Practical Machine Learning

Lecture 1
Introduction to Machine Learning

Dr. Suyong Eum

OSAKA UNIVERSITY

Course Outline

- Lecture program
- Assignments / Assessment
- Subject Website
- Contact detail
- Pre-requisite for the class
- Regarding tutorial sessions

Lecture Program

	Title of each lecture
Week 1	Introduction to Machine Learning
Week 2	Linear models for classification and regression
Week 3	K-means model and Gaussian Mixture Model (GMM)
Week 4	Hidden Markov Model (HMM)
Week 5	Support Vector Machine (SVM) and Kernel trick
Week 6	Principal Component Analysis (PCA)
Week 7	Neural Networks
Week 8	Convolutional Neural Networks (CNN)
Week 9	Tensorflow – CNN implementation
Week 10	Recurrent Neural Networks (RNN) and Long Short Term Memory (LSTM)
Week 11	Tensorflow – RNN/LSTM/GRU implementation
Week 12	Generative Models: Variational Auto Encoder (VAE) and Generative Adversarial Network (GAN)
Week 13	Tensorflow – VAE and DCGAN implementation
Week 14	Reinforcement Learning (RL): Deep Q-Network (DQN) and Policy Gradient (PG: AC/A3C)
Week 15	Tensorflow – DQN/PG/AC/A3C implementation

Non-neural network based machine learning algorithms (Weeks 1–5)

Preprocessing and feature extraction (Week 6)

Neural networks based machine learning algorithms (Weeks 7–13)

Reinforcement Learning (Weeks 14–15)

Assignment & Assessment

- No examination

- There will be three assignments during the course.
  - First two assignments: 2 x 30% = 60%
  - Last assignment: 40%

- Assignment can be done individually or by a group of three or less
  - No advantage or disadvantage in terms of the number of students

- Attendance is not mandatory

Course website

- ❑ Please visit the website regularly.
- ❑ The latest lecture notes and information will be available before the lecture.
- ❑ You can find assignments as well as their relevant information.
- ❑ Please go to:
  - www.suyongeum.com/ML

Contact detail

- Dr. Suyong EUM (Lecturer)

OSAKA University Graduate School of Information Science and Technology
1-5 Yamadaoka, Suita, Osaka, JAPAN, 565-08. B606
email: suyong[at]ist.osaka-u.ac.jp

- Dr. Hua YANG (tutor)

OSAKA University Graduate School of Information Science and Technology
1-5 Yamadaoka, Suita, Osaka, JAPAN, 565-08. B501
email: h-yang[at]ist.osaka-u.ac.jp

Pre-requisite for the class

- Good knowledge of Python
  - All assignments need to be done with python.
  - One python tutorial will be given by the tutor.

- Some mathematics
  - Linear algebra
  - Optimization
  - Probability theory

Regarding tutorials

- There will be four tutorials during the course (1 hour each)
  1) Introduction to Python (Week 2)
  2) Perceptron algorithm (Week 3)
  3) Support Vector Machine (Week 6)
  4) Principal Component Analysis (Week 7)
- We need to decide when and what time to do that
- The completion of the tutorials will help you to do the first assignment.
- A request from the tutor
  - Make sure the installation of “Anaconda” in your laptop before the tutorial.
  - Python 3.x and setting Path appropriately.
  - https://conda.io/docs/user-guide/install/index.html

Lecture Outline

- Machine learning and its short history
- A typical process in the operation of machine learning algorithms with an example
- What you can do after this course

Machine Learning

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

Machine Learning – Tom M. Mitchell, 1997

- Learning is a process to understand an underlying process through a set of observations.

