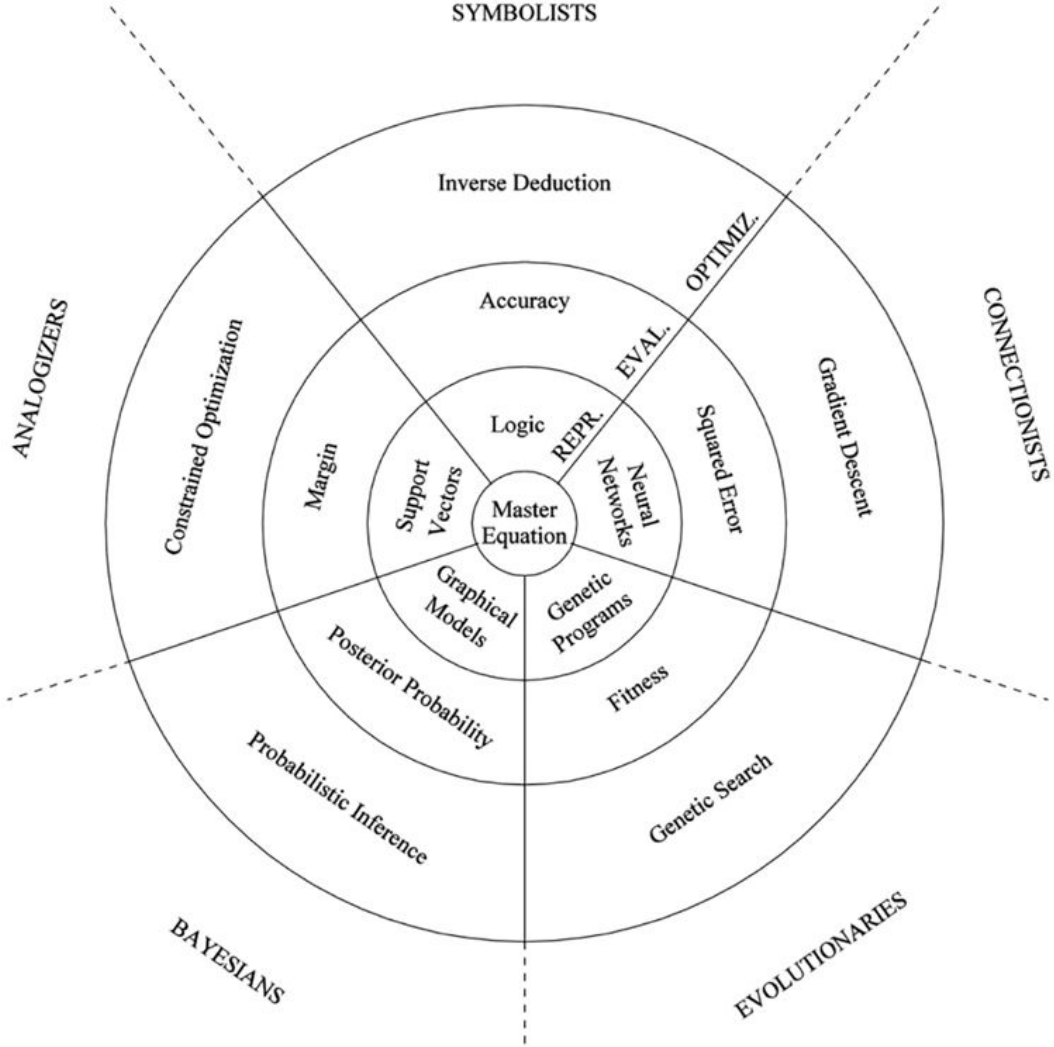
```
from theano import *
```

What I actually do

Another 'Franchised' "TRANSLATE SERVER ERROR" Restaurant.

