Deep Learning and Business Models

Tran Quoc Hoan
The University of Tokyo
VNITC@2015-09-13

Today agendas

- Deep Learning boom
- Essentials of Deep Learning
- Deep Learning - the newest
- Deep Learning and Business Models

What is Deep Learning?

Deep Learning is a new trend of machine learning that enables machines to unravel high level abstraction in large amount of data

Image recognition

Speech recognition

Customer Centric Management

Natural Language Processing

Drug Discovery & Toxicology

Deep Learning Boom

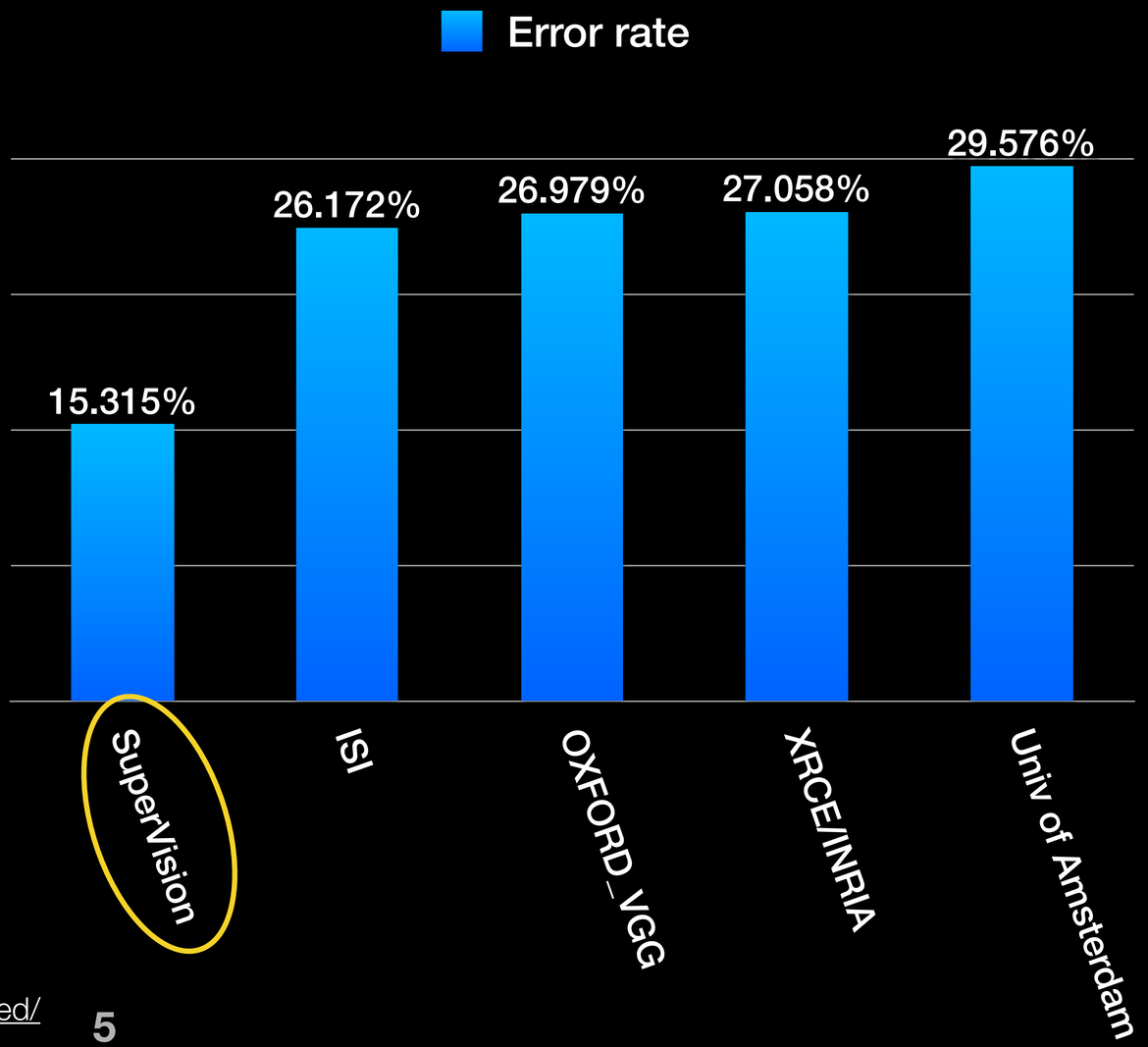
“If we knew what it was we were doing, it would not be called research, would it?”

–Albert Einstein

Big News - ILSVRC 2012

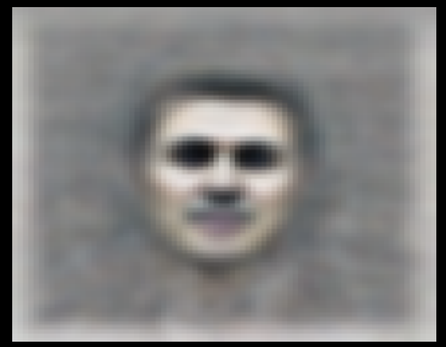
- ILSVRC 2012 (ImageNet Large Scale Visual Recognition)

Image source: http://cs.stanford.edu/people/karpathy/cnnembed/



Big News - Google brain

- **Self-taught** learning with **unlabelled** youtube videos and 16,000 computers



(Human)

(Cat)

AI grandmother cell (2012)

Data and Machine Learning

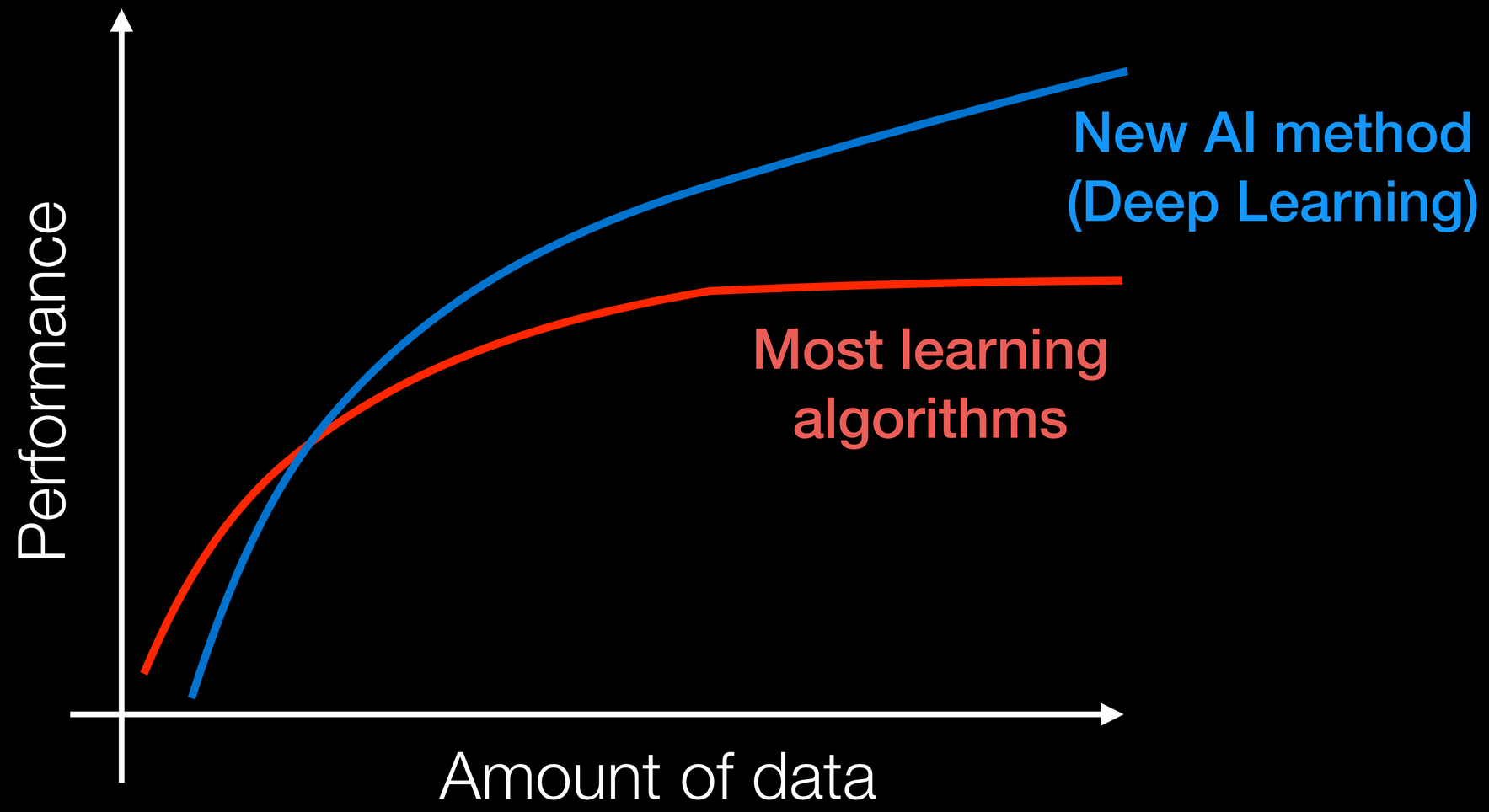


Image source: http://cs229.stanford.edu/materials/CS229-DeepLearning.pdf

New boom for the Giants & Startups

One of 10 breakthrough technologies 2013

(MIT Technology Review)

Google brain project (2012)

Facebook AI Research Lab (2013 Dec.)

Project Adam (2014)

Enlitic

Deepmind

Nervana Systems

Ersatz Labs

Skymind

MetaMind

PFN (Japan)

Pioneers

Theory + algorithms

Geoffrey Hinton
(Toronto, Google)

Yann LeCun
(NewYork, Facebook)

Yoshua Bengio
(Montreal)

Implementation

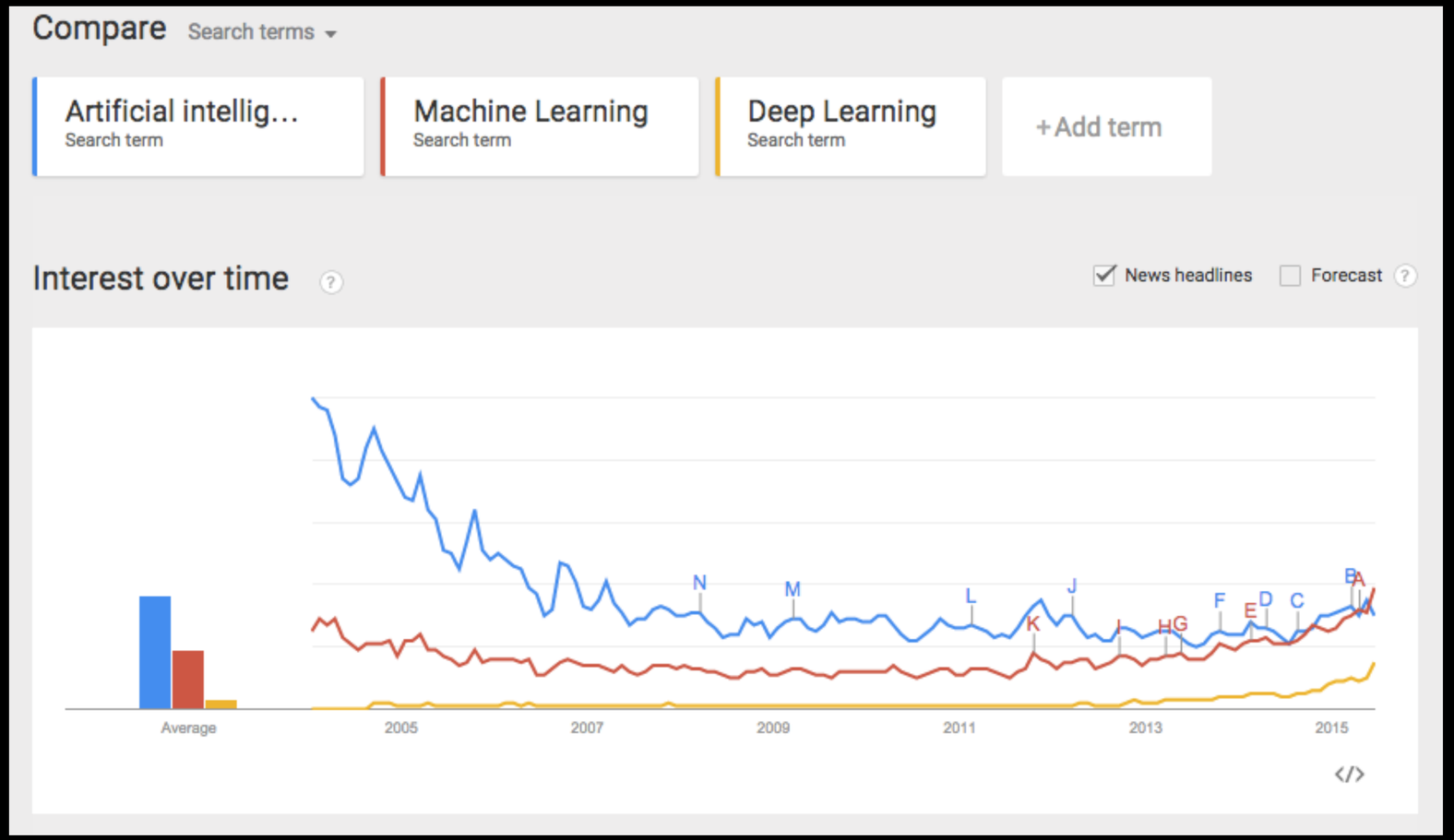
Andrew Ng
(Stanford, Baidu)

Jeffrey Dean
(Google)

Le Viet Quoc
(Stanford, Google)

Sub-summary

- More than 2000 papers in 2014 ~ 2015 with 47 Google services.