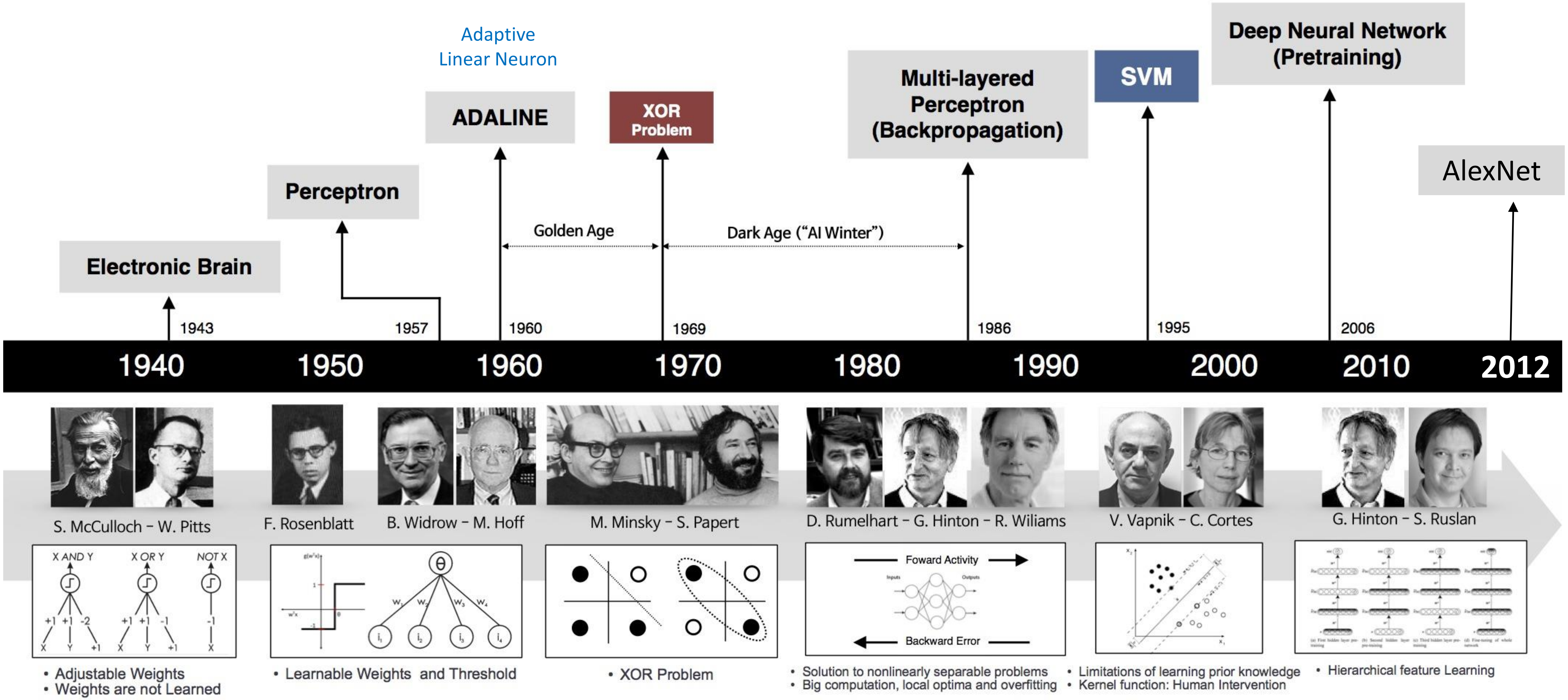
vs

recognition

creation

action

History of Machine Learning



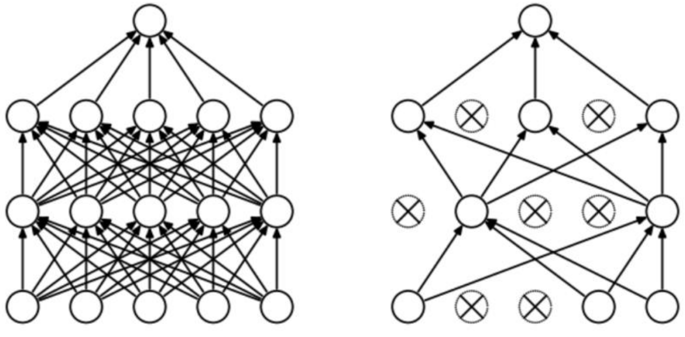
https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html