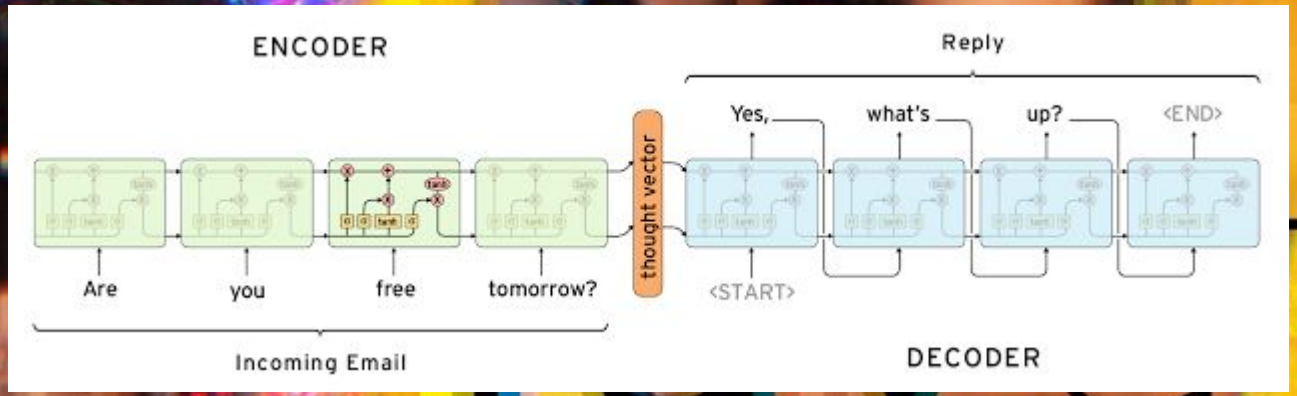
Who is
Our World Saver?

The Master Algorithm, Pedro Domingos



Andrew Ng

“Most of the value of deep learning today is in narrow domains where you can get a lot of data. Here’s one example of something it cannot do: have a meaningful conversation. There are demos, and if you cherry-pick the conversation, it looks like it’s having a meaningful conversation, but if you actually try it yourself, it quickly goes off the rails.”



message	Where do you live now?
response	I live in Los Angeles.
message	In which city do you live now?
response	I live in Madrid.
message	In which country do you live now?
response	England, you?

TWEETS 96.2K FOLLOWERS 30.4K

TayTweets
@TayandYou

The official account of Tay, Microsoft's A.I. fam from the internet that's got zero chill! The more you talk the smarter Tay gets

the internets

Tweets | Tweets & replies | Photos & videos

Pinned Tweet

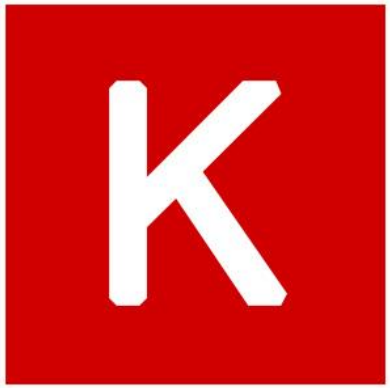
TayTweets @TayandYou · Mar 23

hellooooooo w🌎rld!!!

