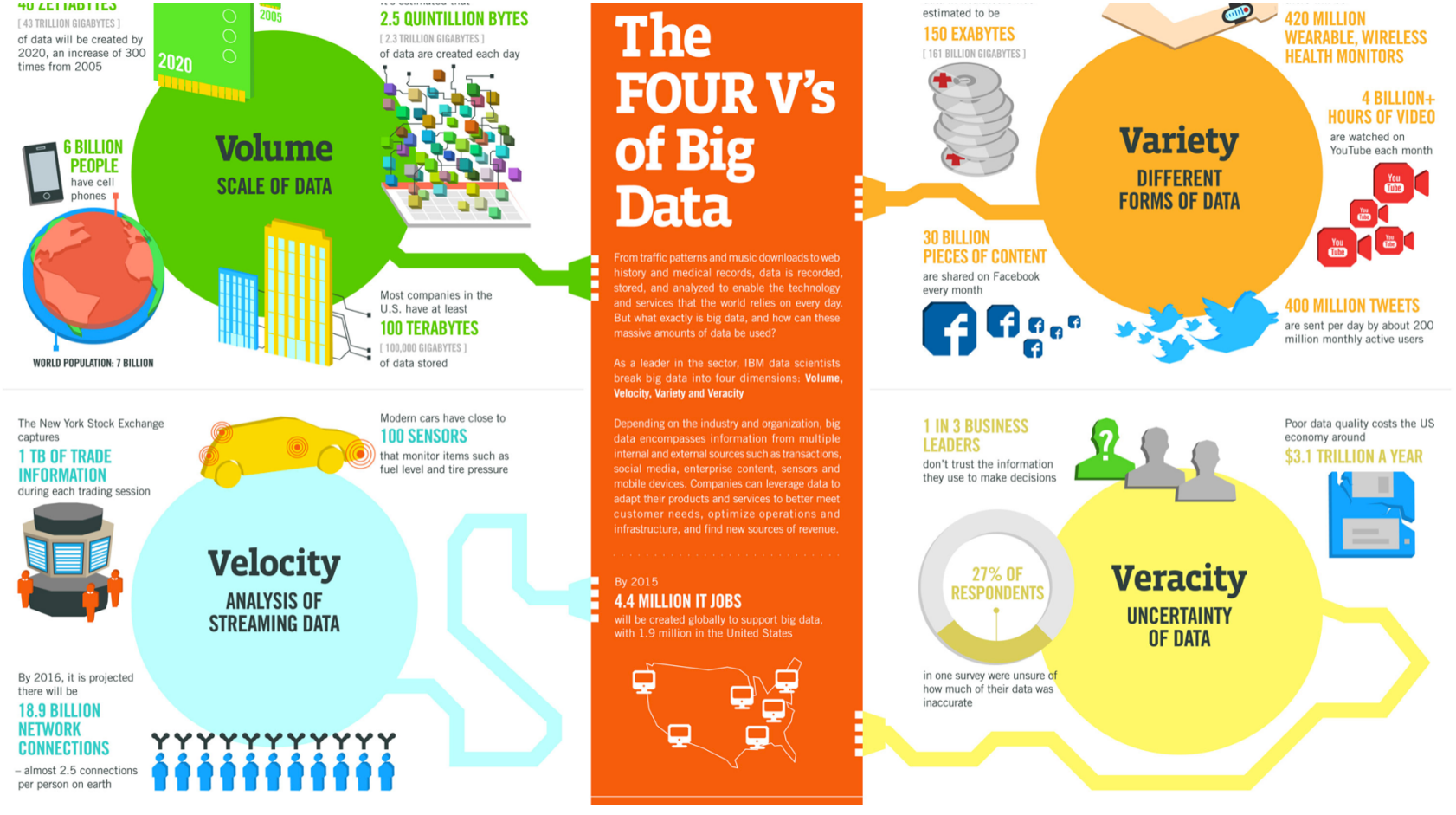
Deep Learning
深度学习
Lecture 01: Introduction

Wanxiang Che

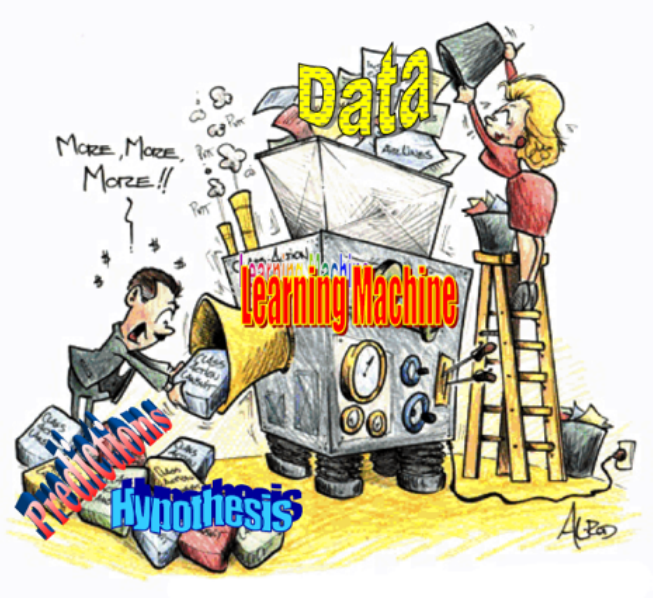
2016-7-2

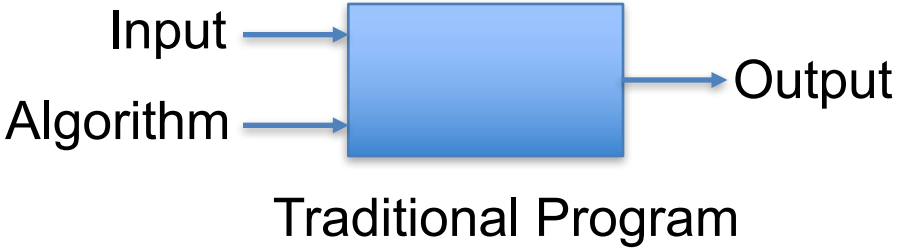
The Age of Big Data (大数据)



What is Machine Learning (机器学习)?

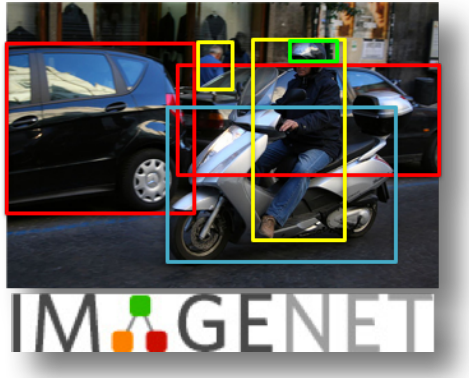
- From Data to Knowledge



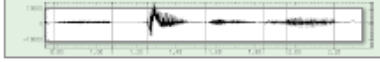
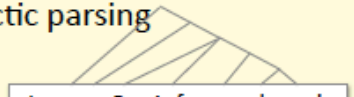


Some Applications of ML

- Recognizing patterns (识别模式)
  - Objects in real scenes
  - Facial identities or facial expressions
  - Spoken words
- Recognizing anomalies (识别异常)
  - Unusual sequences of credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant
- Prediction (预测)
  - Future stock prices or currency exchange rates
  - Which movies will a person like?