The perfect storm

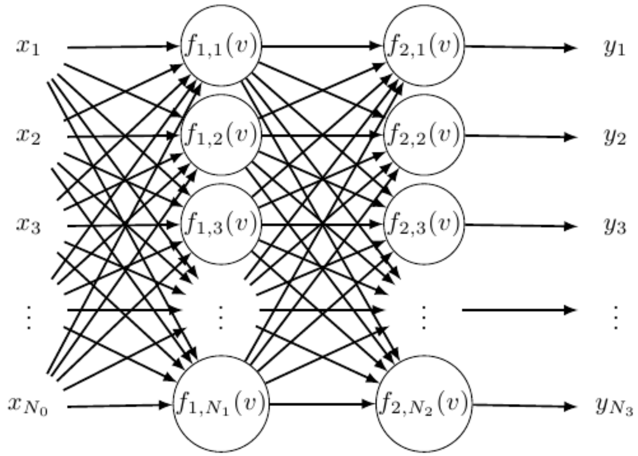
Data

Computation

Algorithms



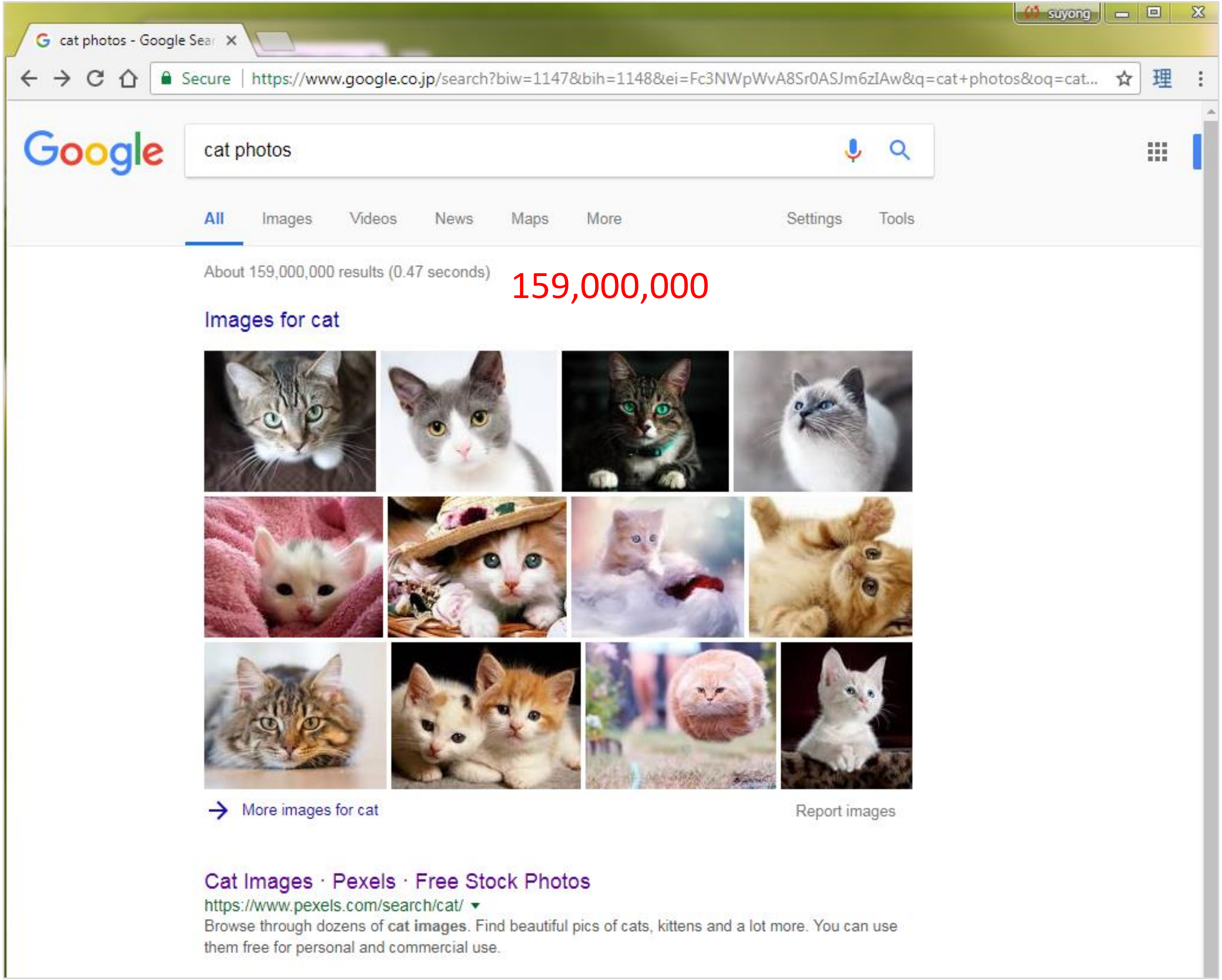