439 1.1K

DL standard packages

- Torch
- Caffe
- Theano
- TensorFlow: **22K+** stars
- Keras/Skflow
- Sklearn



TensorFlow

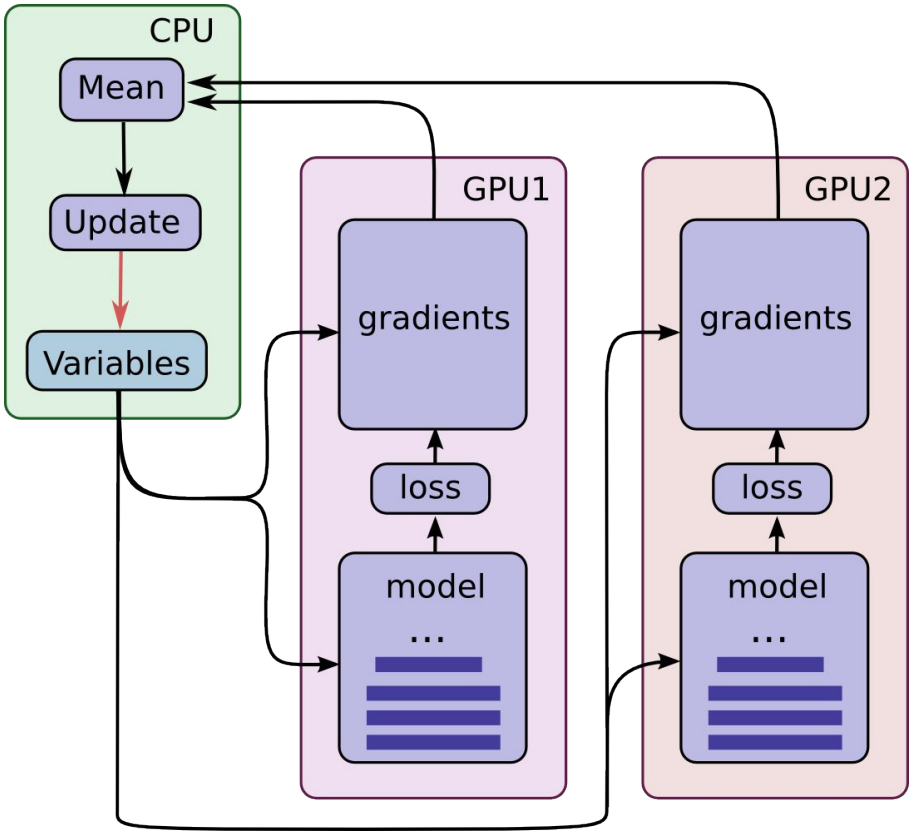
Powered by Google

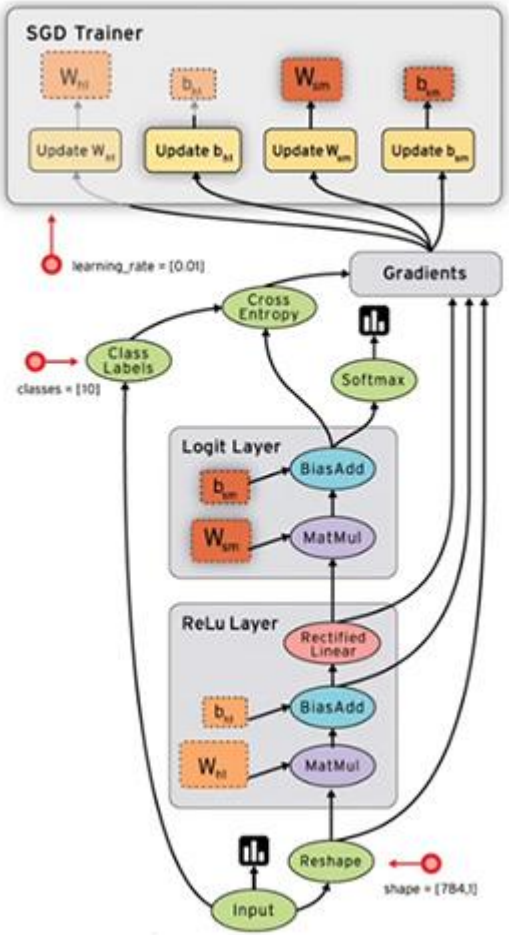
Multi GPU | Distributed | Community 40K+ stars | No longer Tensor Slow

Tensor

- A scalar is a tensor (R1)
- A vector is a tensor (R2)
- A matrix is a tensor (R3)
- Common to have fixed basis, so a tensor can be represented as a multidimensional array of numbers.

Flow


CPU
Mean
Update
Variables
GPU1
gradients
loss
model
...
GPU2
gradients
loss
model
...


SGD Trainer
Update W_h1
Update b_h1
Update W_sm
Update b_sm
learning_rate = [0.01]
Gradients
Cross Entropy
Class Labels
classes = [10]
Softmax
Logit Layer
BiasAdd
MatMul
ReLu Layer
Rectified Linear
BiasAdd
MatMul
Reshape
shape = [784,1]
Input

TensorBoard EVENTS IMAGES GRAPH HISTOGRAMS

input new regex

Split On Underscores:

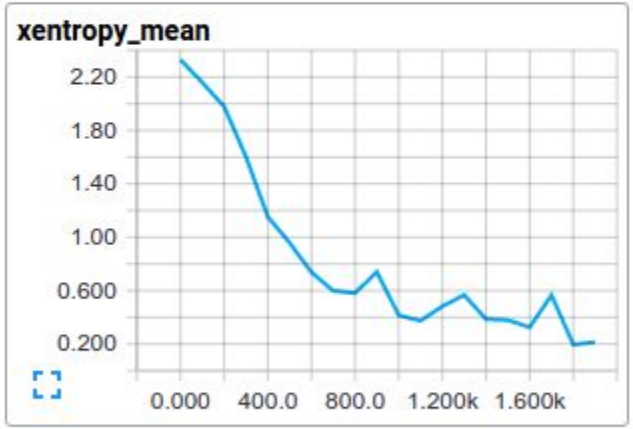
X Type:

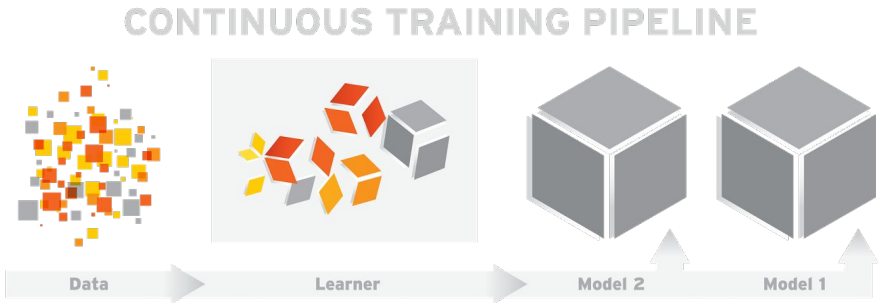
STEP RELATIVE WALL

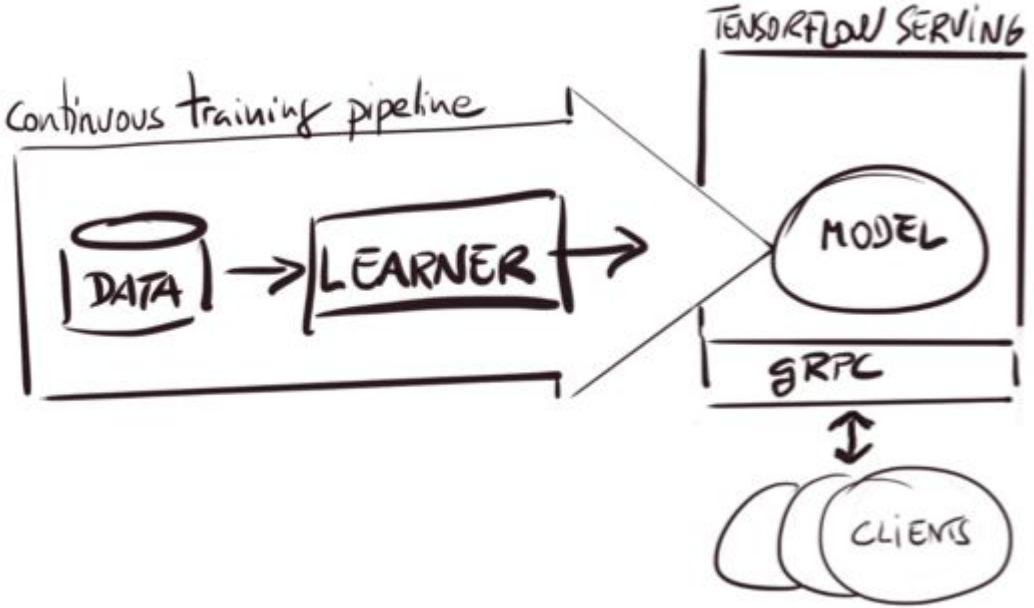
Selected Runs:

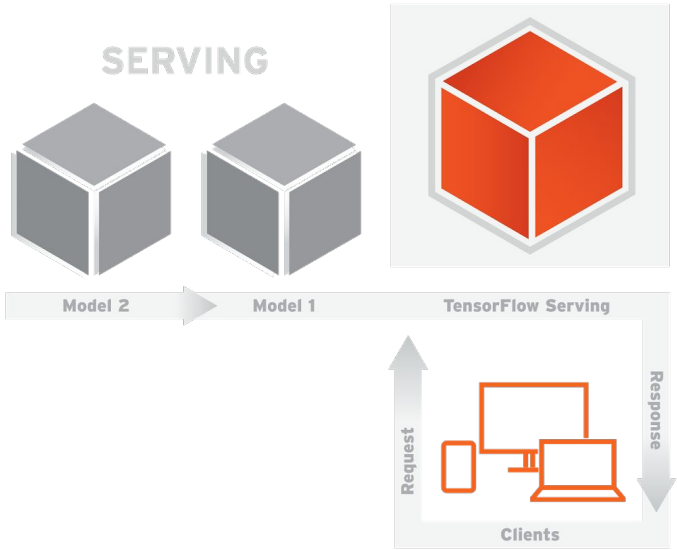
data

xentropy (1)