Essentials of Deep Learning

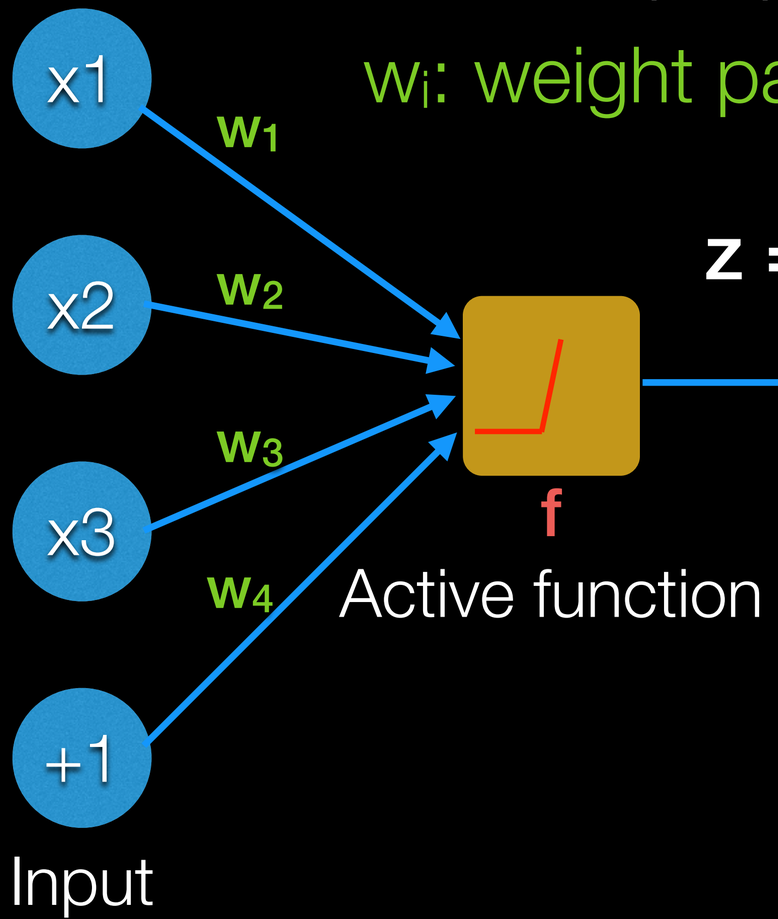
“If we knew what it was we were doing, it would not be called research, would it?”

–Albert Einstein

The basic calculation

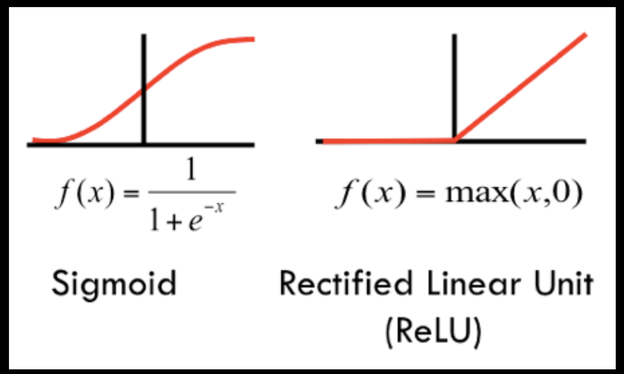
Forward propagation

w_i: weight parameters

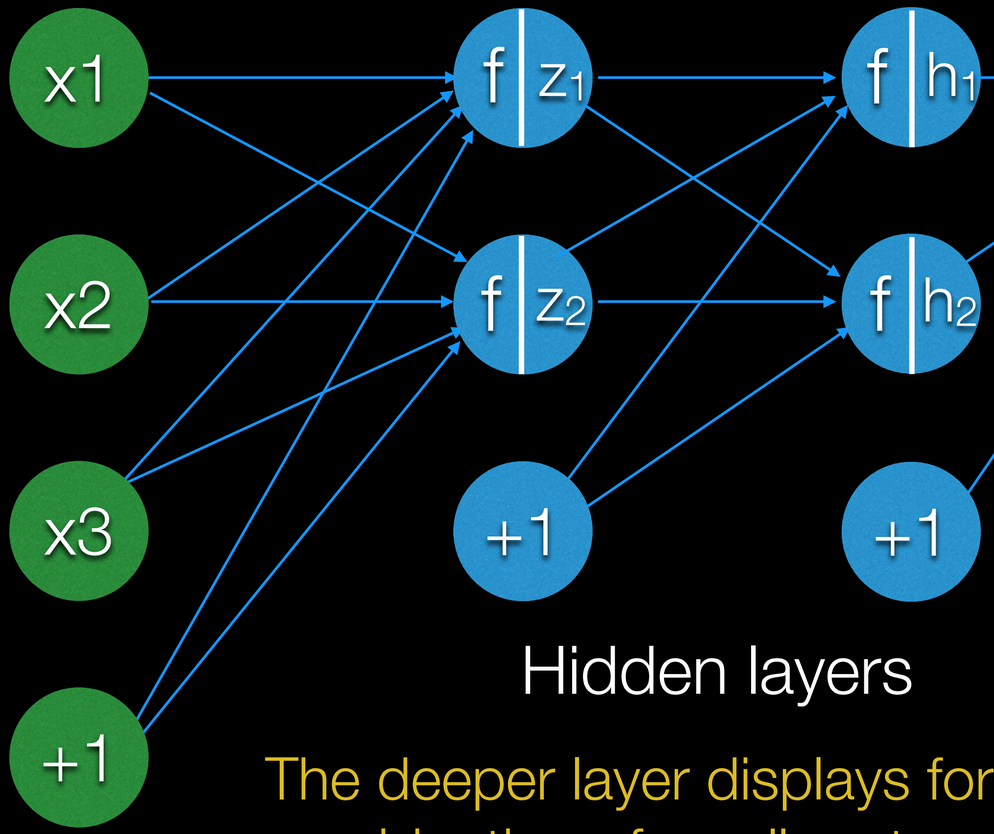


$$z = f(x_1w_1 + x_2w_2 + x_3w_3 + w_4)$$

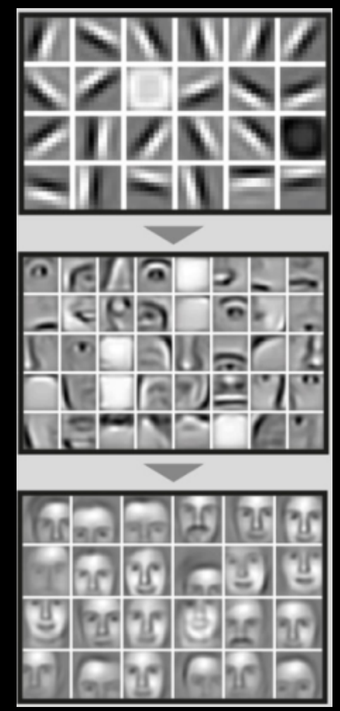
Active function



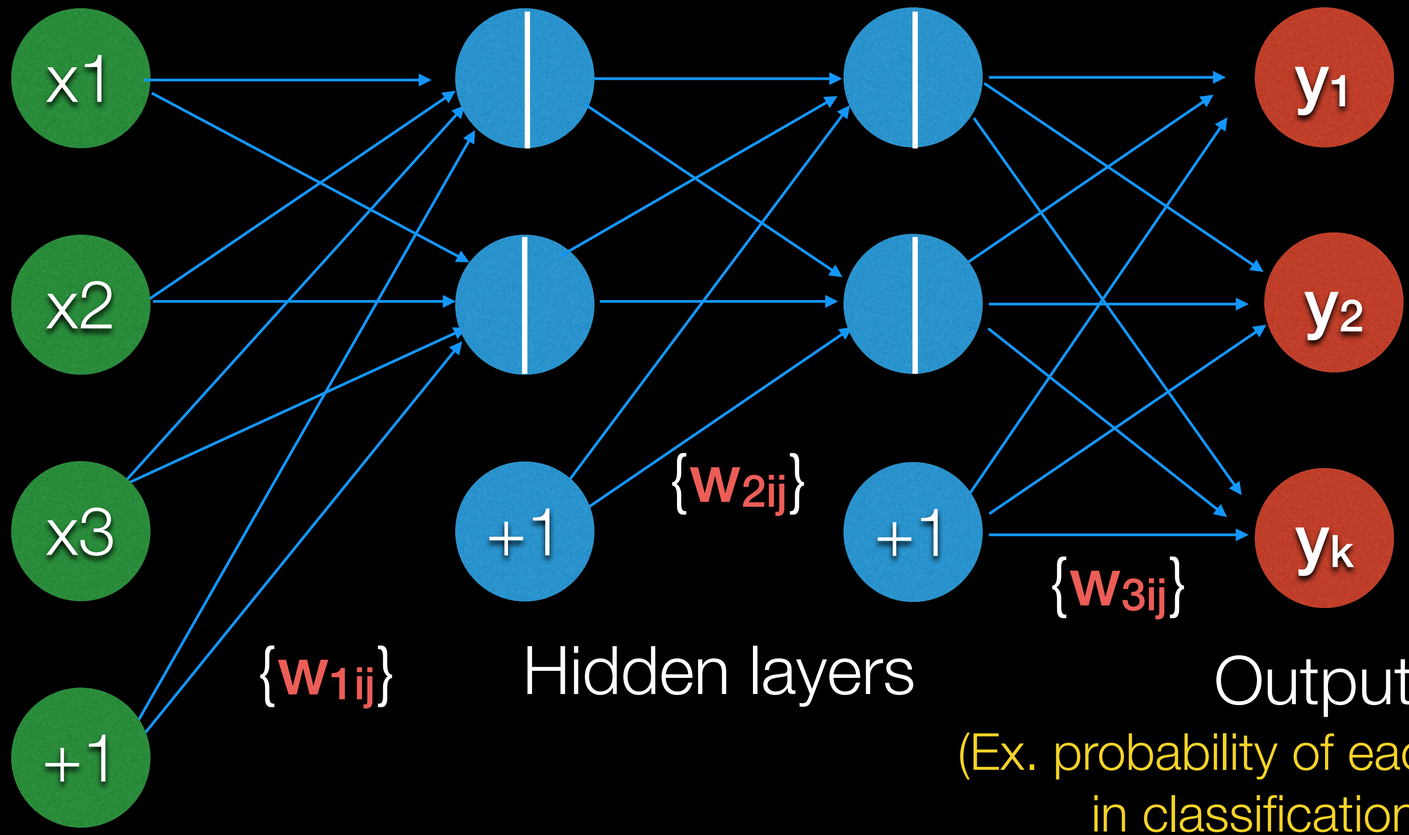
Neural network (1/2)



The deeper layer displays for comprehensive combination of small parts



Neural network (2/2)



(Ex. probability of each in k class in classification task)

Find parameters {**w**} to minimize **E** → How?

Supervisor learning

- Loss function E measures how output labels are similar to true labels.
- For N input data