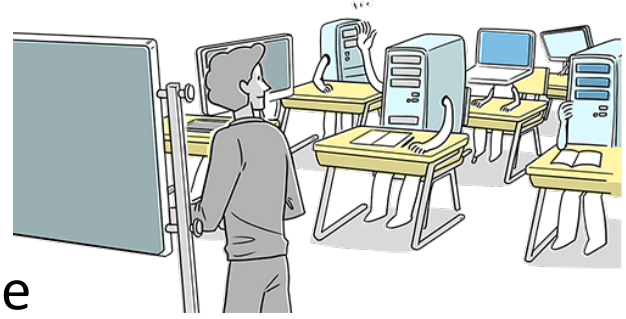


ML for NLP

mostly solved	making good progress	still really hard
Spam detection OK, let's meet by the big ... ✓ D1ck too small? Buy V1AGRA ... ✗	**Sentiment analysis** The pho was authentic and yummy. Waiter ignored us for 20 minutes.	**Semantic search** people protesting globalization Search ...demonstrators stormed IMF offices...
Speech synthesis I'd like you to listen to me 	**Coreference resolution** Obama told Mubarak he shouldn't run again.	**Question answering (QA)** Q. What currency is used in China? A. The yuan 
Part-of-speech (POS) tagging ADJ ADJ NOUN VERB ADV Colorless green ideas sleep furiously.	**Word sense disambiguation (WSD)** I need new batteries for my *mouse*. 	**Textual inference & paraphrase** T. Thirteen soldiers lost their lives ... H. Several troops were killed in the ... YES
Named entity recognition (NER) PERSON ORG LOC Obama met with UAW leaders in Detroit	**Syntactic parsing**  I can see Russia from my house!	**Summarization** Sheen continues rant against ... Sheen is nuts
Information extraction (IE) You're invited to our bunga bunga party, Friday May 27 at 8:30pm in Cordura Hall Party May 27 add	**Machine translation (MT)** Our specialty is panda fried rice. 我们的专长是熊猫炒饭	**Discourse & dialog** Where is Thor playing in SF? Metreon at 4:30 and 7:30

[Christopher Manning]

When to apply ML?



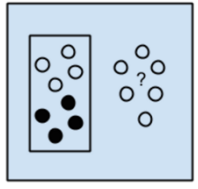
- Humans are unable to explain their expertise
  - e.g. speech recognition, vision, language
- The problem size is to vast for our limited reasoning capabilities
  - e.g. calculating webpage ranks, matching ads to Facebook pages
- Human expertise is absent
  - e.g. navigating on Mars
- Solution changes with time
  - e.g. tracking, temperature control, preferences
- Solution needs to be adapted to particular cases
  - e.g. biometrics, personalization

Types of Learning Task

Machine Learning

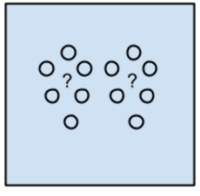
Supervised Learning
有指导学习

Learn to predict an output when given an input



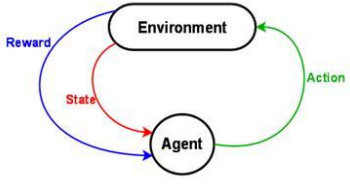
Unsupervised Learning
无指导学习

Discover a good internal representation of input



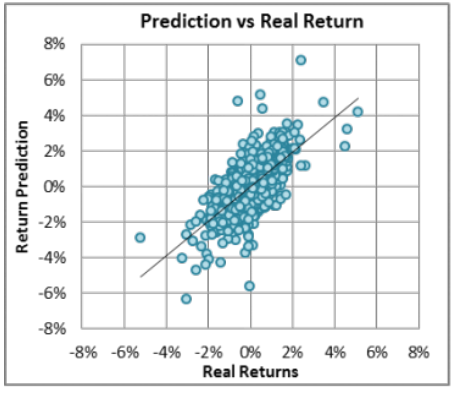
Reinforcement Learning
强化学习

Learn to select an action to maximize future reward



Two Types of Supervised Learning

- Training case/example/data
  - Input vector ***x*** and a target output *y*
- Regression (回归): The target output is a real number
  - The price of a stock in 6 months time
  - The temperature at noon tomorrow
- Classification (分类): The target output is a class label
  - The decision of email spam
  - The class of a news document

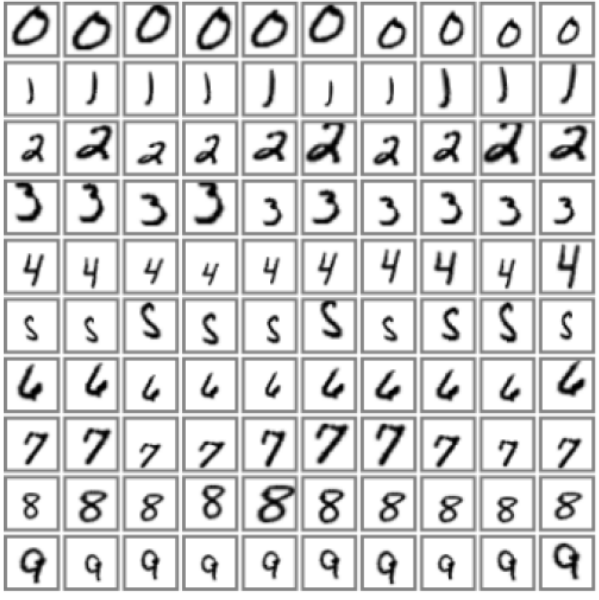


How Supervised Learning Typically Works

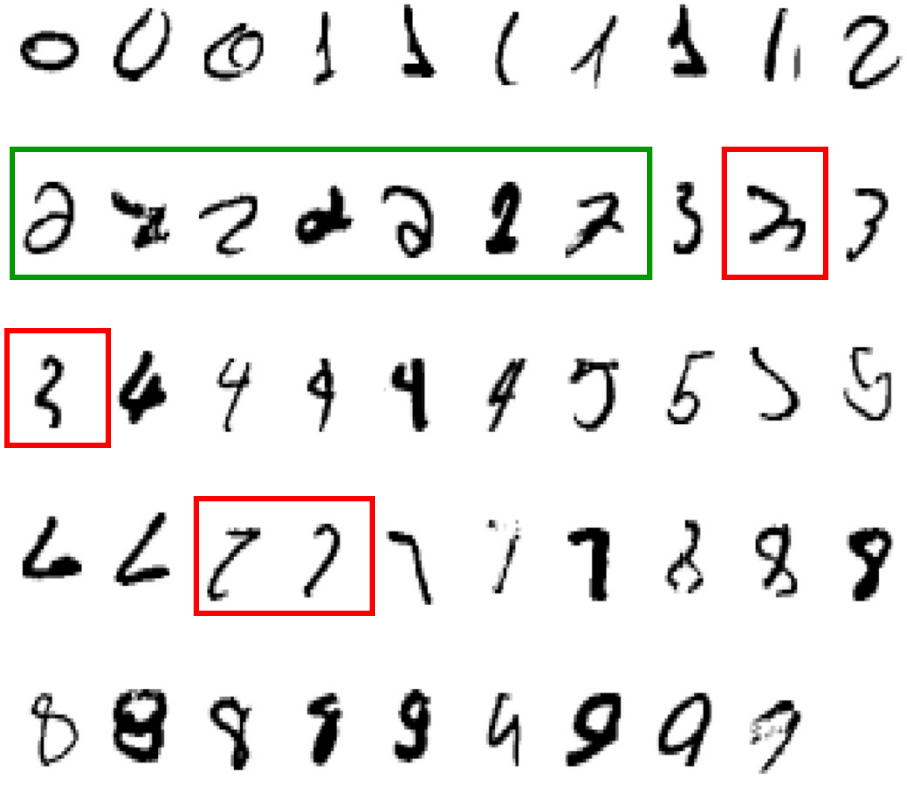
- Choosing a model-class: $y = f(\boldsymbol{x}; \boldsymbol{W})$
  - f is a way of using some numerical parameters (参数), $\boldsymbol{W}$, to map each input vector, $\boldsymbol{x}$, into a predicted output y
- Learning/optimizing (学习/优化)
  - Adjusting the parameters to reduce **loss/cost** (the discrepancy between the target output, y, on each training case and the actual output, y', produced by the model)