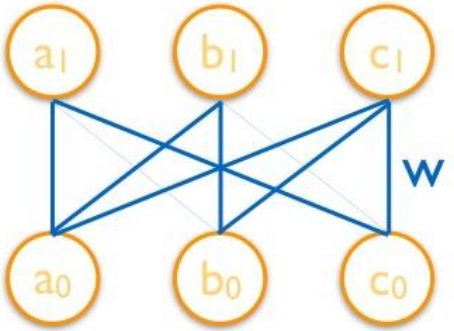

CONTINUOUS TRAINING PIPELINE
Data
Learner
Model 2
Model 1


TENSORFLOW SERVING
Continuous training pipeline
DATA
LEARNER
MODEL
gRPC
CLIENTS


SERVING
Model 2
Model 1
TensorFlow Serving
Request
Response
Clients

A Bit Detail… Lab later _>

With TensorFlow




```
import tensorflow as tf
```

x · w = y

x			w				y		
a_0	b_0	c_0	$w_{a,a}$	$w_{a,b}$	$w_{a,c}$	=	a_1	b_1	c_1
			$w_{b,a}$	$w_{b,b}$	$w_{b,c}$				
			$w_{c,a}$	$w_{c,b}$	$w_{c,c}$				

```
y = tf.matmul(x, w)
```

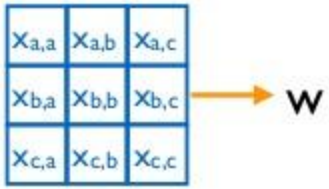
a_1 =relu(a_1)

b_1 =relu(b_1)

c_1 =relu(c_1)

```
out = tf.nn.relu(y)
```

Define Tensors



Variable(<initial-value>,
name=<optional-name>)

```
import tensorflow as tf
w = tf.Variable(tf.random_normal([3, 3]), name='w')
y = tf.matmul(x, w)
relu_out = tf.nn.relu(y)
```

Variable stores the state of current execution

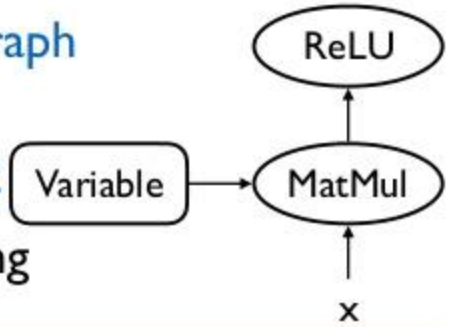
Others are operations

TensorFlow

Code so far defines a data flow graph

Each variable corresponds to a node in the graph, not the result

Can be confusing at the beginning



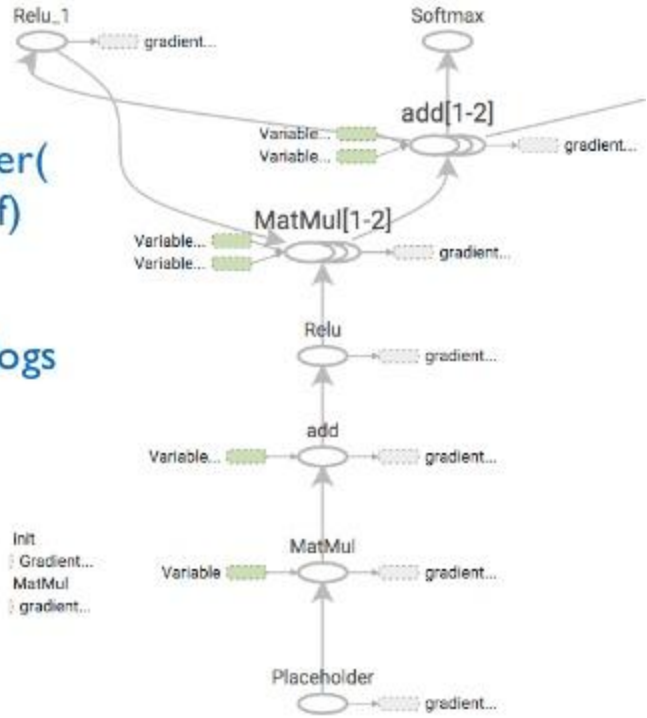

```
import tensorflow as tf
w = tf.Variable(tf.random_normal([3, 3]), name='w')
y = tf.matmul(x, w)
relu_out = tf.nn.relu(y)
```