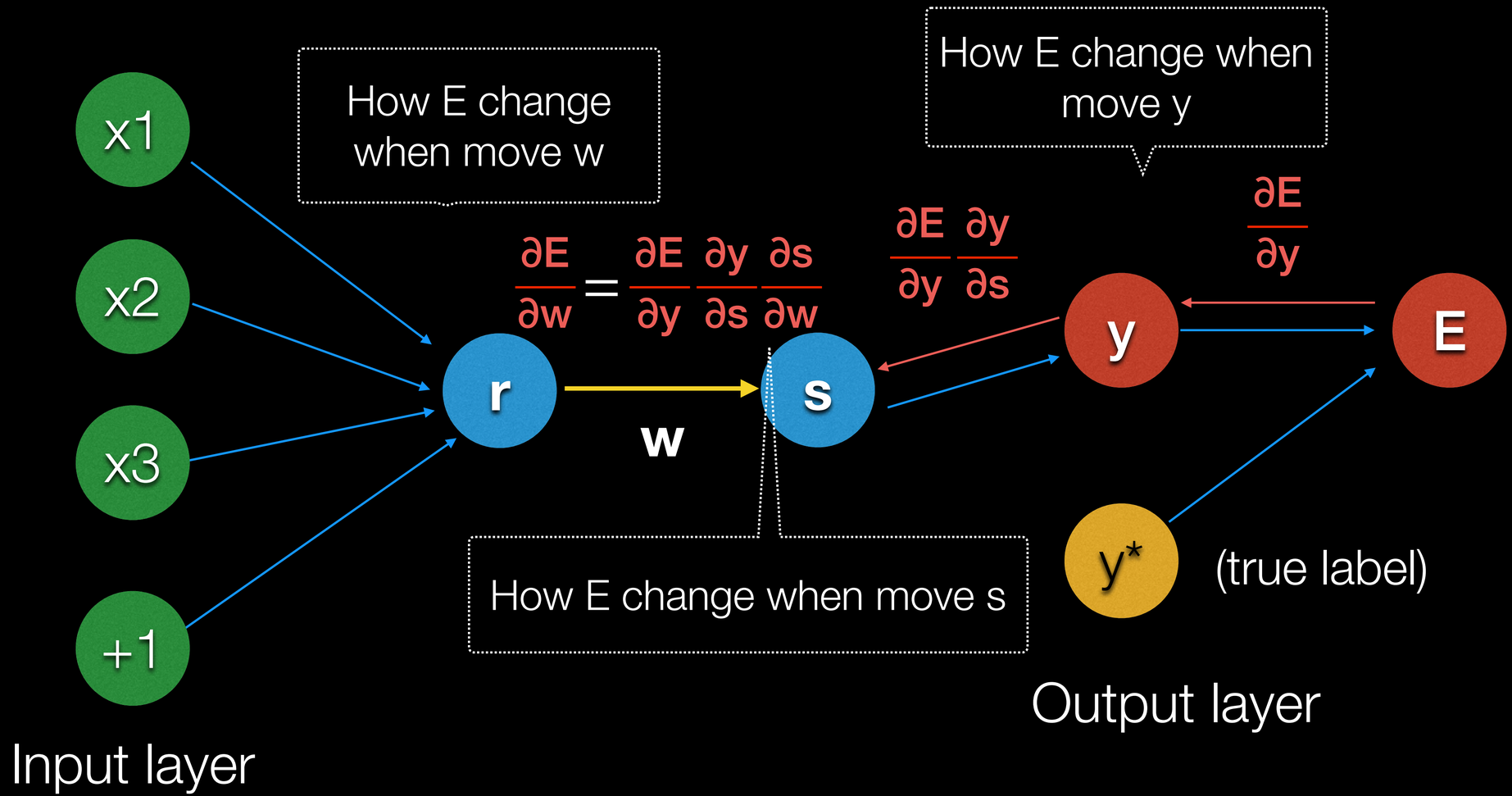
$$E = \sum_{n=1}^{N} E_n$$

$$E_n = \frac{1}{2} \sum_{k} (y_{nk} - t_{nk})^2$$

$$E_n = -\sum_{k} t_{nk} \ln y_{nk}$$

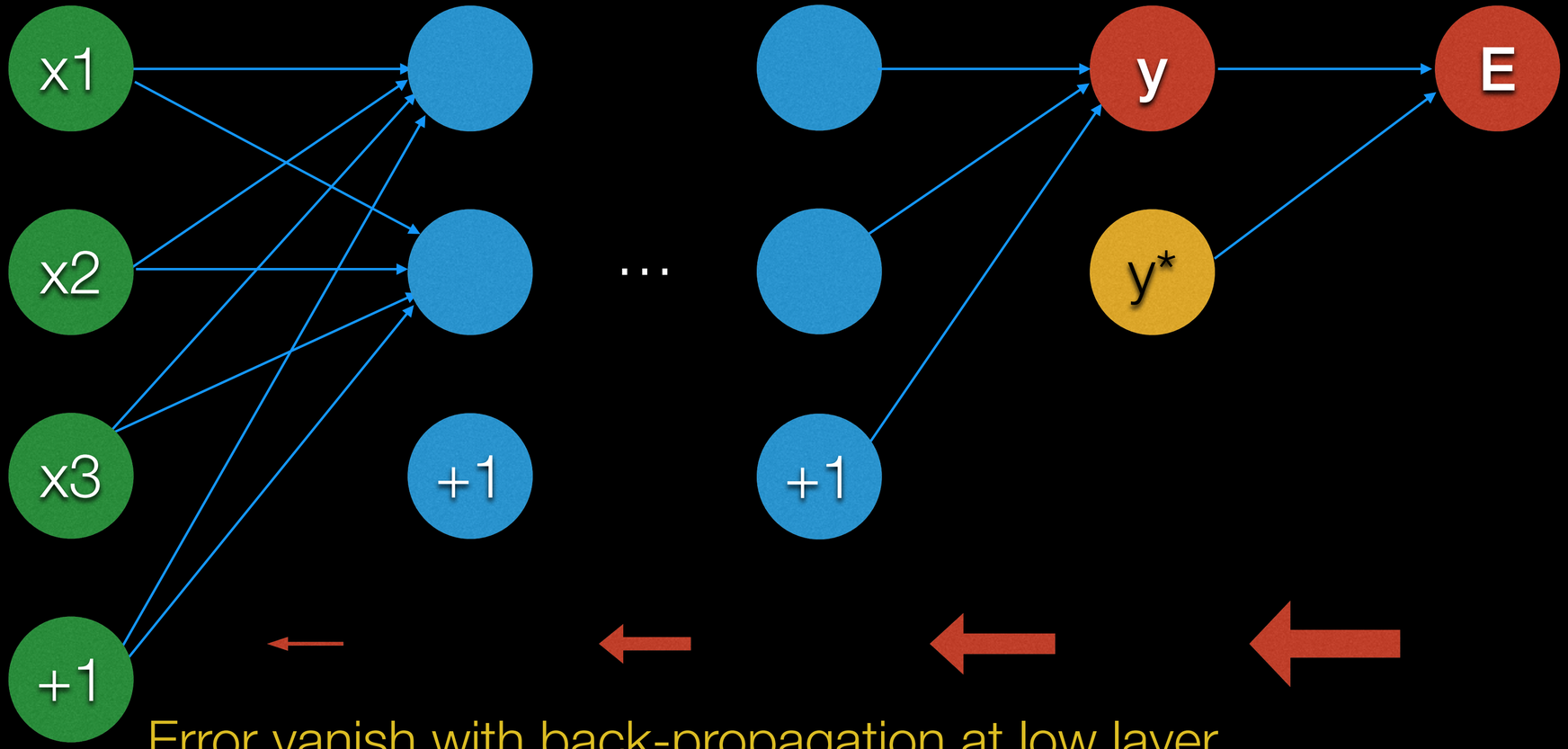
Backward propagation

Backward from loss to update parameters



Vanishing gradient problem

- Neural network architecture reached the limit before deep learning boom

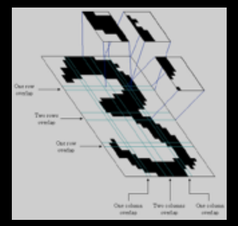


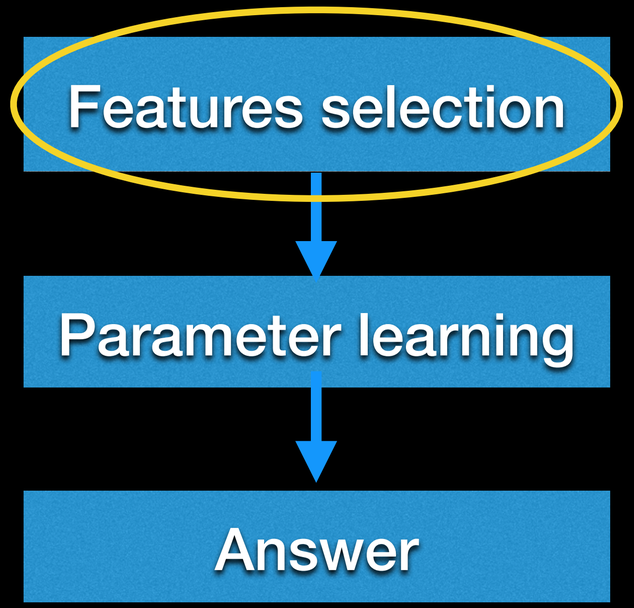
Error vanish with back-propagation at low layer

Features representation

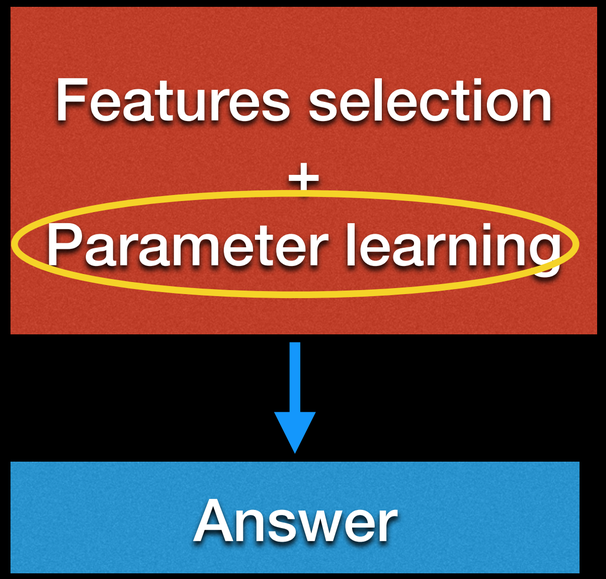
- Deep learning reduced the human's time-consuming of features selection process in classification task

Previous methods



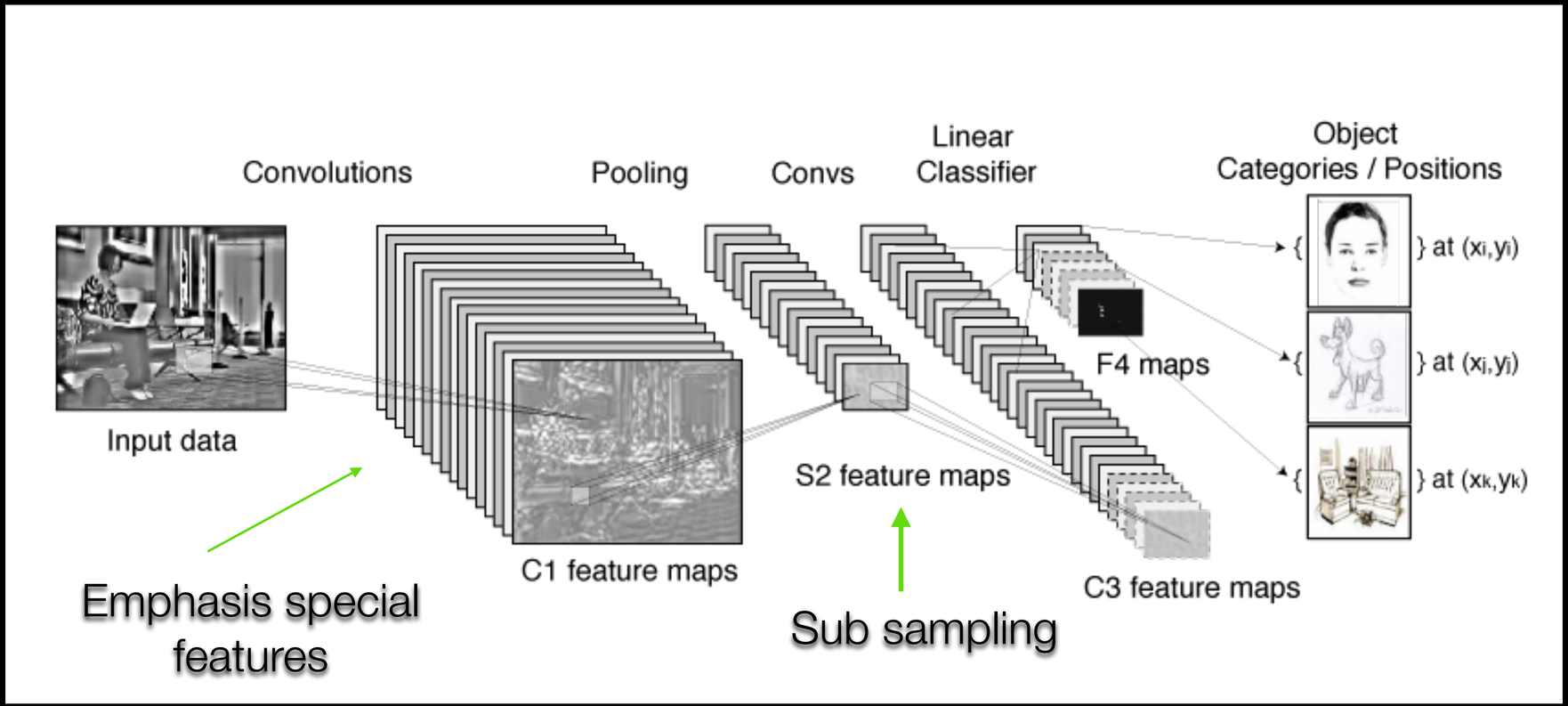


Deep Learning



Convolutional Neural Network

The most popular kind of deep learning model



ConvNet diagram from Torch Tutorial

How to train parameters

- Trainable parameters
  - Values in each kernel of convolutional layer
  - Weigh values and bias in fully connection layer

Stochastic Gradient Descent (SGD) (others: AdaGrad, Adam,…)

Parameters update

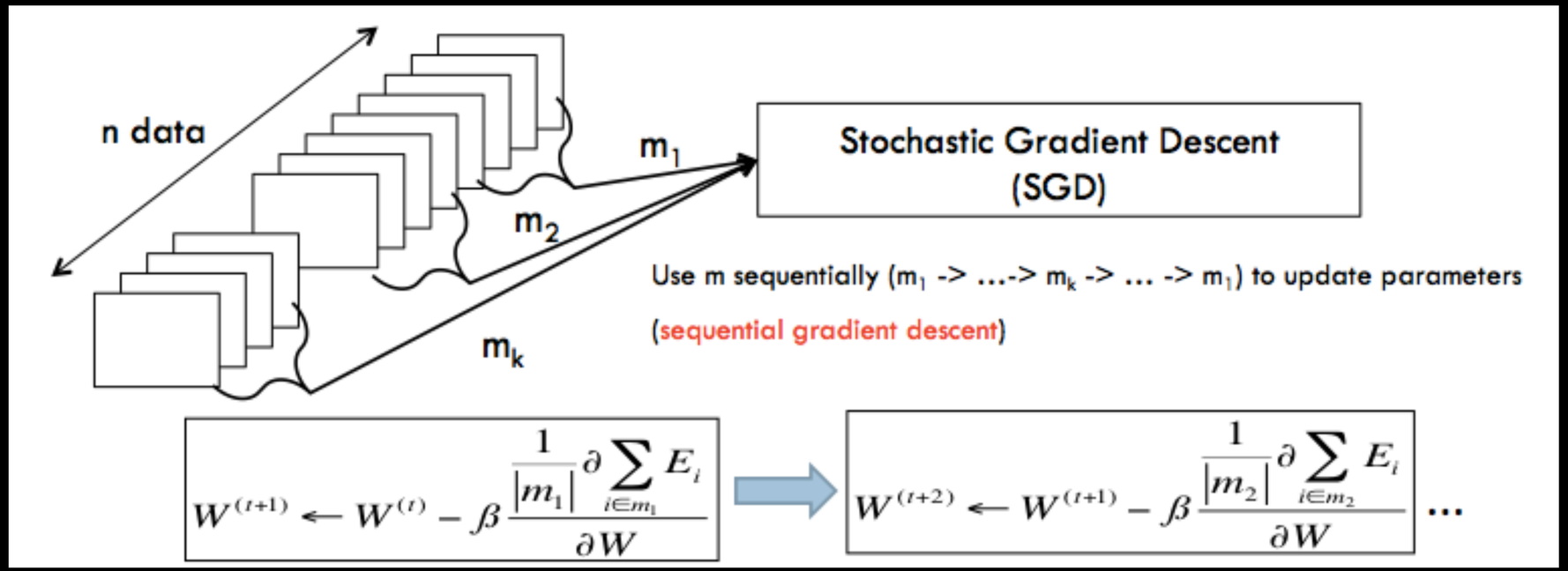
Backward propagation

$$W \leftarrow W - \beta \frac{\partial E}{\partial W}$$

Learning rate

SGD - Mini batch

- Batch learning: update all samples in each update (over-fitting)
- Mini batch: update parameters with some samples