A Standard Example of ML

- The MNIST (Modified NIST) database of hand-written digits recognition
  - Publicly available
  - A huge amount about how well various ML methods do on it
  - 60,000 + 10,000 hand-written digits (28x28 pixels each)



Very hard to say what makes a 2

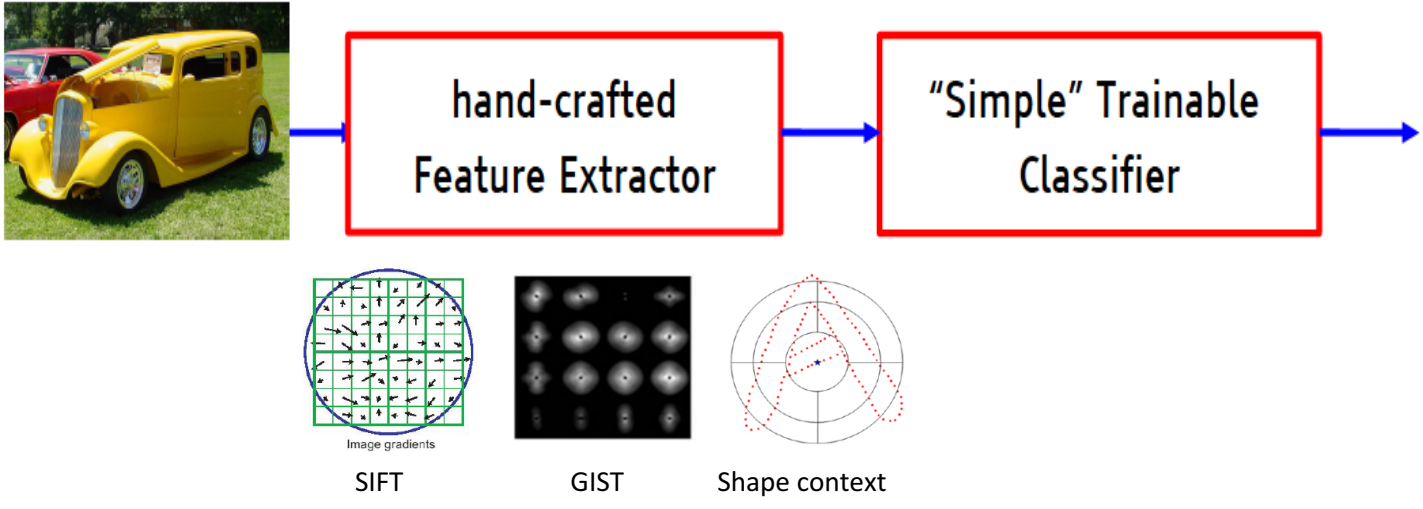


Beyond MNIST: ImageNet

- 1,000 object classes in 1.3 million images
- History of record
  - 2010: 47% error (top-1) and 25% error for (top-5)
  - 2012: Less that 40% error (top-1) and less than 20% (top-5) with a very deep neural net (Krizhevsky et. al. 2012)
  - 2015: Microsoft won 2015 LSVRC challenge
    - 152 layers (CNN), 19.38% error (top-1), 3.57% error (top-5) !!

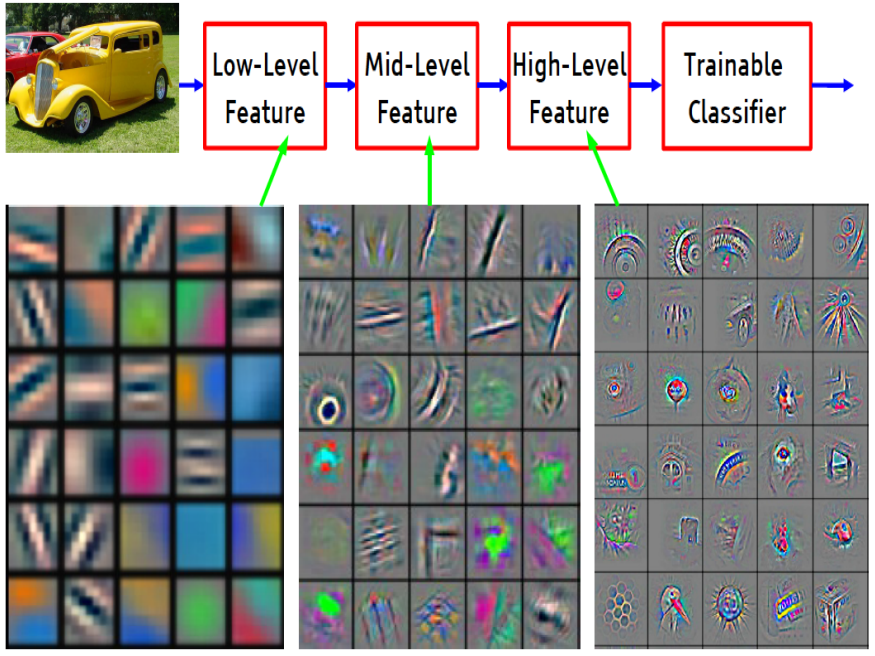
Traditional Model (before 2012)

- Fixed/engineered features + trainable classifier (分类器)
  - Designing a feature extractor requires considerable efforts by experts



Deep Learning (after 2012)

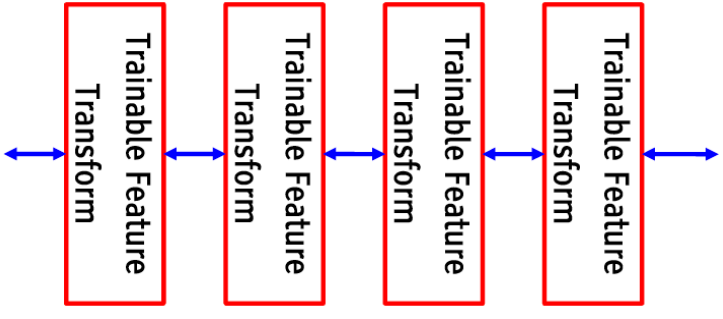
- Learning Hierarchical Representations
- DEEP means more than one stage of non-linear feature transformation



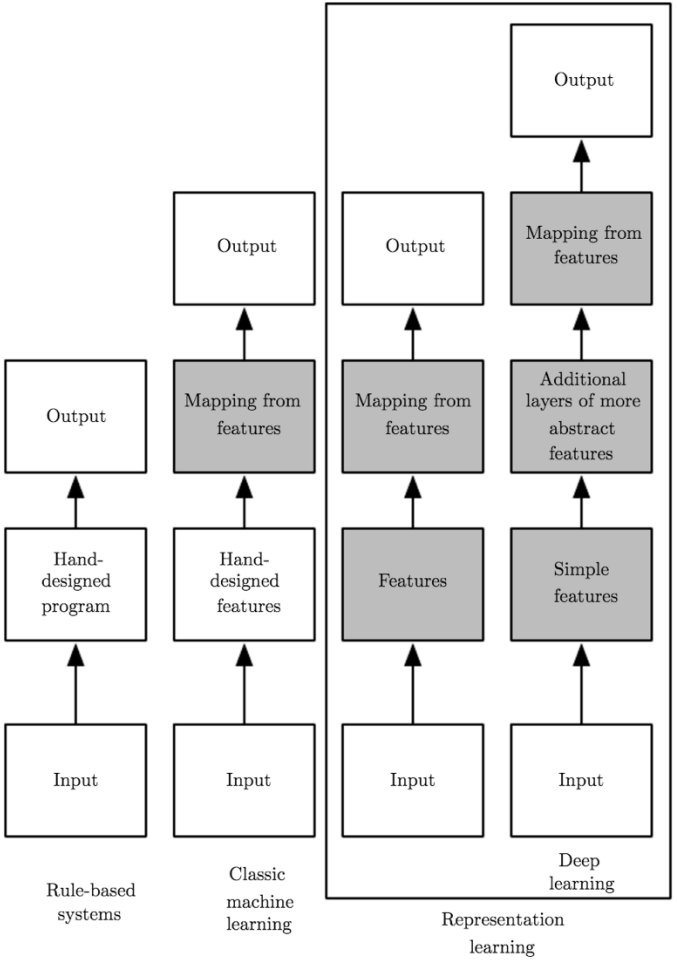
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Trainable Feature Hierarchy