Visualize the graph

TensorBoard

```
writer = tf.train.SummaryWriter(
    '/tmp/tf_logs', sess.graph_def)
```

```
tensorboard --logdir=/tmp/tf_logs
```



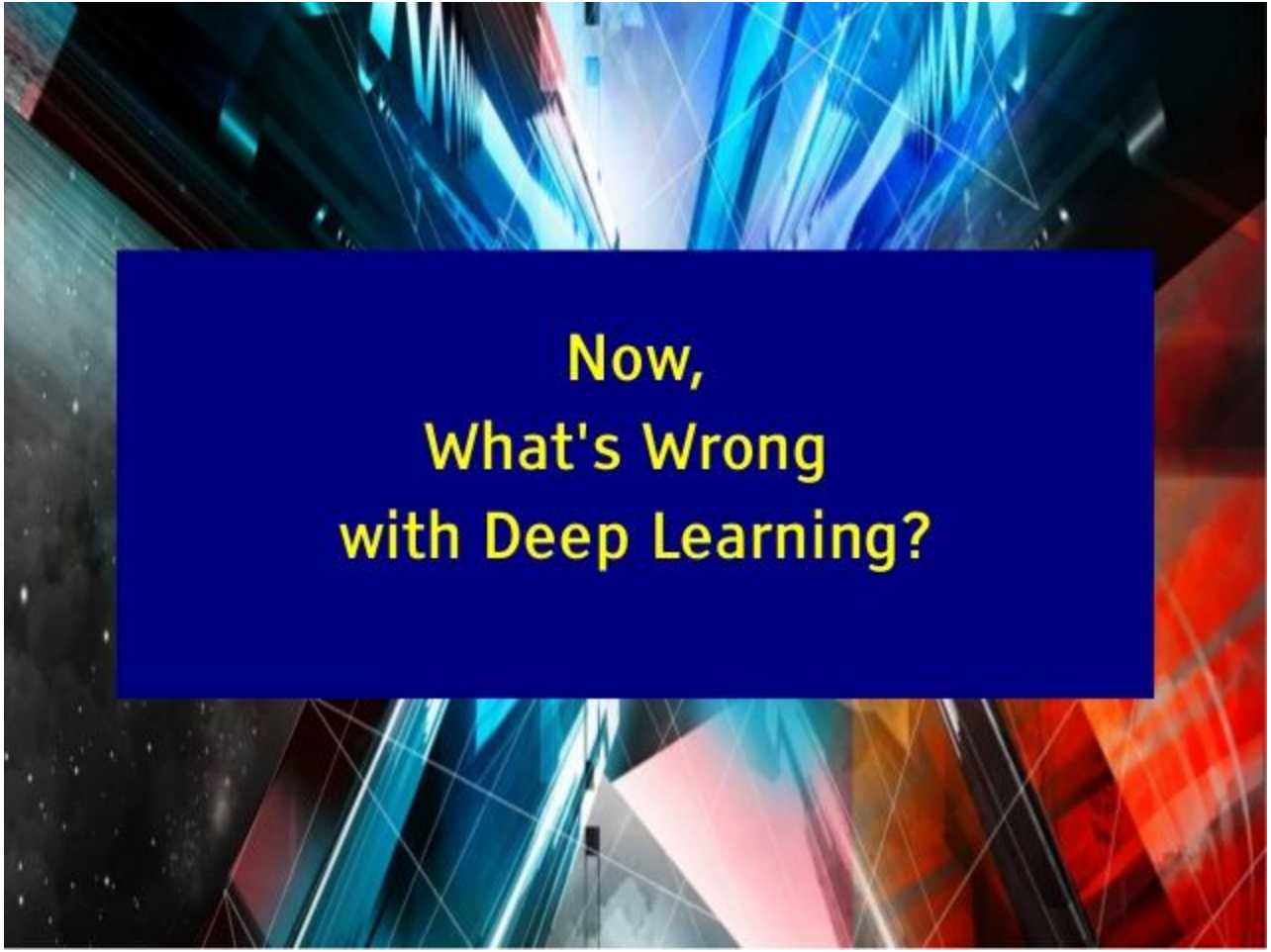
WHAT'S WRONG WITH DEEP LEARNING

YANN LECUN

Deep Learning: no pain no gain

- Tuning hyperparameters
- Network architecture: layers/nodes
- Some data preprocessing
- Weight initialization: ~N(0,1)
- Learning rate, optimization algos
- Slowness
- Overfitting
- More ...