Training trick

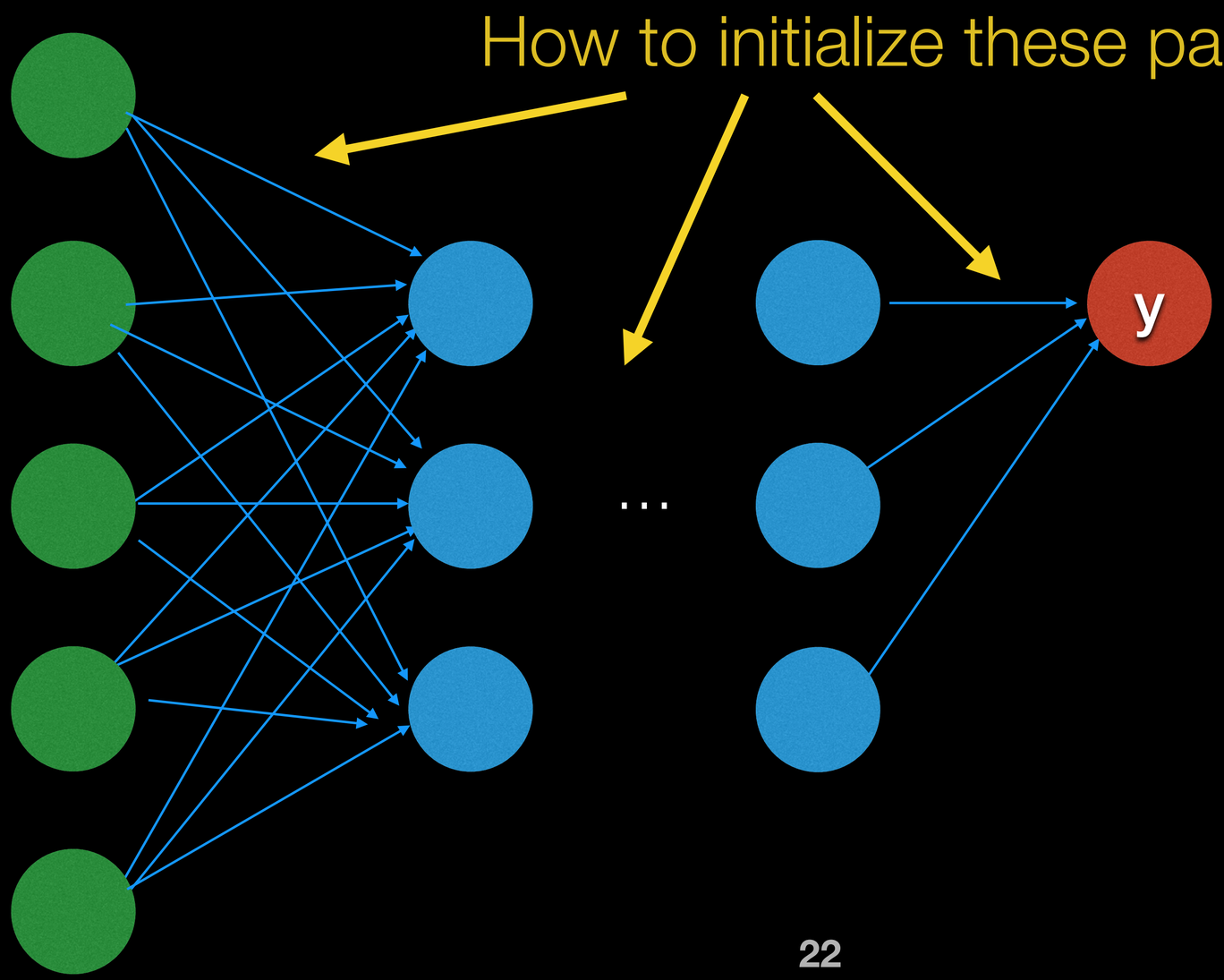
- Data normalization
- Reduce learning rate after some iterations
- Momentum SGD

$$\Delta W^{(t-1)} = W^{(t-1)} - W^{(t-2)}$$

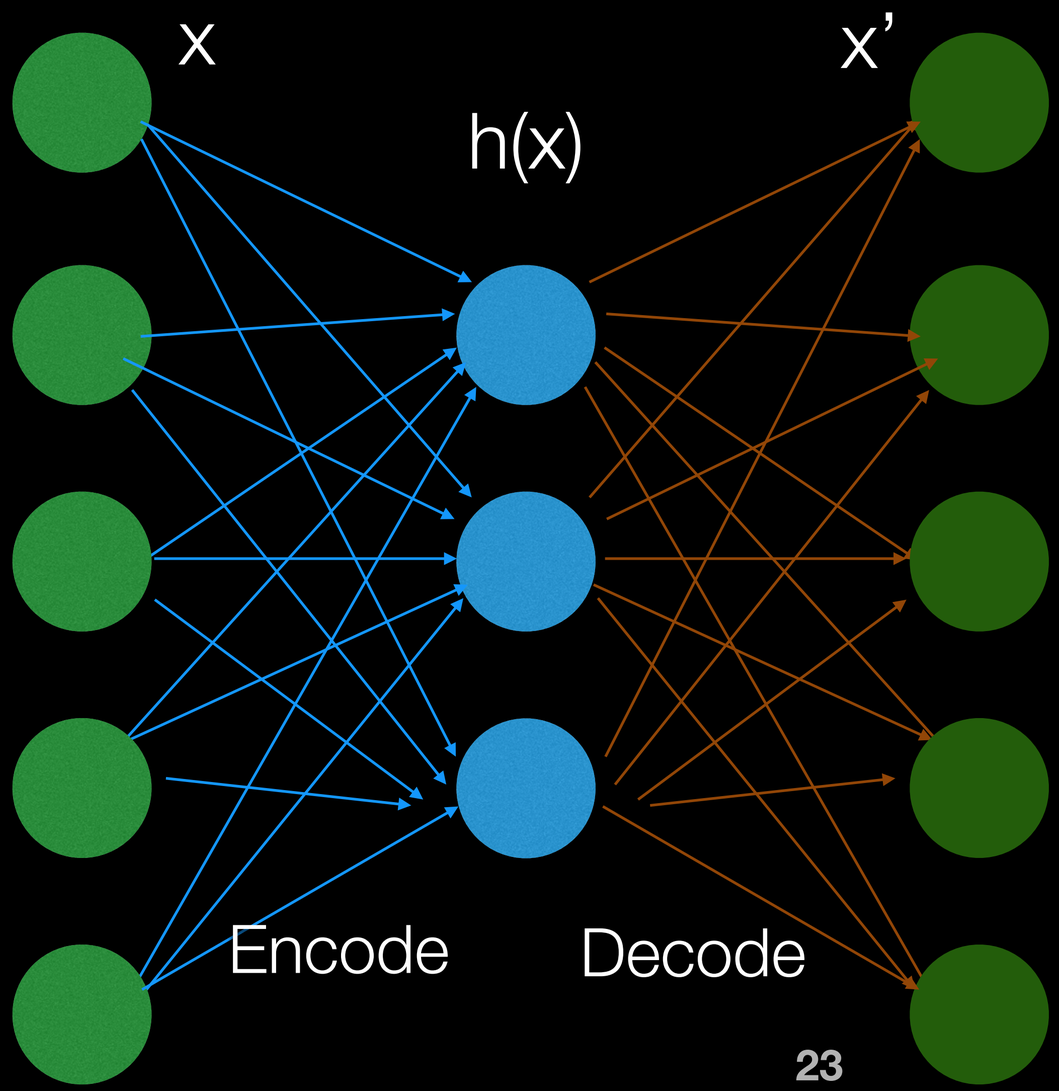
$$W^{(t+1)} \leftarrow W^{(t)} - \beta \nabla E_t + \mu \Delta W^{(t-1)}$$

$$\Delta W^{(t)} \leftarrow \mu \Delta W^{(t-1)} - \beta \nabla E_t$$

Auto-encoder (1/2)



Auto-encoder (2/2)



The ability of reconstruct input makes a good initialization for parameters

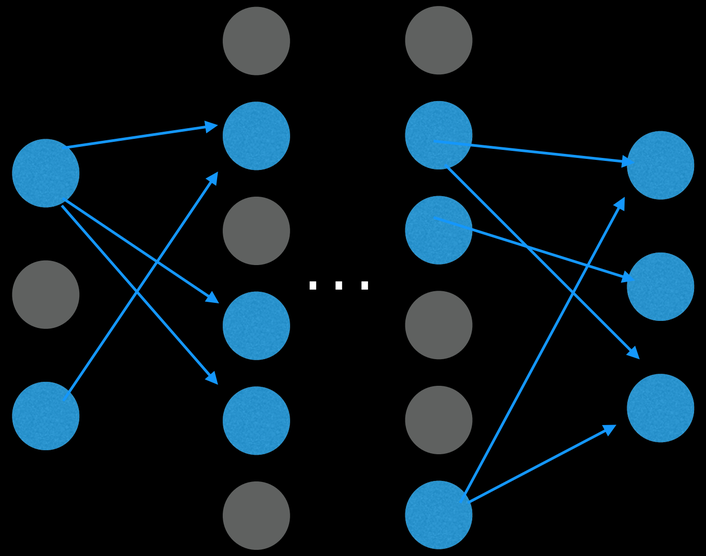
Reconstruct error

$$E = \|x - x'\|$$

Robust algorithms

- Data augmentation: more noisy data → more robust model

- Nodes in network don't need to active all (mimic human brain) → Dropout concept



Deep Learning - The Newest

“If we knew what it was we were doing, it would not be called research, would it?”