- Hierarchy of representations with increasing level of abstraction
- Each stage is a kind of trainable feature transform (转换)
- Image recognition
  - Pixel → edge → texton → motif → part → object
- Text
  - Character → word → phrase → clause → sentence → document
- Speech
  - Sample → spectral band → sound → … → phone → phoneme → word



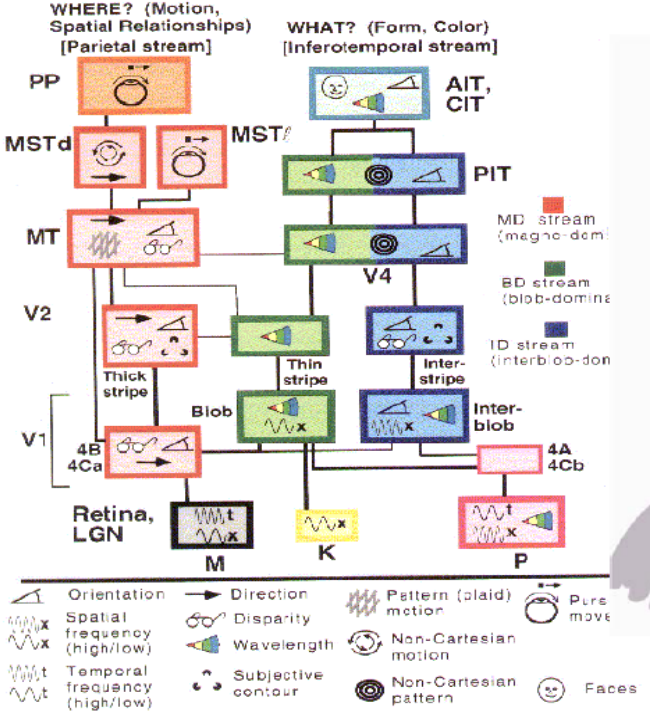
Different AI Disciplines

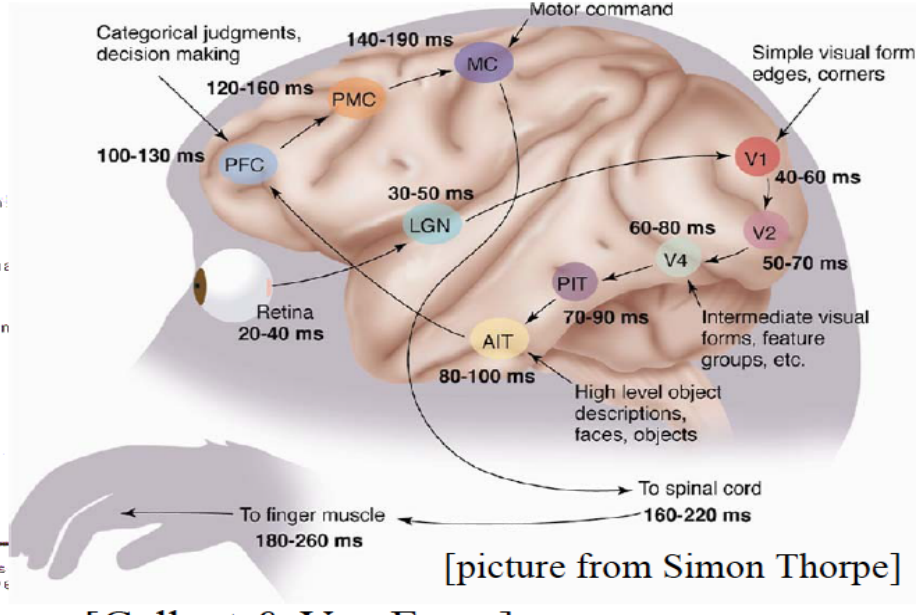


[Goodfellow et al. 2016]

The Mammalian Visual Cortex is Hierarchical

- The ventral (recognition) pathway in the visual cortex has multiple stages
- Retina - LGN - V1 - V2 - V4 - PIT - AIT ....
- Lots of intermediate representations

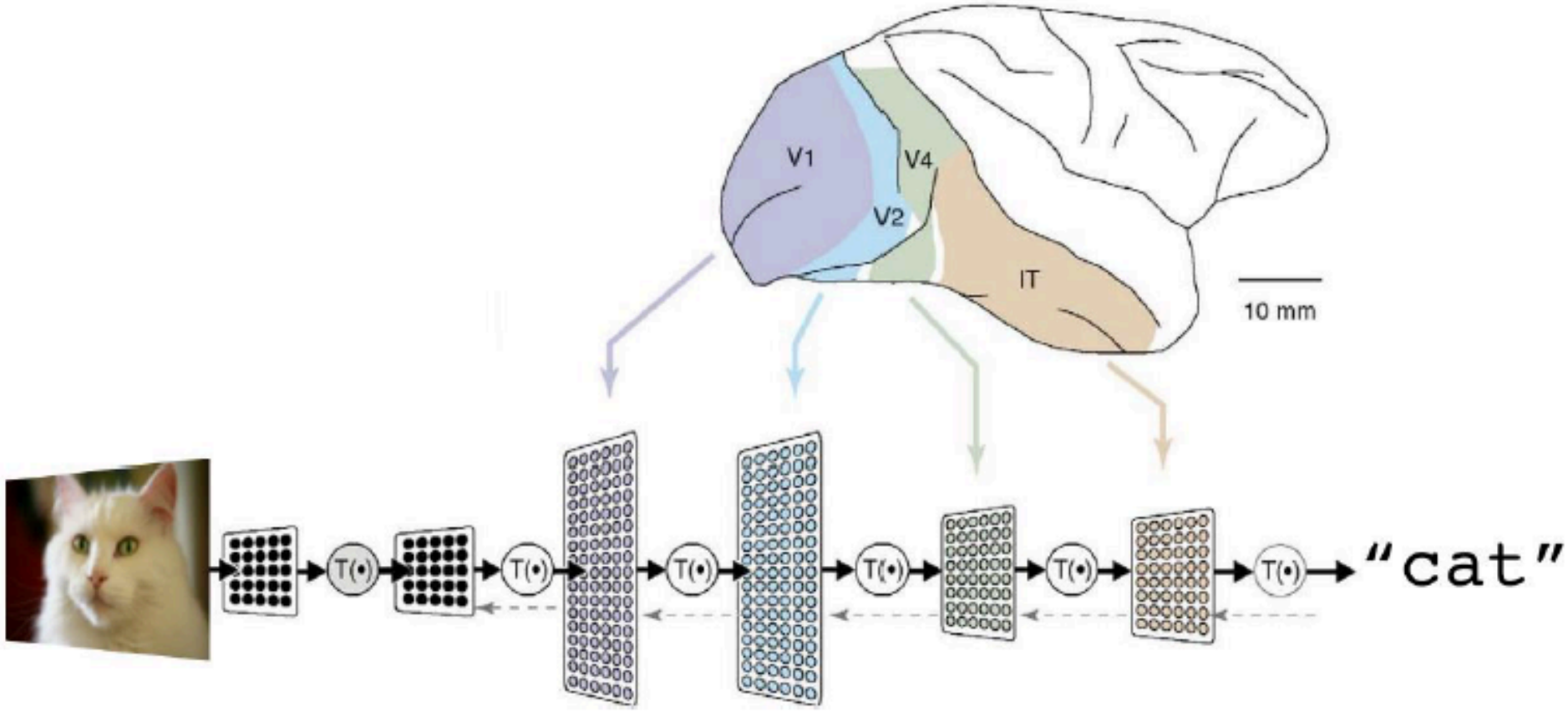




[picture from Simon Thorpe]