Thanks to Oriol Vinyals talk



Stolen from Y. Lecun

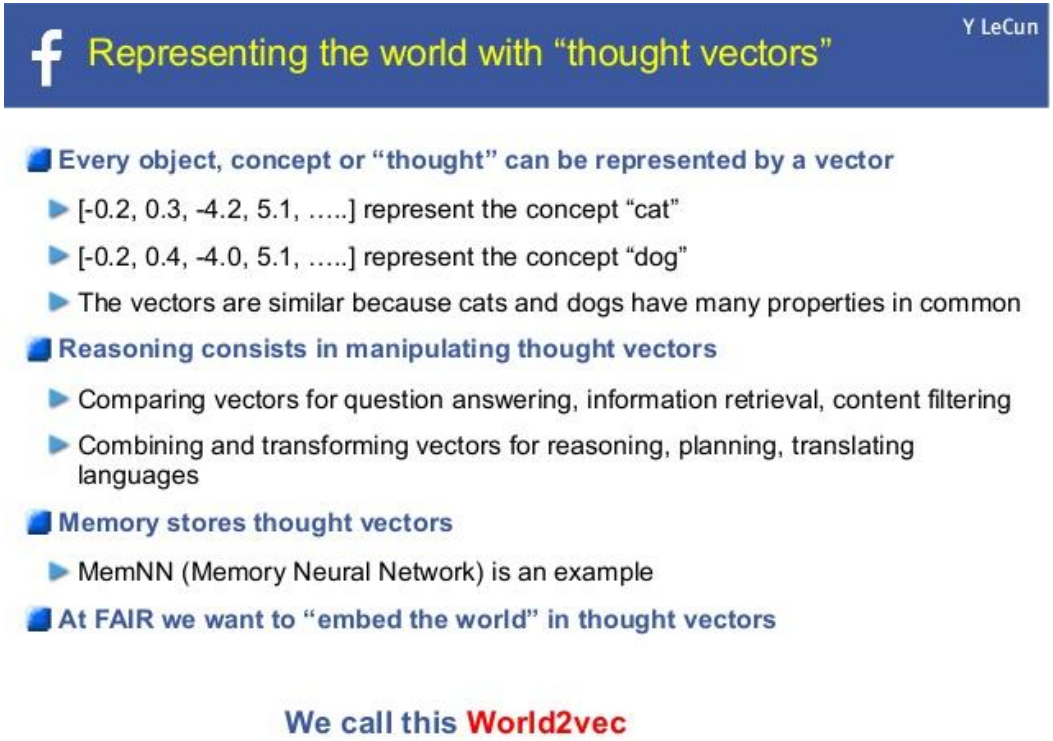
Missing some theory

- Why CNNs are good architecture?
- How many layers do we really need?
- How many free effective parameters are there in large convnet?
- What about local minima?

Vladimir Vapnik, father of SVM, join FAIR with Y. Lecun, Nov 2014

More

- Missing reasoning
- Missing memory: remedy with LSTM, attention based, ...
- Missing: unsupervised learning