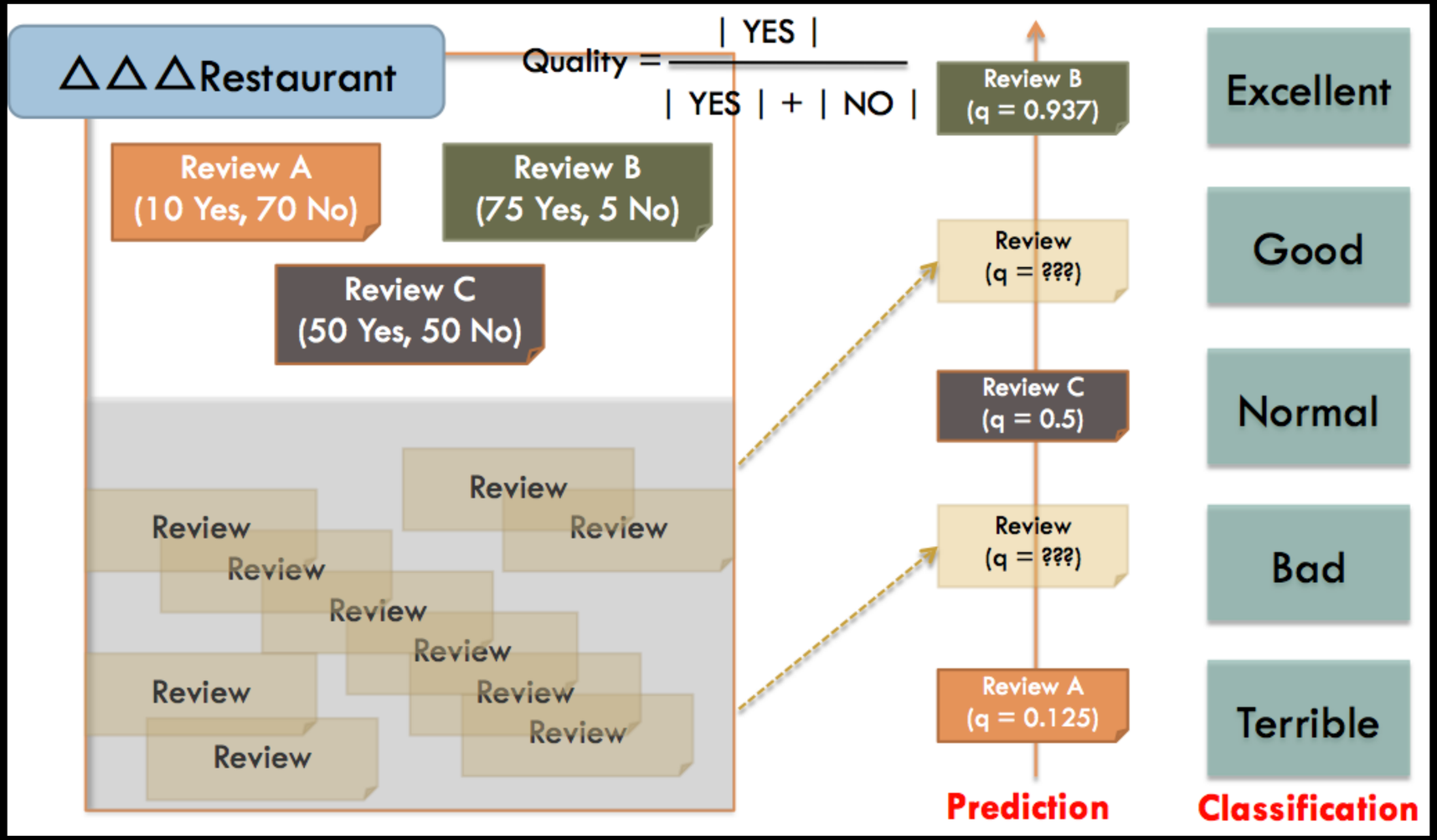
–Albert Einstein

In Services

Google Translate App

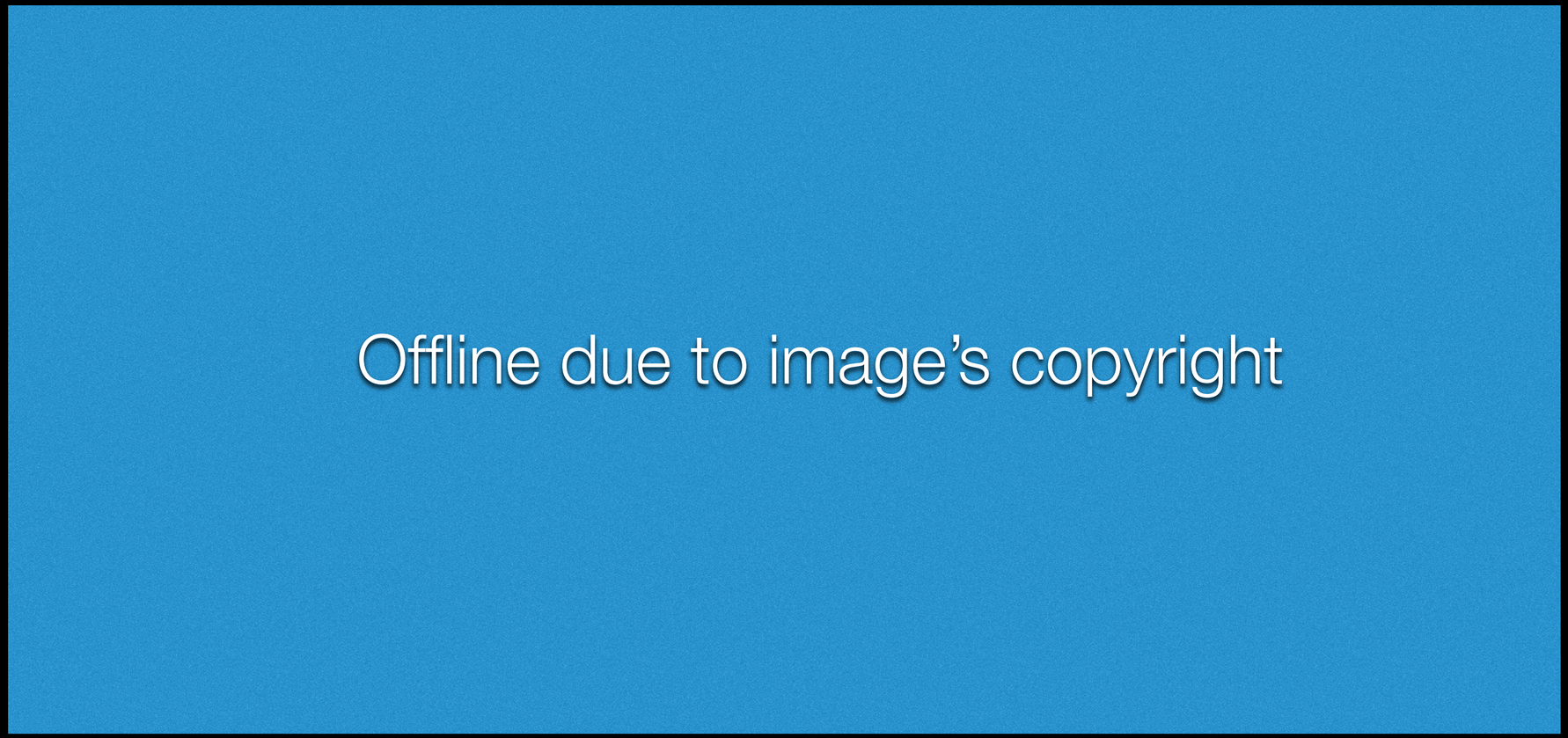
Deep learning inside (work offline)
July 29, 2015 version

In Services



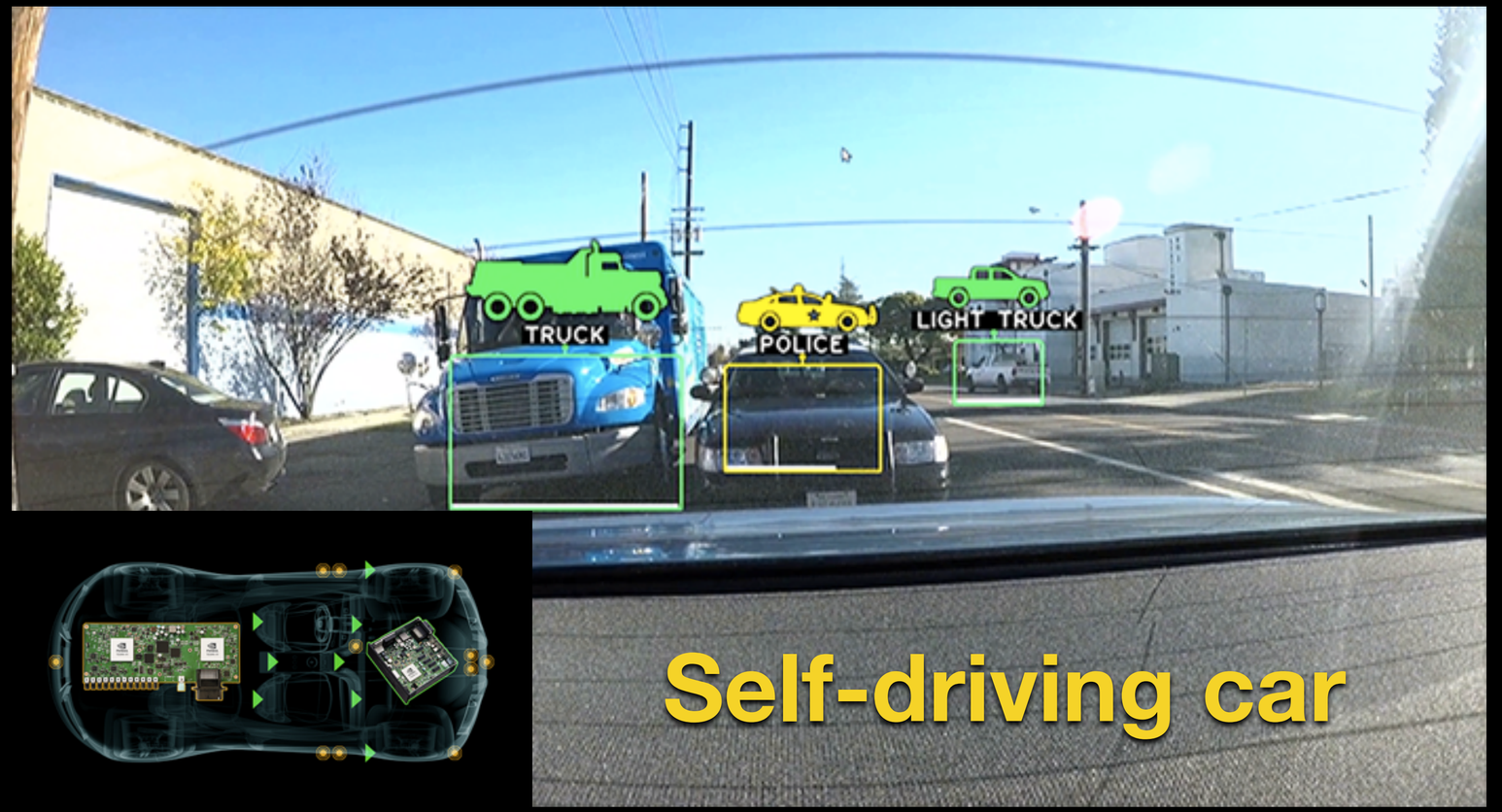
My implemented service http://yelpio.hongo.wide.ad.jp/

In Medical Imaging



Segmentation

In Automatic Devices



http://blogs.nvidia.com/blog/2015/02/24/deep-learning-drive/

In Speech Interfaces

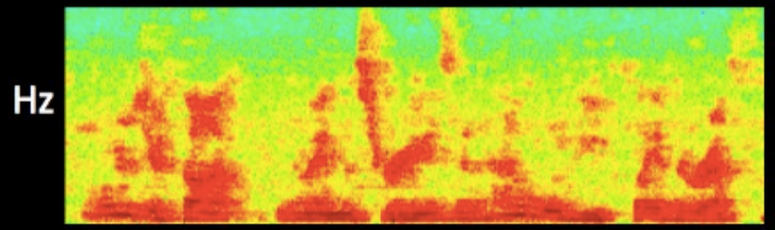
99% is much different with 95%
(Andrew Ng. - Baidu)

Deep Speech – Mandarin

Mandarin is a *tonal* language

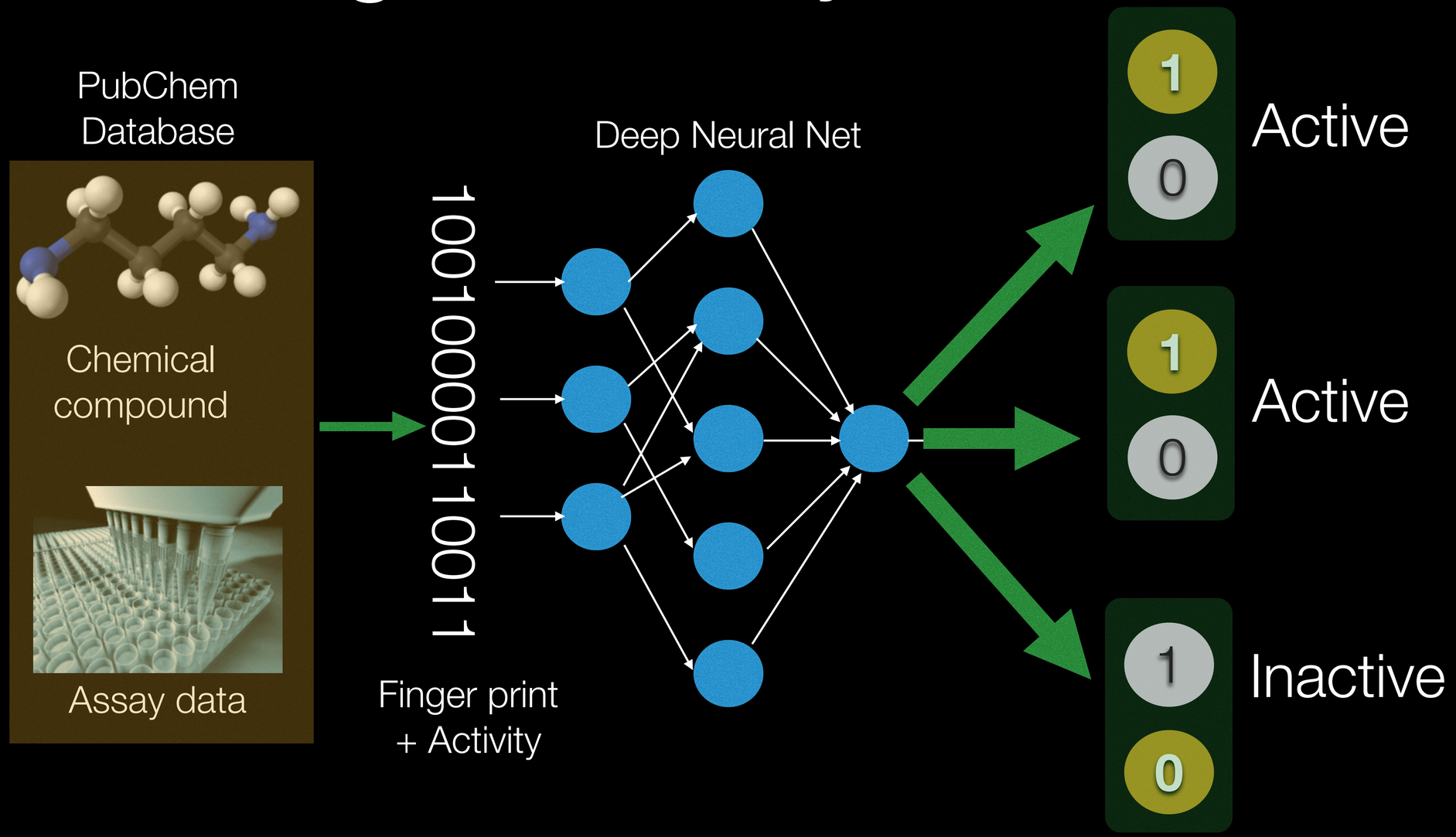
mā má mǎ mà ·ma
妈 麻 马 骂 吗

Model can learn pitch from Spectrogram



https://medium.com/s-c-a-l-e/how-baidu-mastered-mandarin-with-deep-learning-and-lots-of-data-1d94032564a5

In Drug Discovery



http://on-demand.gputechconf.com/gtc/2015/presentation/S5813-Nobuyuki-Ota.pdf

Applications in IoT(Internet of Things)

- More difficult tasks in AI, robotics, information processing
- Huge amount of time series data and states of sensor & devices data
  - **RNN (recurrent neural network)**
- Difficult to get supervisor data
  - **VAE (variational auto encoder)**
- Take action in conditions, environments
  - **Deep Reinforcement Learning**