[Gallant & Van Essen]

Deep Learning Architecture



Deep Learning are Not New

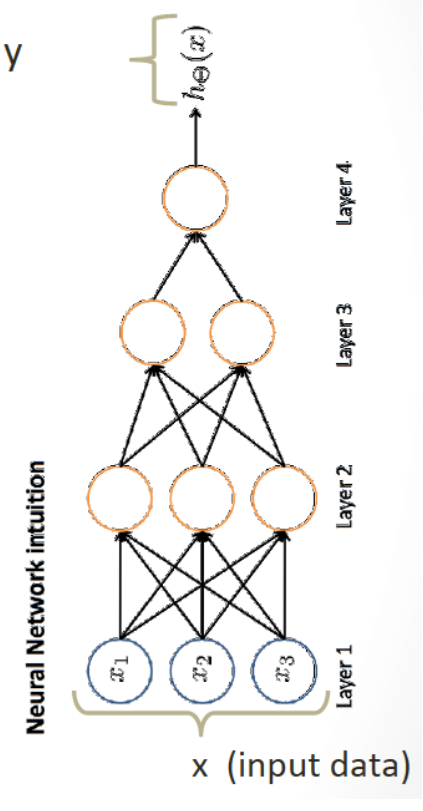
- 1980s technology (Neural Networks, 神经网络)

(label) y

Supervised learning

- Given x and y, learn p(y|x)
- Is this photo, x, a “cat”, y?

x =



About Neural Networks

- Pros
  - Simple to learn p(y|x)
  - Results OK for shallow nets
- Cons
  - Does not learn p(x)
  - Trouble with > 3 layers
  - Overfitts
  - Slow to train

The 2006 Breakthrough

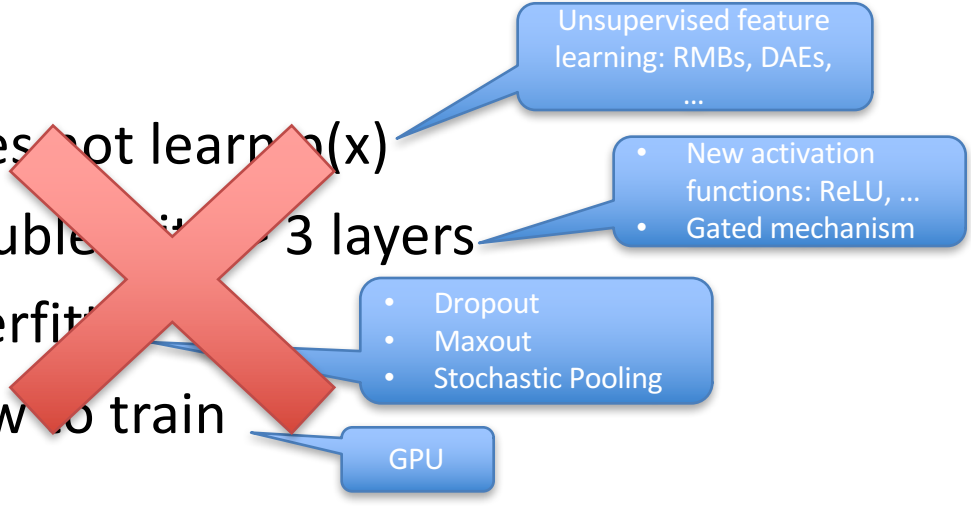
1. Mainstream: shallow nets (3 or fewer layers) on small data
2. Slow training on CPUs
3. Near universal sigmoid neuron nonlinearities
4. Parameters initialized with random weights
5. Could only learn discriminative p(y|x) no generative p(x)
6. Neural Nets: Yet another machine learning algorithm (yamla)
7. Some convolutional networks (LeCun et al.)

Hinton et al.'s RBMs

1. Mainstream: deep nets (6+ layers) on "Big Data"
2. Fast training on GPUs
3. Autoencoder/RBMs to learn generative p(x)
4. Parameters initialized with generative model
5. Rise of the ReLU nonlinearity
6. Dropout prevents overfitting
7. Deep nets outcompetes the best SOTA in the world (Image or speech recognition)
8. Deep learning moves out of academic to Google, Facebook, and so on

Deep Learning beats NN

- Pros
  - Simple to learn p(y|x)
  - Results good for shallow nets
- Cons
  - Does not learn p(x)
  - Trouble with > 3 layers
  - Overfitting
  - Slow to train



Deep Learning - History

1958 Perceptron

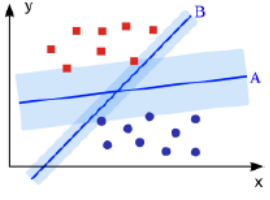
1969
Perceptron criticized

1974 Backpropagation

awkward silence (AI Winter)

Convolution Neural Networks for Handwritten Recognition
1998

1995
SVM reigns



2006
Restricted Boltzmann Machine

Google Brain Project on 16k Cores
2012

2012
AlexNet wins ImageNet
IMAGENET

Deep Learning Big Guys

Yann LeCun
NYU & FB

Geoff Hinton
U. Toronto & Google

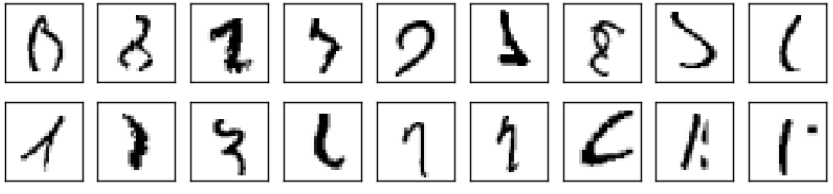
Yoshua Bengio
University of Montreal

Andrew Ng
Stanford & Baidu

Jürgen Schmidhuber
Swiss AI Lab & NNAISENSE

Results on MNIST

- Naïve Neural Network
  - 96.59%
- SVM (default settings for libsvm)
  - 94.35%
- Optimal SVM [Andreas Mueller]
  - 98.56%
- The state of the art: Convolutional NN (2013)
  - 99.79%