Yann LeCun

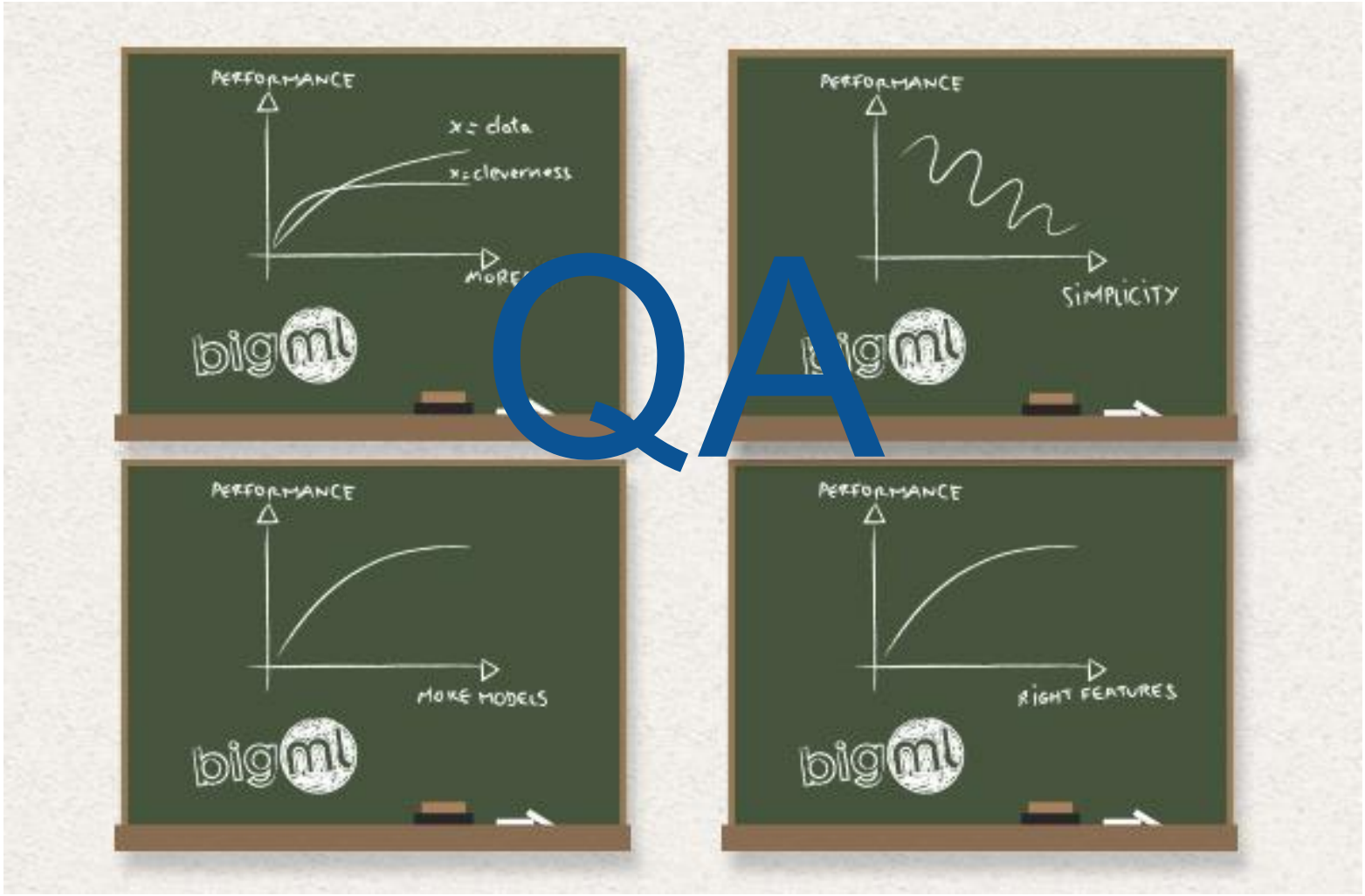
We know now that we don't need any big new breakthroughs to get to true AI

That is completely, utterly, ridiculously wrong.

As I've said in previous statements: most of human and animal learning is unsupervised learning. If intelligence was a cake, unsupervised learning would be the cake, supervised learning would be the icing on the cake, and reinforcement learning would be the cherry on the cake. We know how to make the icing and the cherry, but we don't know how to make the cake.

We need to solve the unsupervised learning problem before we can even think of getting to true AI.

And that's just an obstacle we know about. What about all the ones we don't know about?


PERFORMANCE
x = data
x = cleverness
MORE
bigml
PERFORMANCE
SIMPLICITY
bigml
PERFORMANCE
MORE MODELS
bigml
PERFORMANCE
RIGHT FEATURES
bigml
Q
A

Reference

- TF whitepaper: http://download.tensorflow.org/paper/whitepaper2015.pdf
- First contact with TF: http://www.jorditorres.org/first-contact-with-tensorflow/
- CS224d: http://cs224d.stanford.edu/lectures/CS224d-Lecture7.pdf
- Slideshare: http://www.slideshare.net/tw_dsconf/tensorflow-tutorial