dropout



Backpropagation

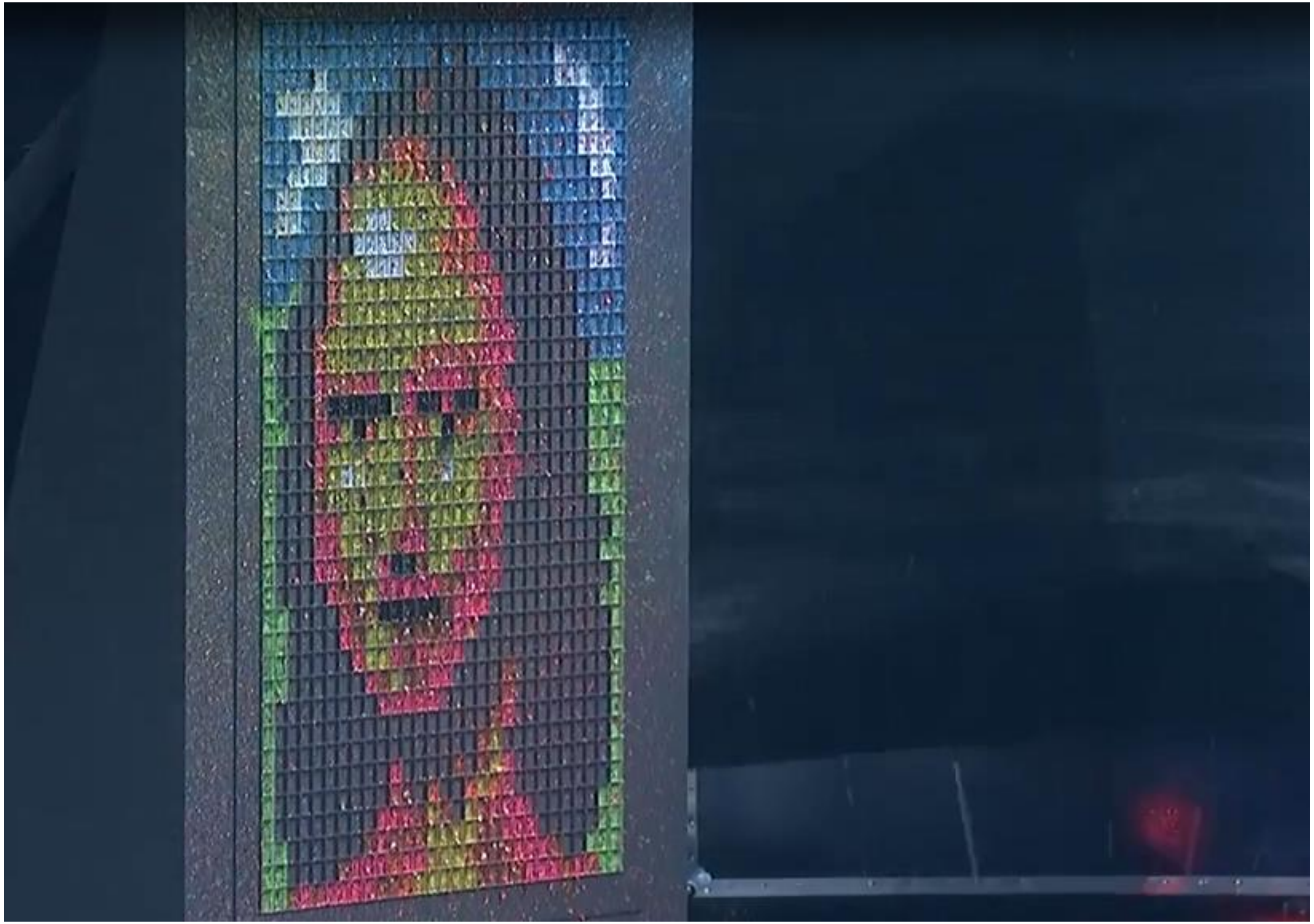
https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html