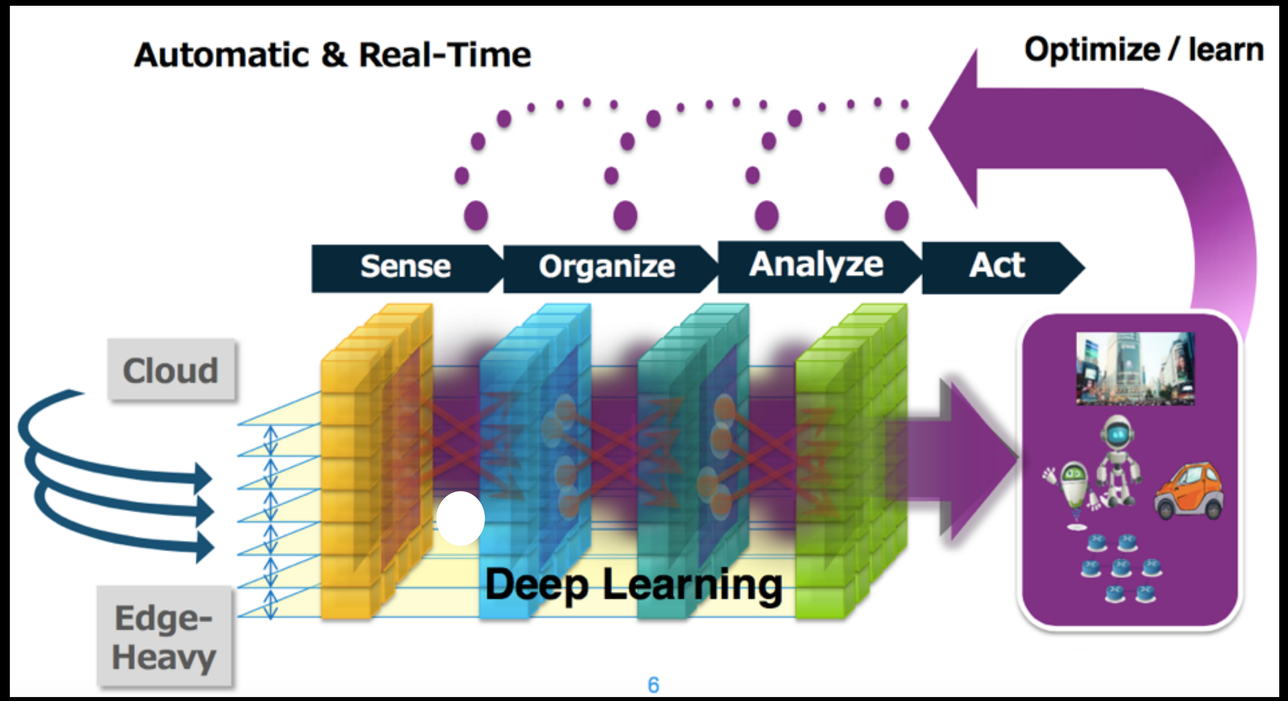
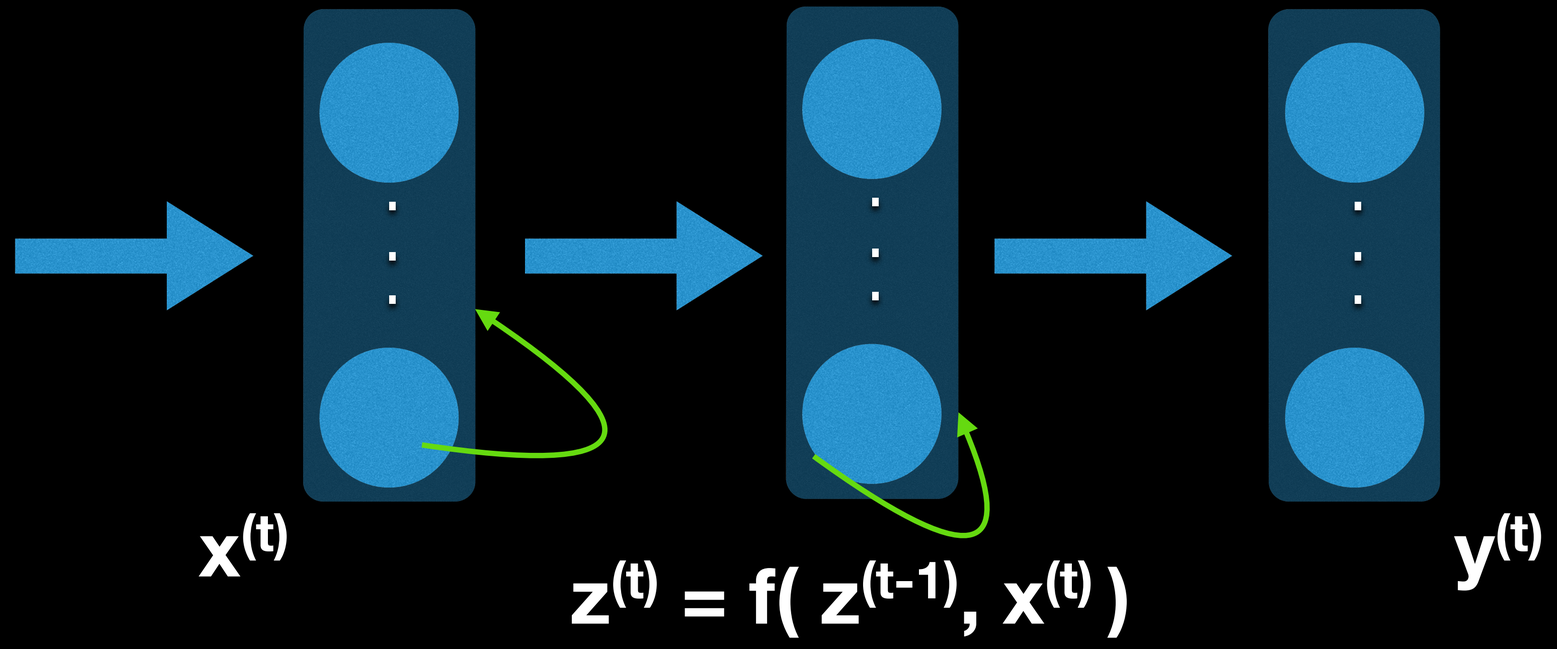


Image source:
http://on-demand.gputechconf.com/gtc/2015/presentation/S5813-Nobuyuki-Ota.pdf

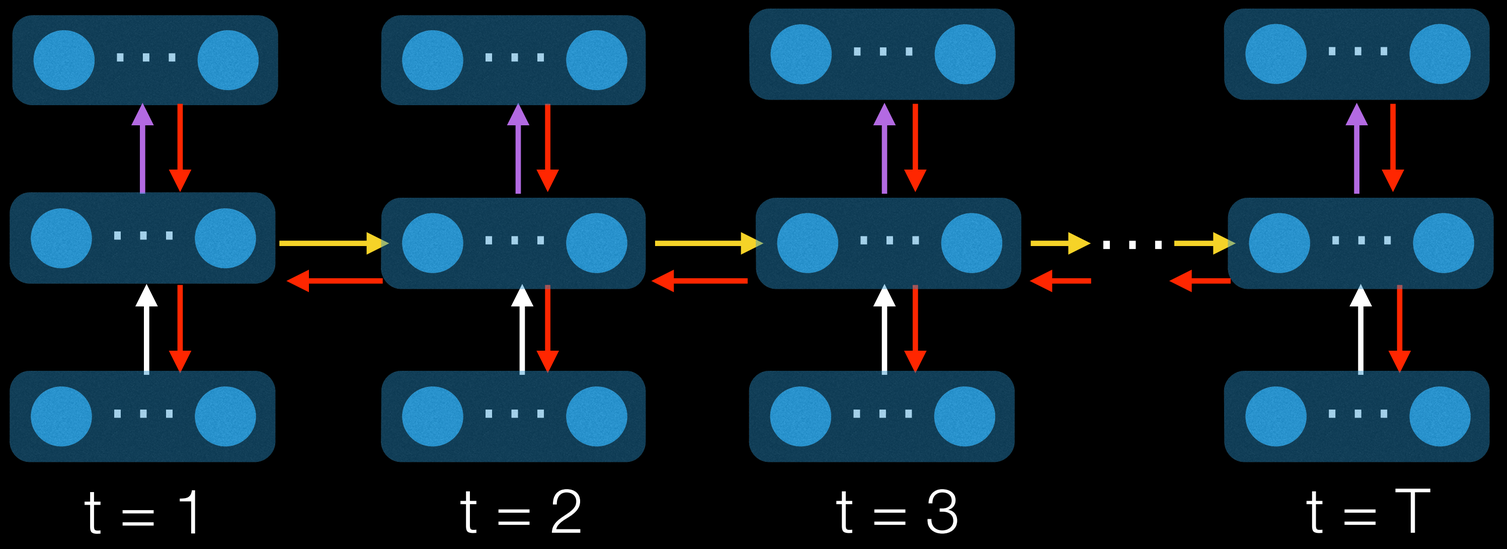
RNN (1/3)

- Recurrent Neural Network: loop inside neural network
- Represent for time series data, sequence of inputs (speech model, natural language model,…)



RNN (2/3)

- Time series extension of RNN



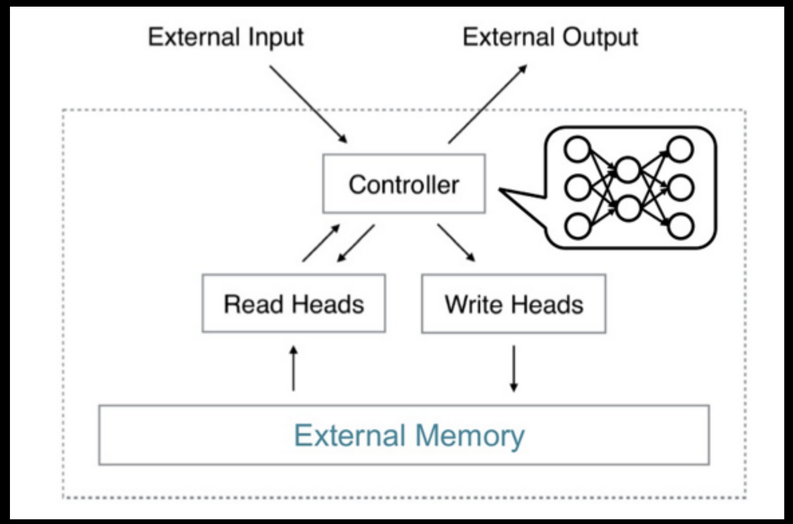
How to learn parameters?

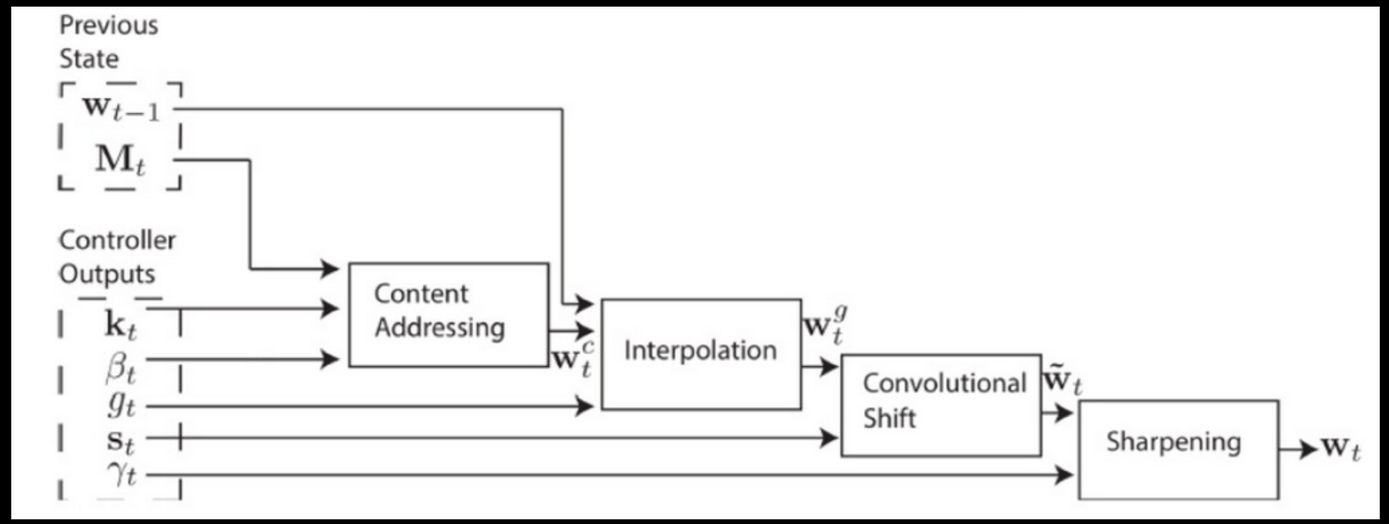
Back propagation through time (BPTT)
+ Long Short-Term Memory

RNN (3/3)

- Neural Turing Machine [A. Graves, 2014]
  - Neural network which has the capability of coupling the external memories
  - Applications: COPY, PRIORITY SORT