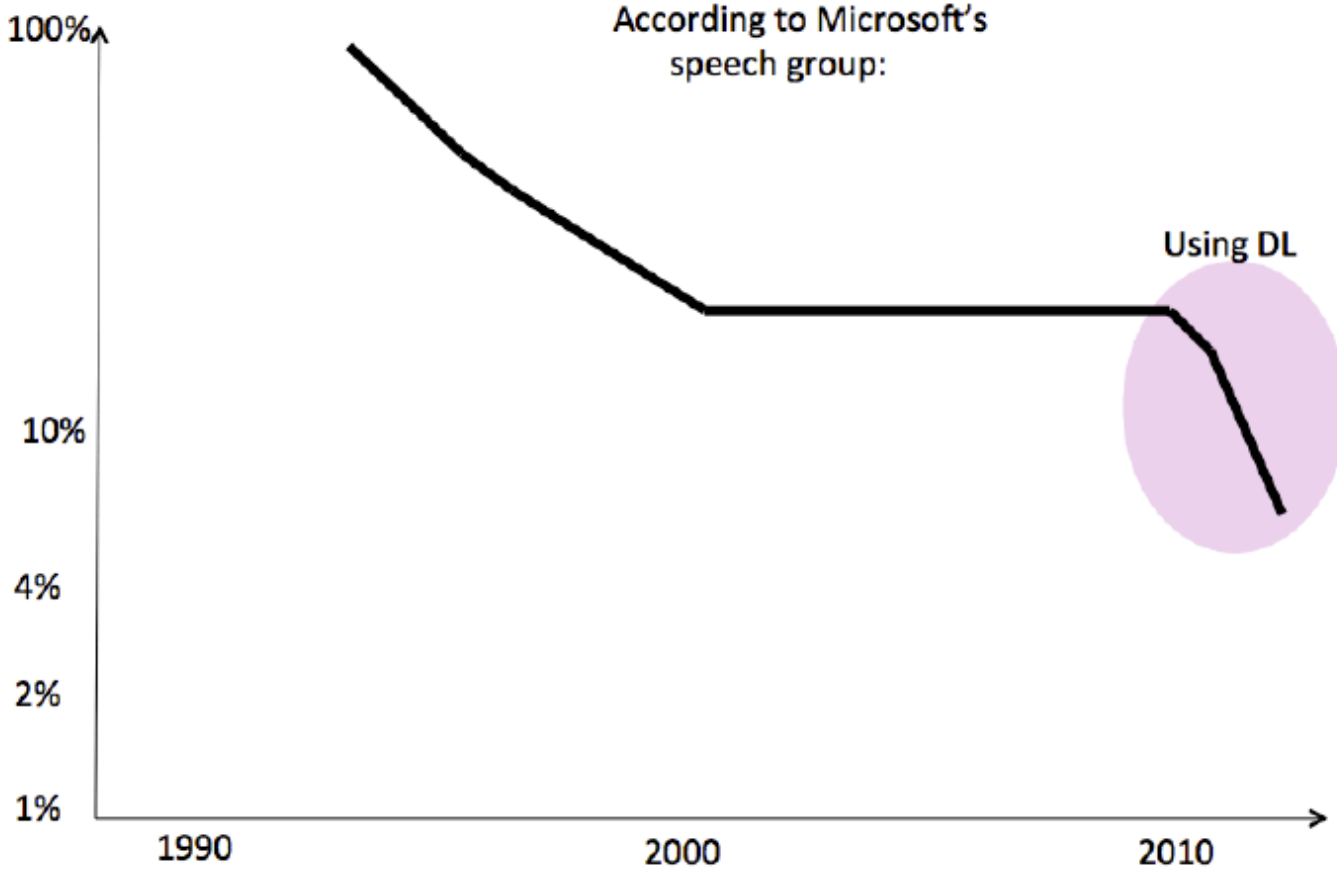


Deep Learning Wins

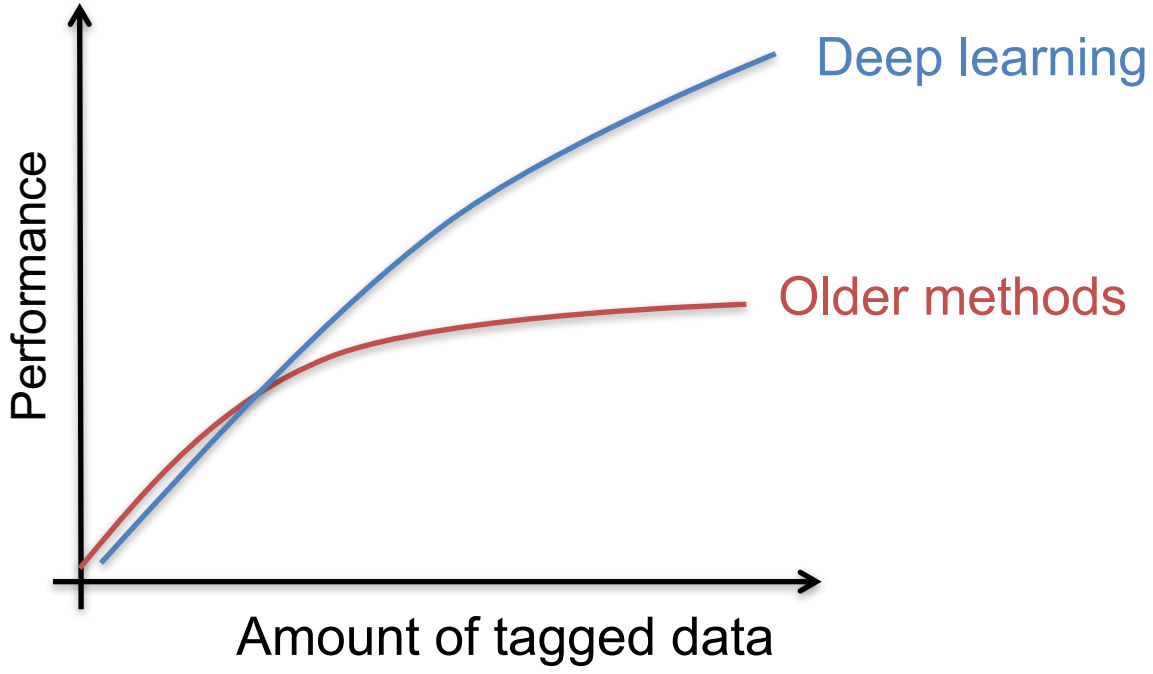
9. MICCAI 2013 Grand Challenge on Mitosis Detection
8. ICPR 2012 Contest on Mitosis Detection in Breast Cancer Histological Images
7. ISBI 2012 Brain Image Segmentation Challenge (with superhuman pixel error rate)
6. IJCNN 2011 Traffic Sign Recognition Competition (only our method achieved superhuman results)
5. ICDAR 2011 offline Chinese Handwriting Competition
4. Online German Traffic Sign Recognition Contest
3. ICDAR 2009 Arabic Connected Handwriting Competition
2. ICDAR 2009 Handwritten Farsi/Arabic Character Recognition Competition
1. ICDAR 2009 French Connected Handwriting Competition. Compare the overview page on handwriting recognition.