Data



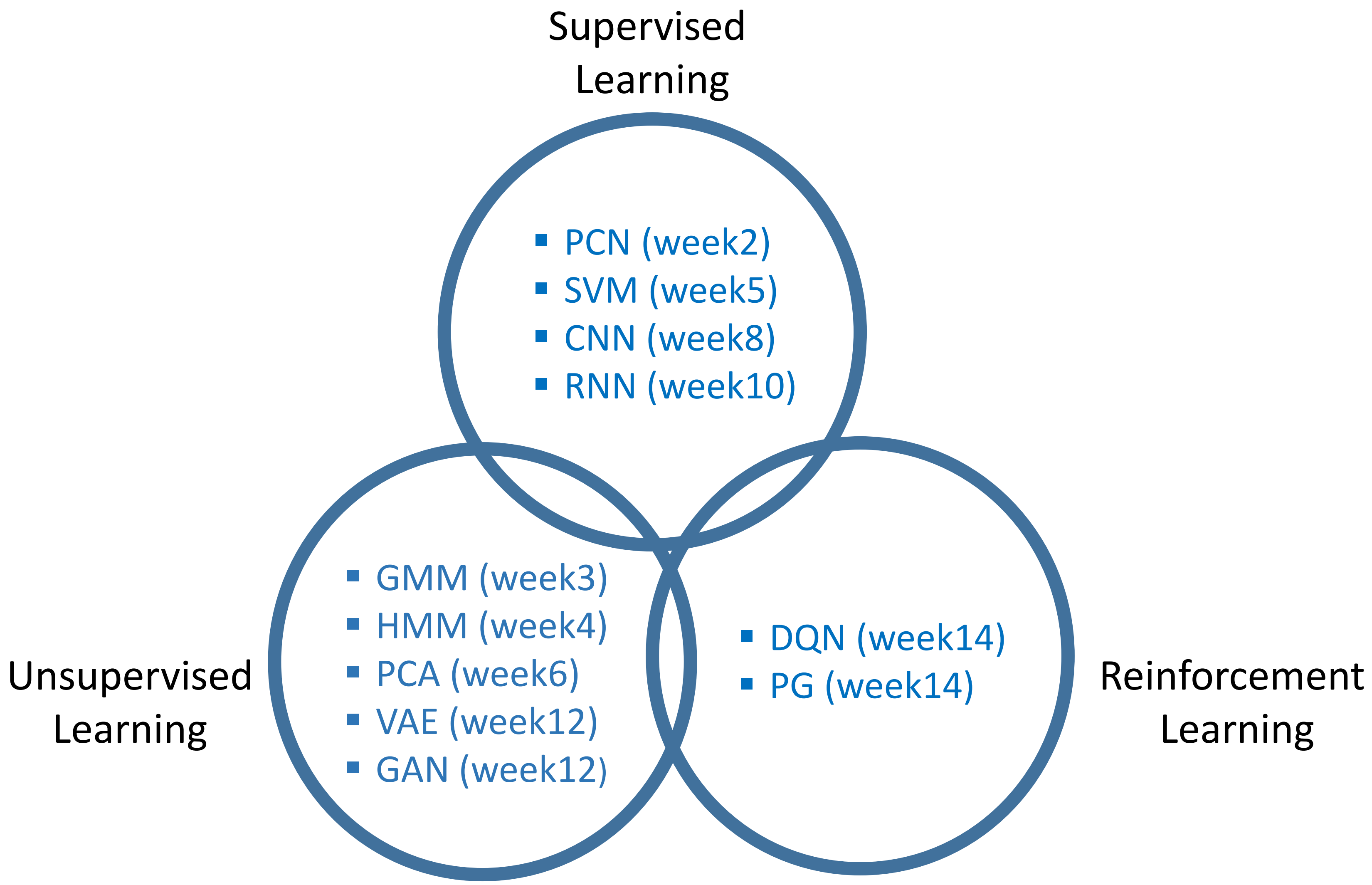
159,000,000

CPU vs GPU demo



https://www.youtube.com/watch?v=-P28LKWTzrI

Types of machine learning algorithms

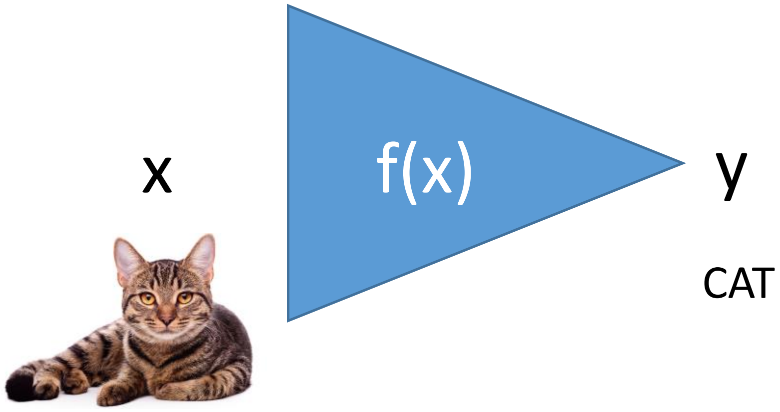


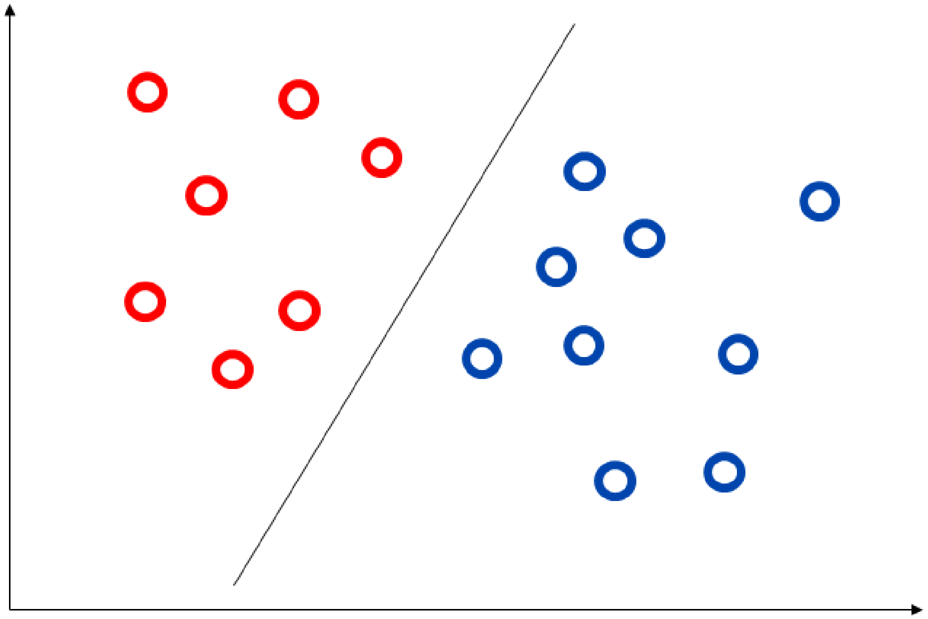
Supervised Learning

- ❑ Input + Output with Label
- ❑ Supervised learning is learning from by a knowledgeable external supervisor.

- *<u>Question -> Answer</u>*

y = f(x)