NN with parameters to coupling to external memories

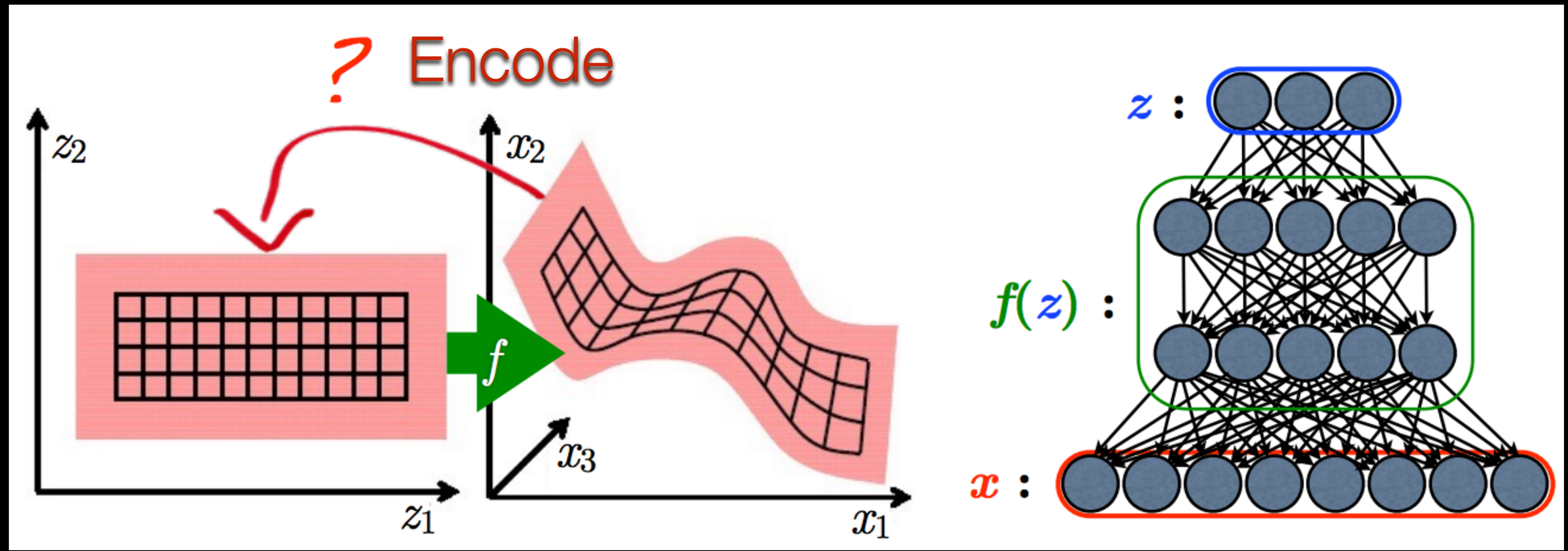




VAE (variational auto encoder)

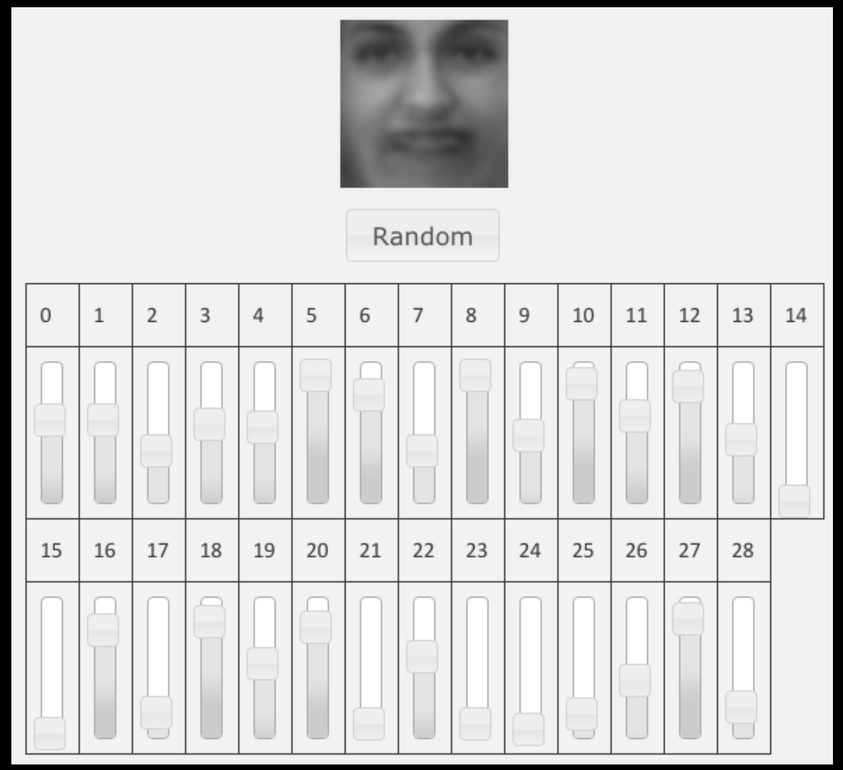
- Learn a mapping from some latent variable z to a complicated distribution on x

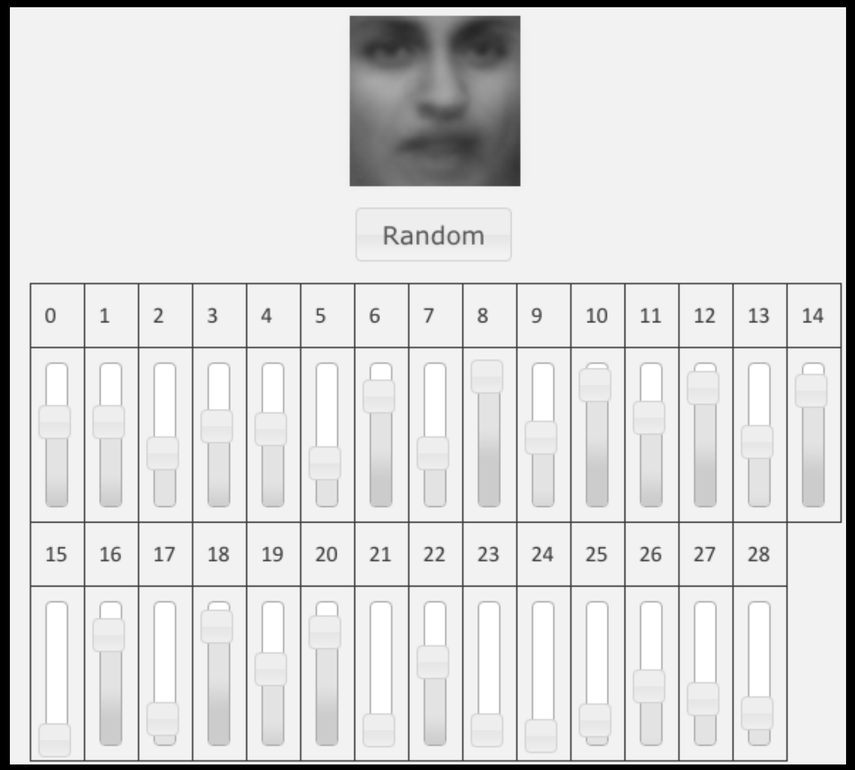
  p(x) = ∫p(x,z)dz where p(x, z) = p(x|z)p(z)
  p(z) = something simple and p(x|z) = f(z) = neural network



VAE

- Example: generate face image from 29 hidden variables





http://vdumoulin.github.io/morphing_faces/online_demo.html

VAE

- Example: learning from number and writing style
- Experiment (PFN): given a number x, produce outputs as remained numbers with writing style of x
- VAE: useful in half-supervised learning process (especially when training data are not enough)

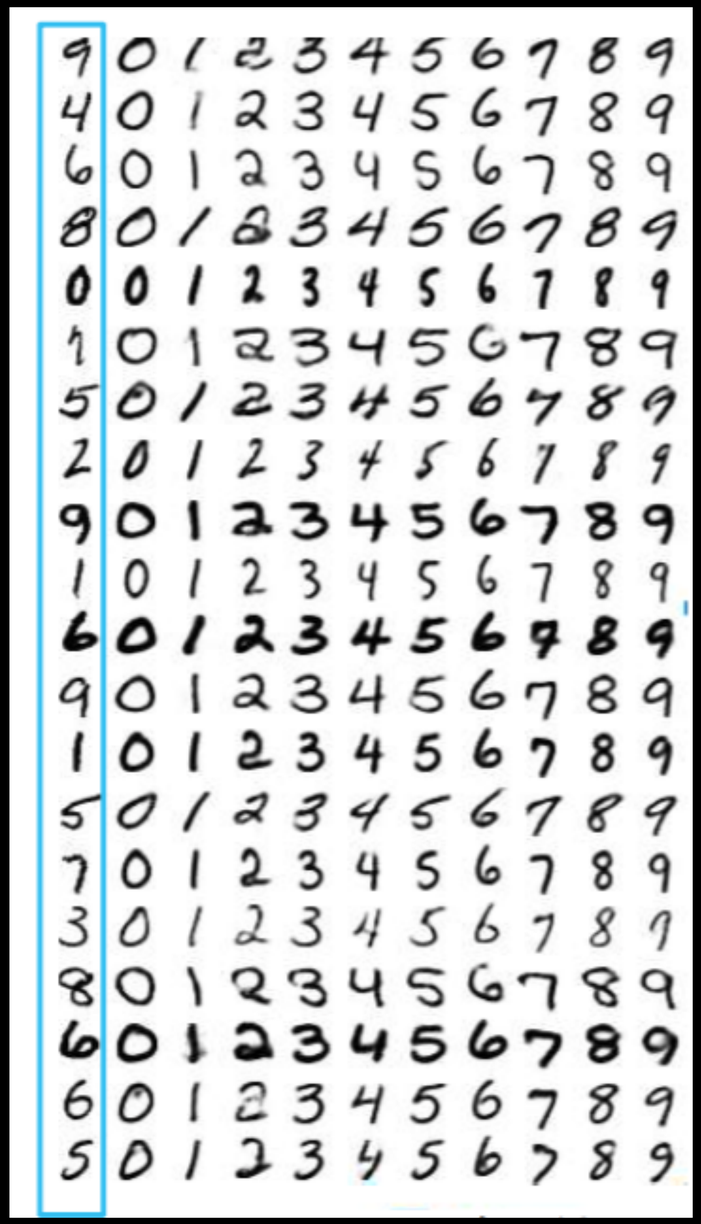


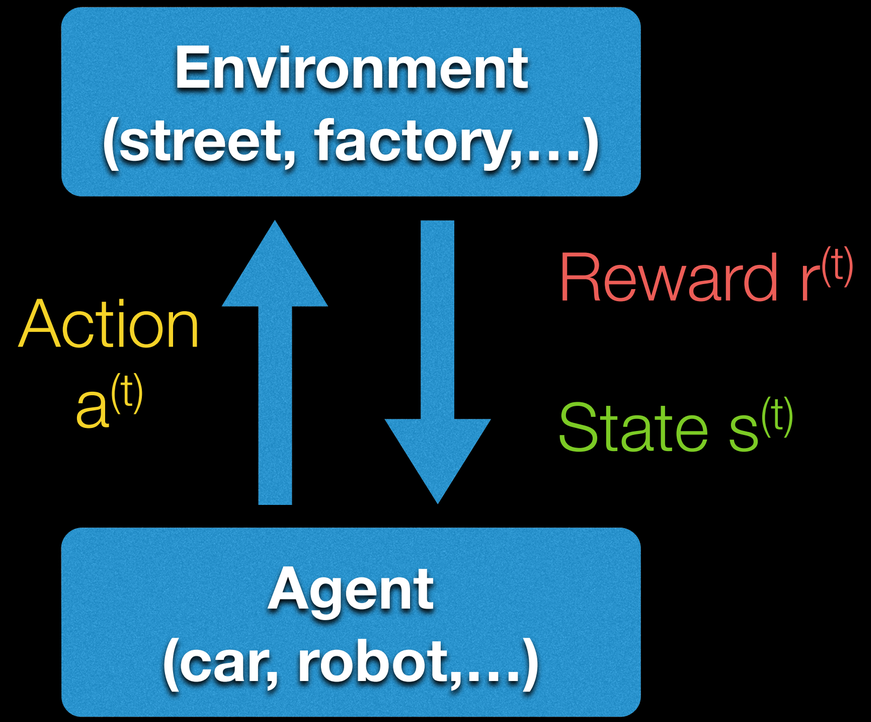
Image from ディープラーニングが活かすIoT, Preferred Networks, Inc. 2015/06/09 Interop 2015 seminar

Reinforcement Learning (RL)

Total rewards in the future

$$R = r^{(t)} + \mu r^{(t+1)} + \mu^2 r^{(t+2)} + \ldots$$
$$(\mu < 1)$$

Design next action

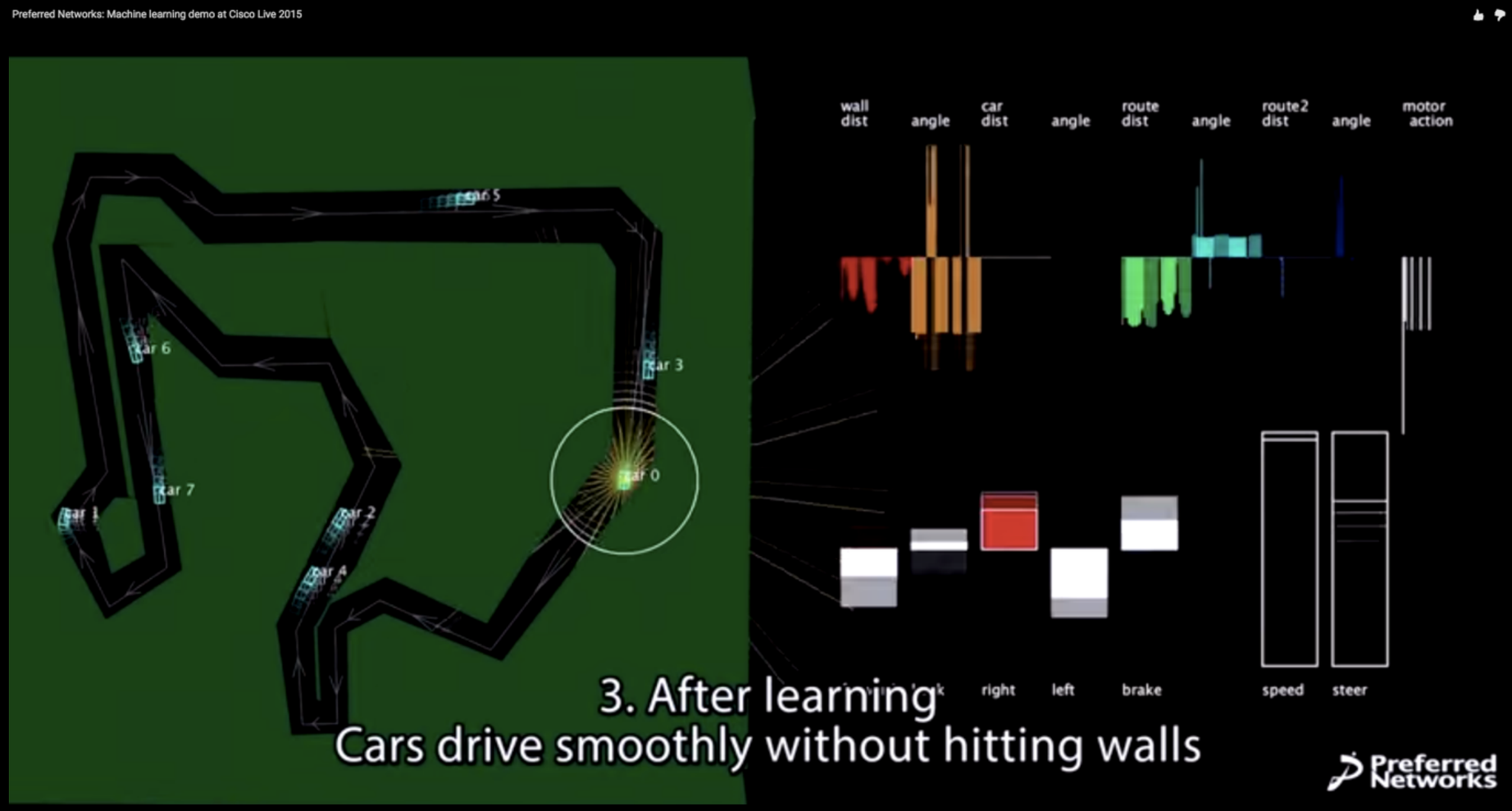


Deep Reinforcement Learning (RL)