- http://people.idsia.ch/~juergen/deeplearning.html

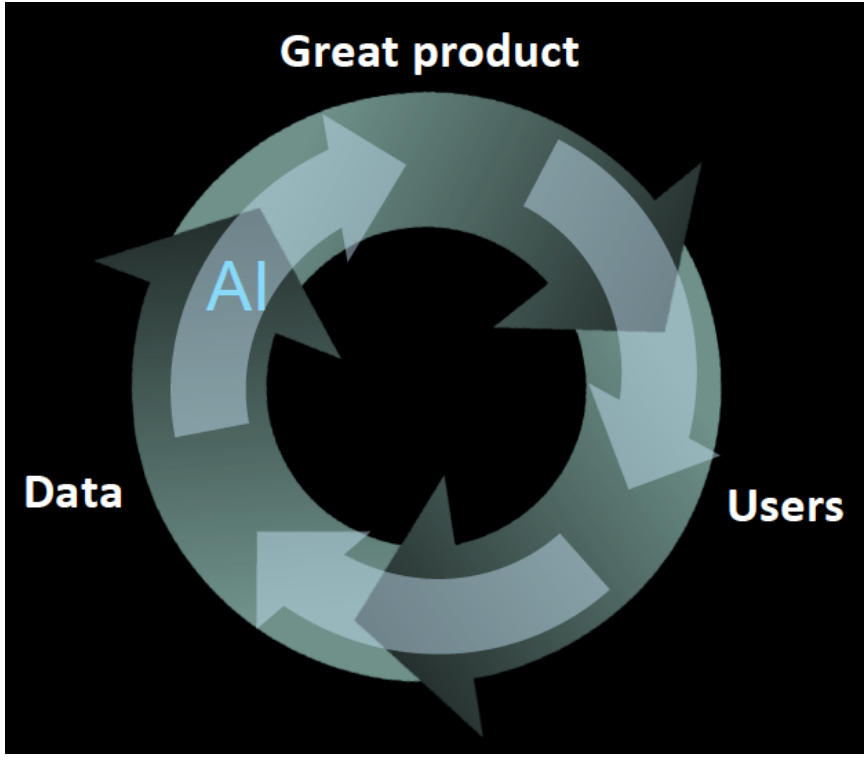
Deep Learning for Speech Recognition



Power of deep learning



Andrew Ng: Virtuous circle of AI



Speech Recognition

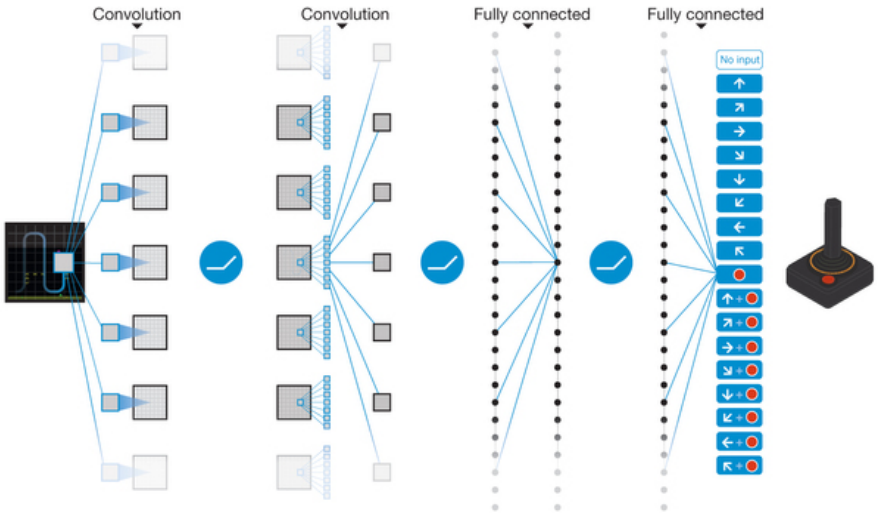
- Microsoft's Real-time Speech Translation

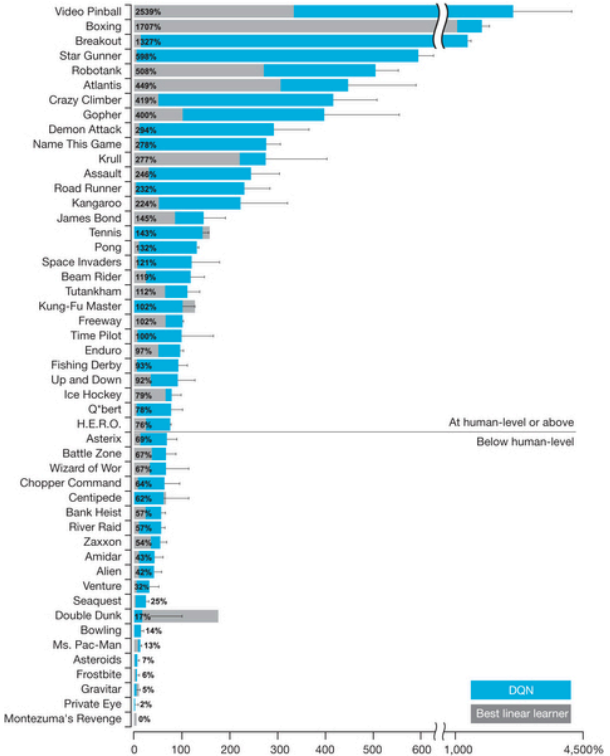
Rick Rashid in Tianjin, China, October, 25, 2012

Skype real-time translator

Google DeepMind

- Human-level control through deep reinforcement learning. Nature 518, 529–533 (26 February 2015)
  - 23/43/49





AlphaGo

AlphaGo

MIT Technology Review

10 BREAKTHROUGH TECHNOLOGIES 2013

Introduction | The 10 Technologies | Past Years

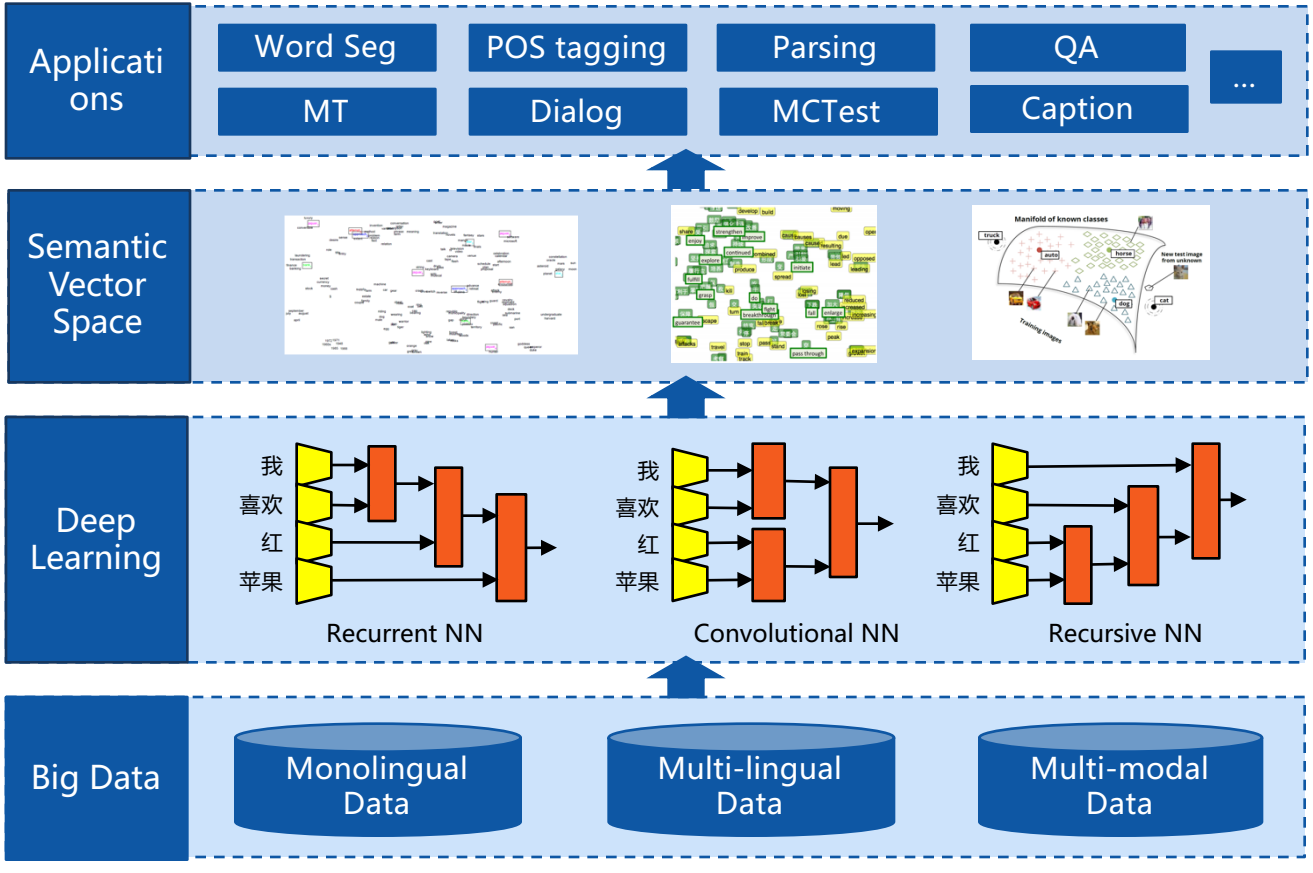
Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.