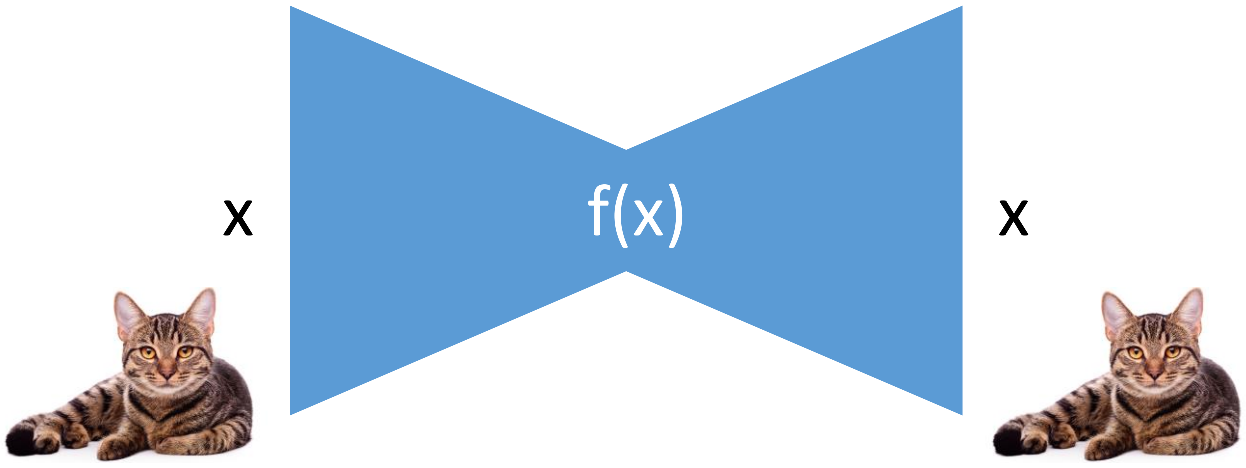


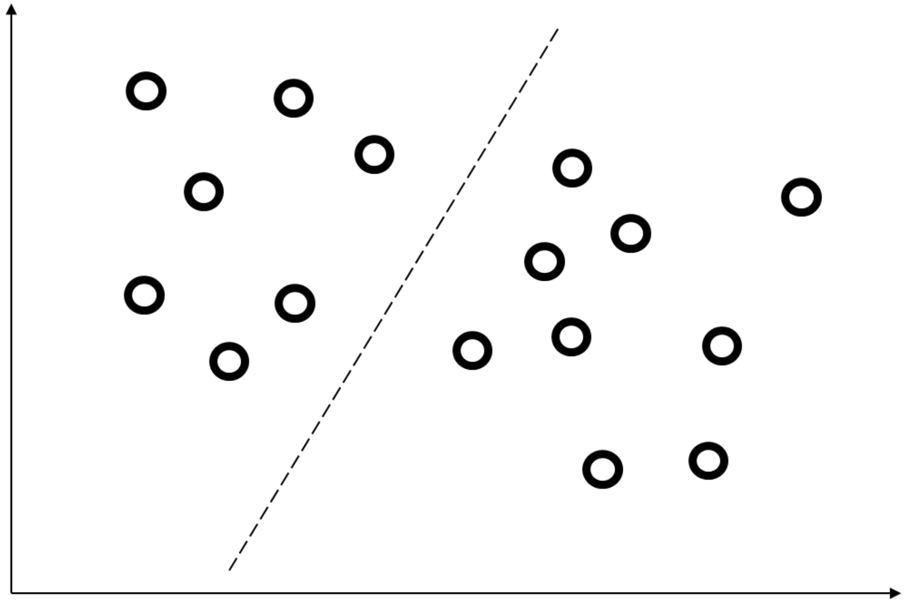
Unsupervised Learning

- ❑ Input + Output without Label
- ❑ Feature Learning

  - *<u>Question -> Question</u>*

    $x = f(x)$



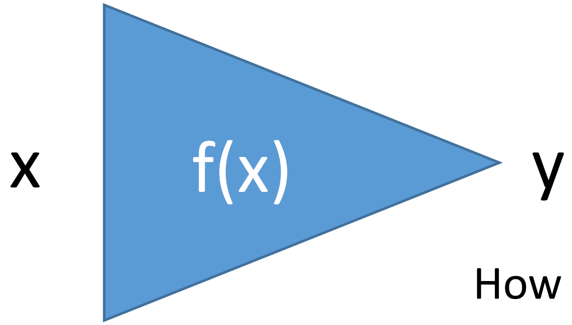


Reinforcement Learning

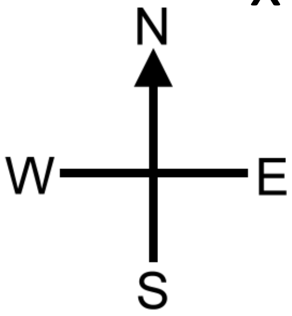
- Input + partial output with its quality: in some sense similar to supervised learning
- An action is rewarded/penalized to take a better action next time

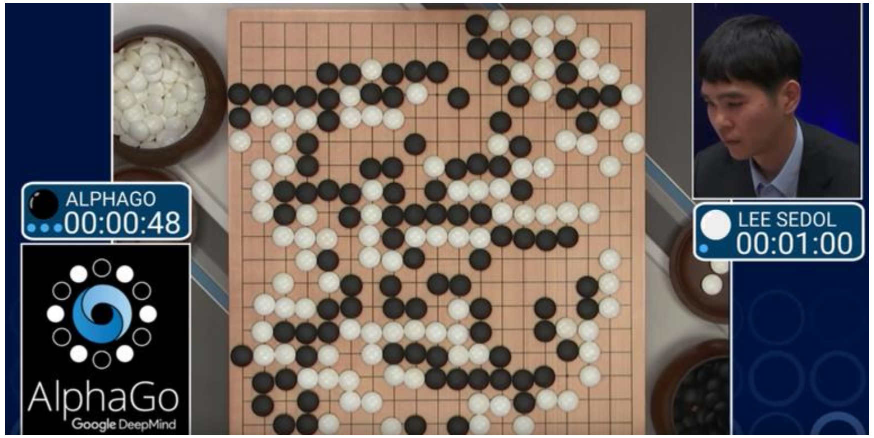
- ***<u>Carrot and stick</u>***

y = f(x)



How good?

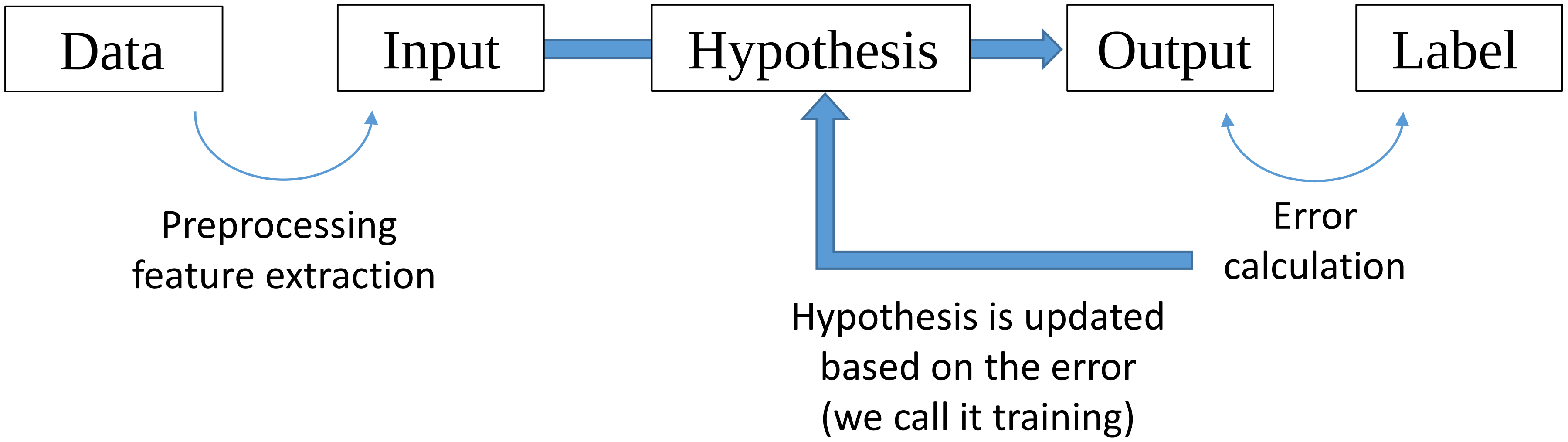




A typical process in the operation of machine learning algorithms with an example