- Q - learning: need to know future expectation reward Q(s, a) of action a at state s (Bellman update equation)
  $Q(s, a) \leftarrow Q(s, a) + \beta (r + \mu \max_{a'} Q(s', a') - Q(s, a))$
- Deep Q-Learning Network [V. Mnih, 2015]
  - Get Q(s, a) by a deep neural network Q(s, a, w)
    - Maximize: $L(w) = E[(r + \mu \max Q(s', a', w) - Q(s, a, w))^2]$
  - Useful when there are many states

PFN - demo

https://www.youtube.com/watch?v=RH2TmreYkdA



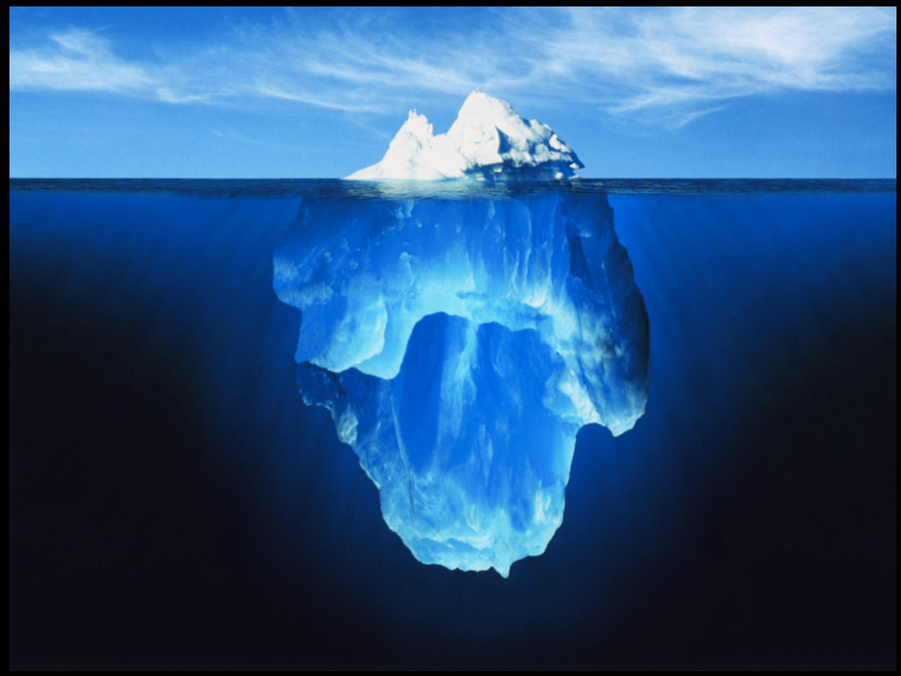
Deep Learning and Business Models

“If we knew what it was we were doing, it would not be called research, would it?”

–Albert Einstein

The future of the field

- Better hardware and bigger machine cluster
- Better implementations and optimizations
- Understand video, text and signals
- Develop in business models



Deep learning is just the tip of the iceberg

Model 1 - Framework development

- Bridge the gap between algorithms and implementation
- Provide common interfaces and understandable API for users

Decaf / Caffe
a Berkeley Vision Project

torch

theano

Pylearn2

Chainer

Model 2 - Building a deep-able hardware

- New hardware (chipset) architecture for Deep Learning

http://www.bdti.com/InsideDSP/2015/04/21/Synopsys

https://www.qualcomm.com/news/snapdragon/2015/03/02/cognitive-computing-and-custom-cpu-drive-next-gen-snapdragon-processors

https://developer.nvidia.com/devbox

Model 3 - Deep Intelligences for IoT

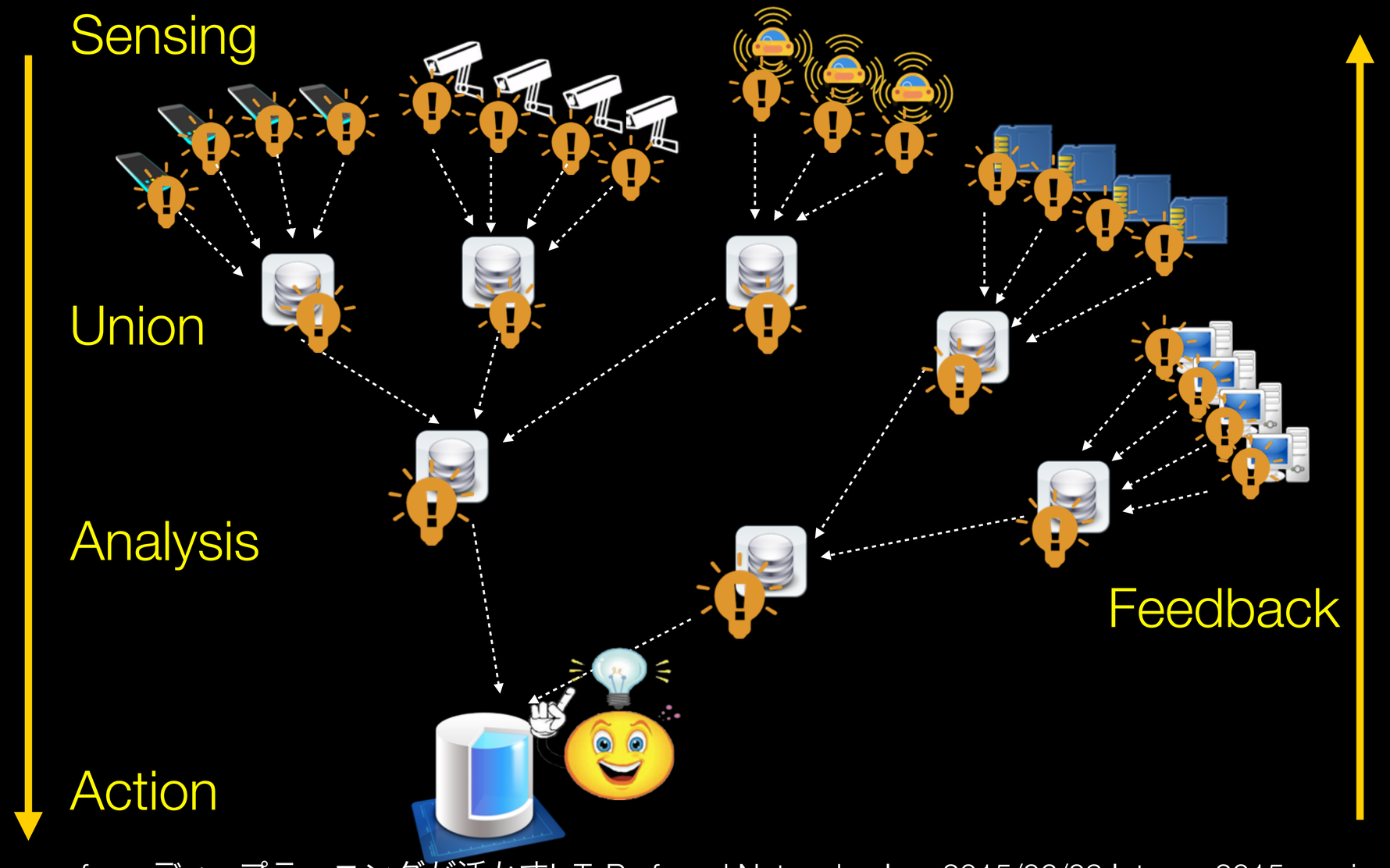
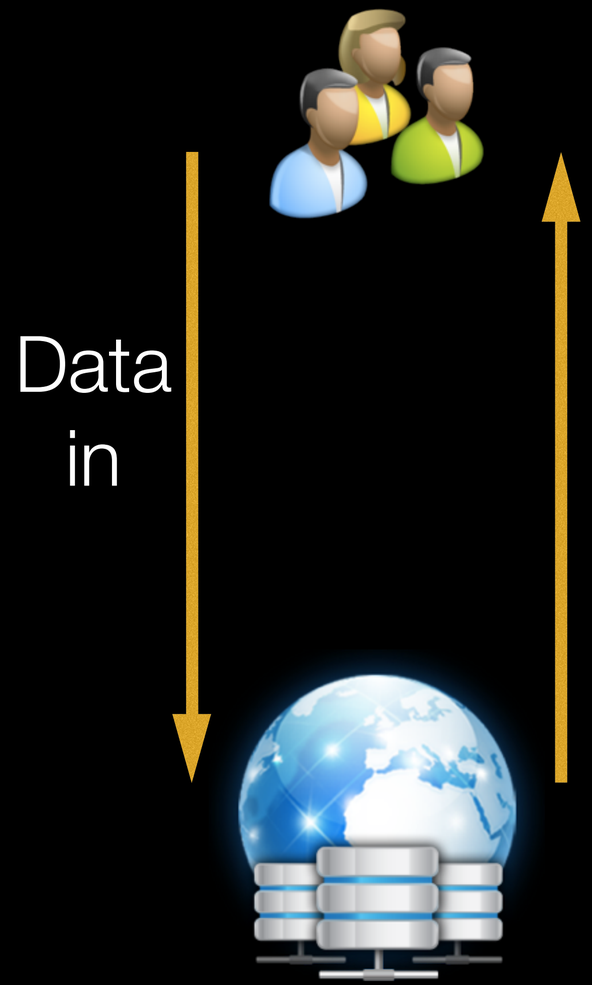
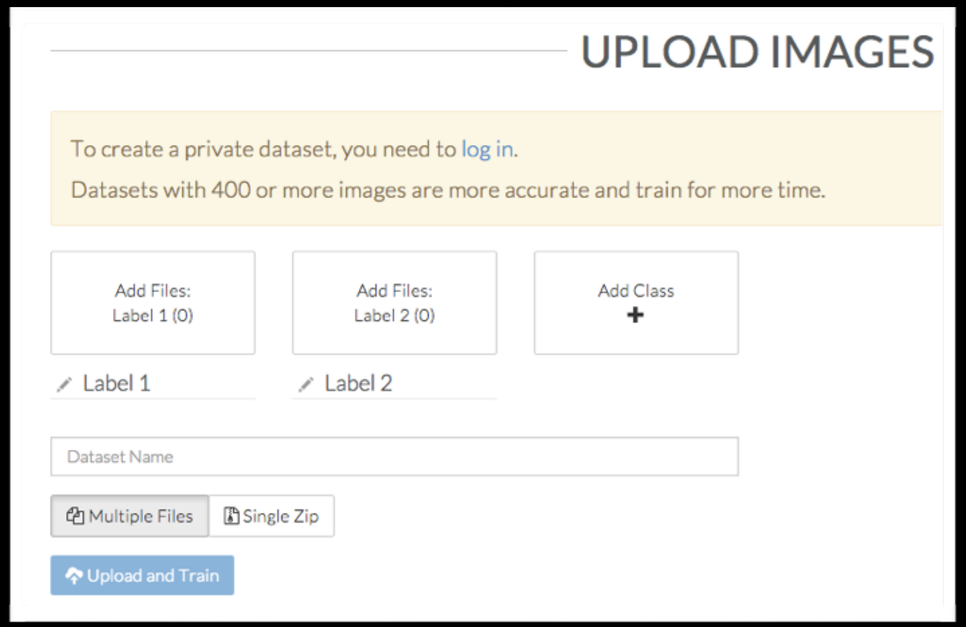


Image from ディープラーニングが活かすIoT, Preferred Networks, Inc. 2015/06/09 Interop 2015 seminar

Model 4 - Host API and deep learning service





MetaMind

https://www.metamind.io/vision/train

https://www.metamind.io/language/train

Usability + Tuning + Scale out

Model 5 - Personal Deep Learning

Applied previous solutions
Almost free

Facial verification & tagging

http://www.adweek.com/socialtimes/deepface/433401

Summary

- Deep Learning has the capability to find patterns among data by enabling wide range of abstraction.
- Deep Learning shown significant results in voice and image recognition compared with conventional machine learning methods.
- Deep Learning has potential applications and business models in some important key sectors.

The challenges are your ideas !

“Thank you for listening.”

- Some good materials for learning
  - Papers, code, blog, seminars, online courses
  - For beginner: 深層学習 (機械学習プロフェッショナルシリーズ)
  - Deep Learning - an MIT Press book in preparation
    - http://www.iro.umontreal.ca/~bengioy/dlbook/