- http://www.technologyreview.com/featuredstory/513696/deep-learning/

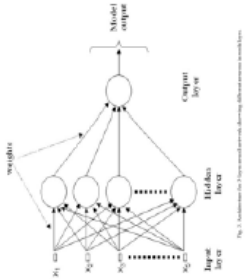
Big Players

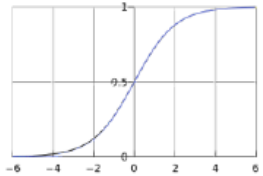
Deep Learning for NLP



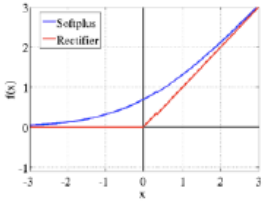
Past | Present

p(y|x), yamla





Tan-1, logistic

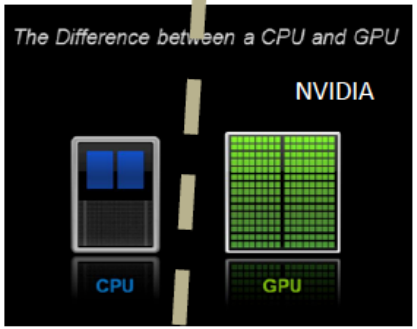


ReLU, maxout

Data

Labels

Backprop, feature engineering



Backprop, dropout, SGD, aSGD, autoencoders, wake-sleep, curriculum learning, ?...

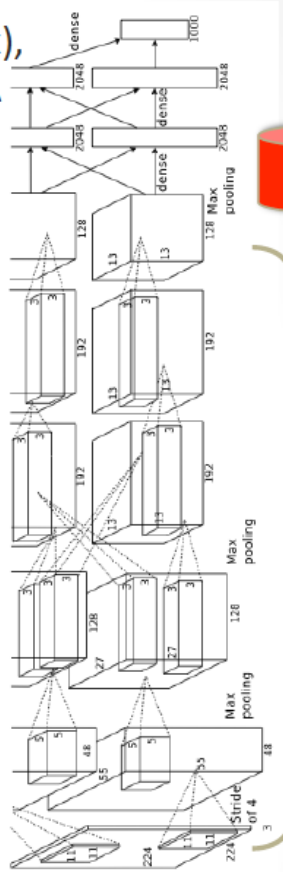
C, Matlab

Caruana-like academic evaluations

Kaggle

CUDA, Torch, Cudamat, Theano, pylearn2

p(y|x), SOTA

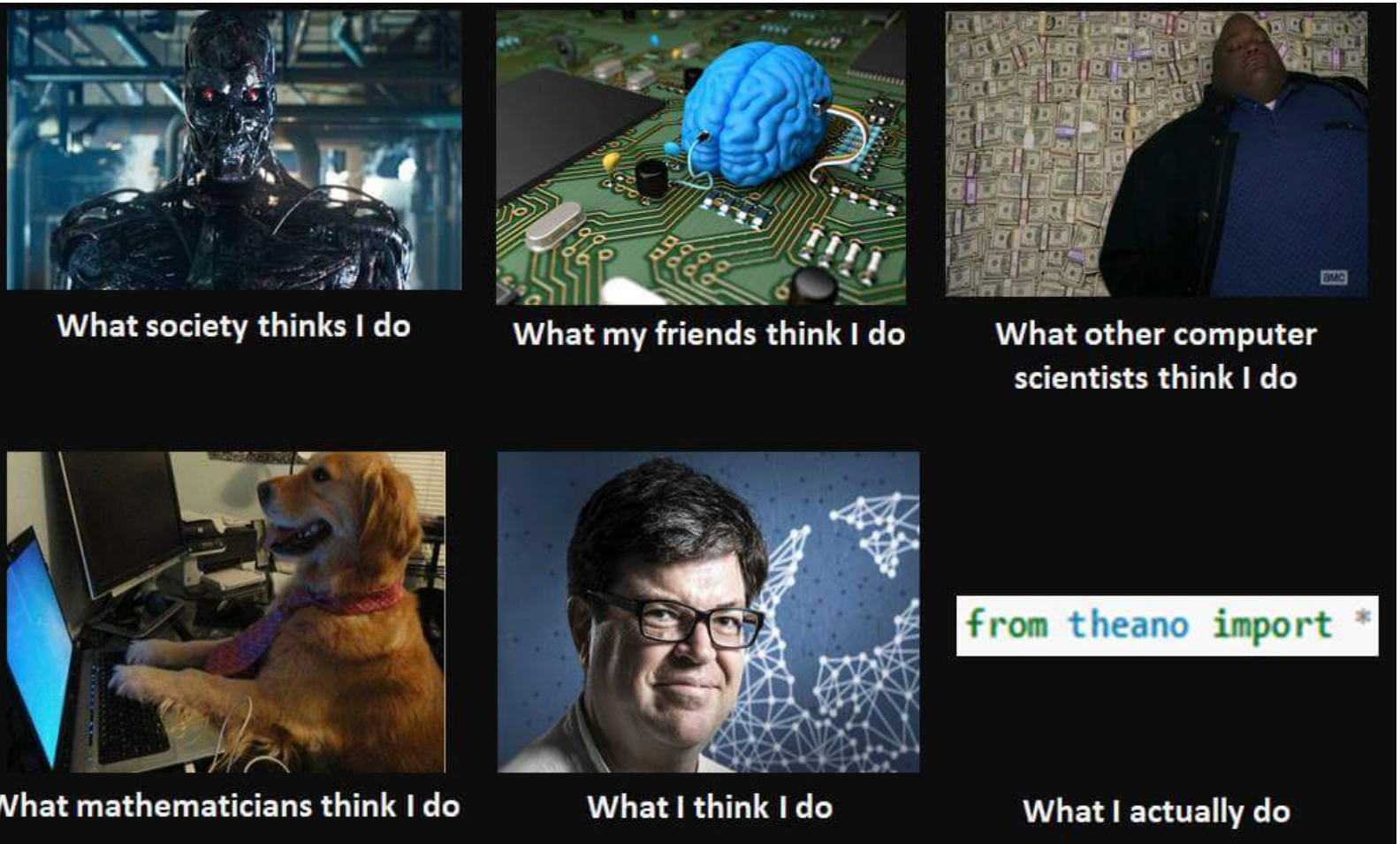


Labels

p(x)

Data VVVV

What am I doing?



Textbooks

- Ian Goodfellow, Yoshua Bengio and Aaron Courville. Deep Learning. MIT Press. 2016
  - http://www.iro.umontreal.ca/~bengioy/dlbook/
- Michael Nielsen. Neural Networks and Deep Learning
  - http://neuralnetworksanddeeplearning.com
- Stanford Courses
  - CS224d: Deep Learning for Natural Language Processing
    - http://cs224d.stanford.edu/
  - CS231n: Convolutional Neural Networks for Visual Recognition
    - http://cs231n.stanford.edu/
  - Andrew Ng. Unsupervised Feature Learning and Deep Learning
    - http://ufldl.stanford.edu/tutorial/