Components of machine learning

- Raw data (including label)
- Input (features: dimension of a data point)
- Hypothesis (a function approximating a target function)
- Output
- Label

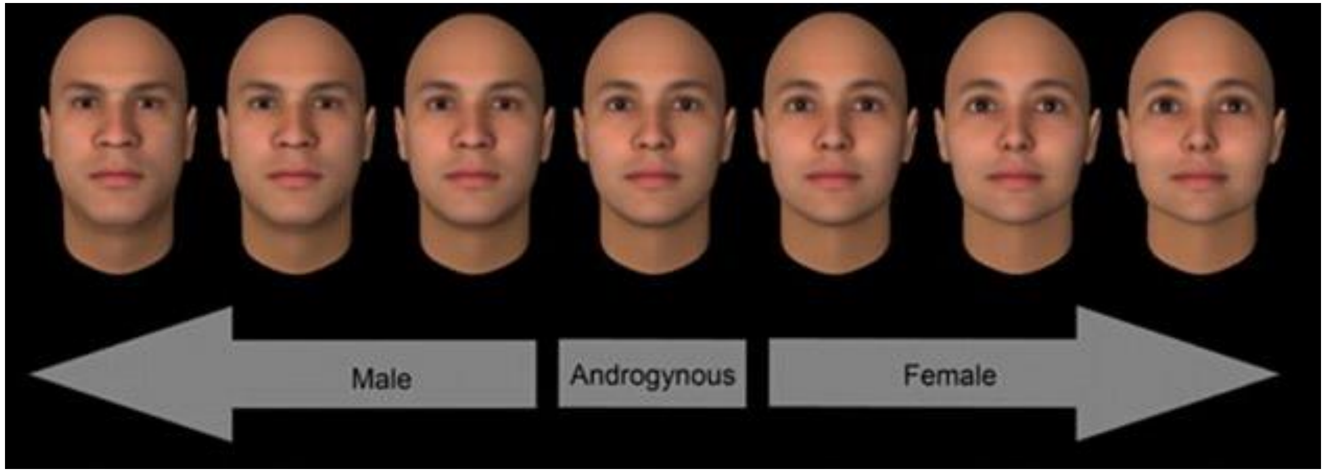


An example: male / female classification

Given data per person

- Height: 170cm
- Weight: 52kg
- Foot size: 25cm
- Hand size: 20cm
- Nose height: 1.5cm
- Eye size: 2.5cm
- Hair length: 5cm

- *Male*

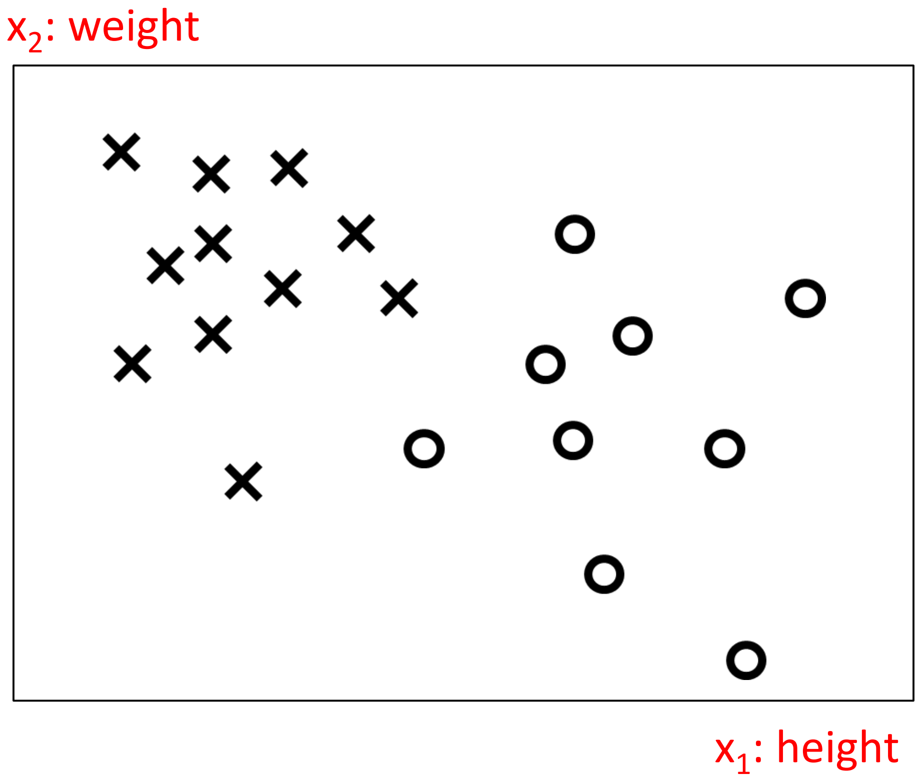


Components	Notation	Description
Data	(x1, y1), (x2, y2), … , (xn, yn)	(a vector data, male/female)
Input	X	xn={x_1,x_2, … , x_d}: d dimensions
Output	Y	Output data from hypothesis
Hypothesis	g: X → Y	(hypothesis) A model
Target function	f: X → Y	Unknown

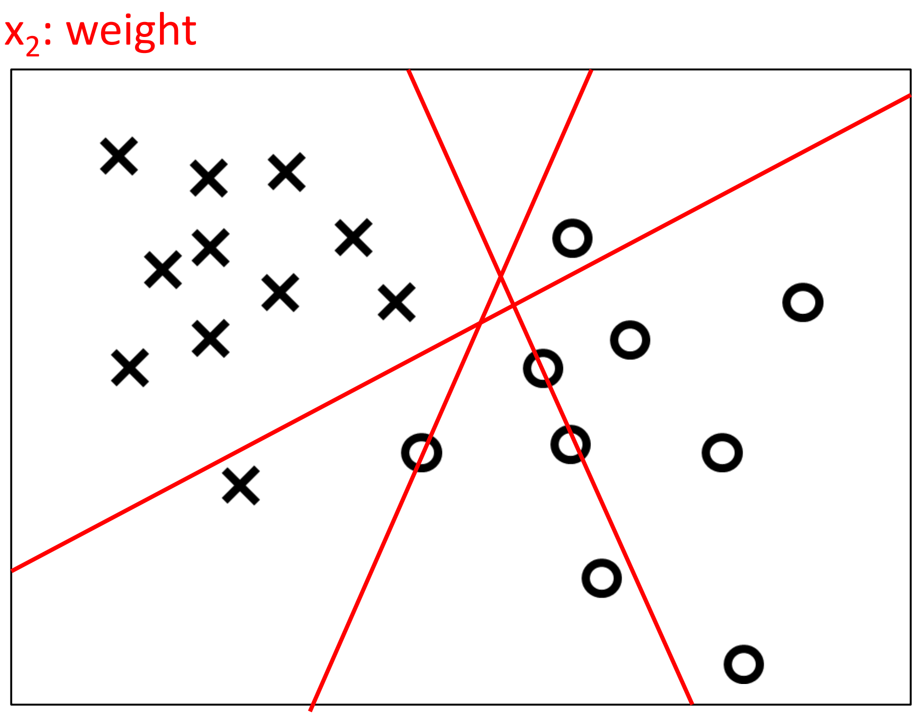
An example: feature extraction

- Which features are important to tell that the given object is male or female?
- Assuming you chose two features and then you plot the data points

- Given data per person
  - Height: 170cm
  - Weight: 52kg
  - Foot size: 25cm
  - Hand size: 20cm
  - Nose height: 1.5cm
  - Eye size: 2.5cm
  - Hair length: 5cm
  - *Male*



An example: hypothesis (model) selection



Female if $\sum_{i=1}^{2} w_i x_i < threshold$

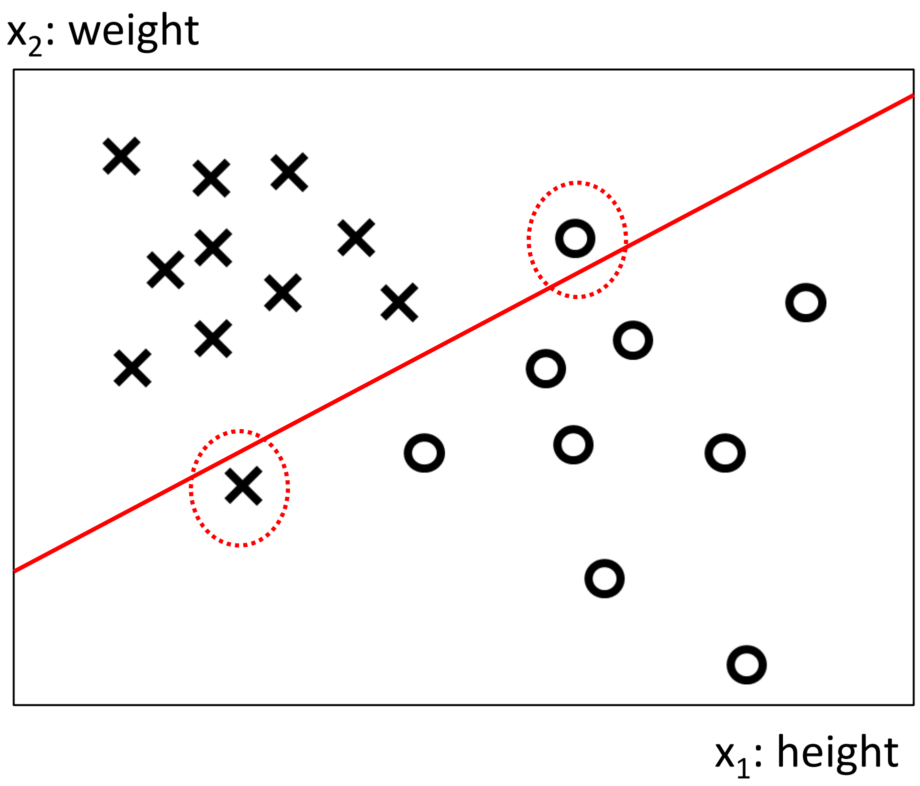
Male if $\sum_{i=1}^{2} w_i x_i > threshold$

$$h(x) = sign\left(\left(\sum_{i=1}^{2} w_i x_i\right) - threshold\right)$$

$$h(x) = sign\left(\sum_{i=0}^{d} w_i x_i\right)$$

$$h(x) = sign\left(\mathbf{w}^{T}\mathbf{x}\right)$$

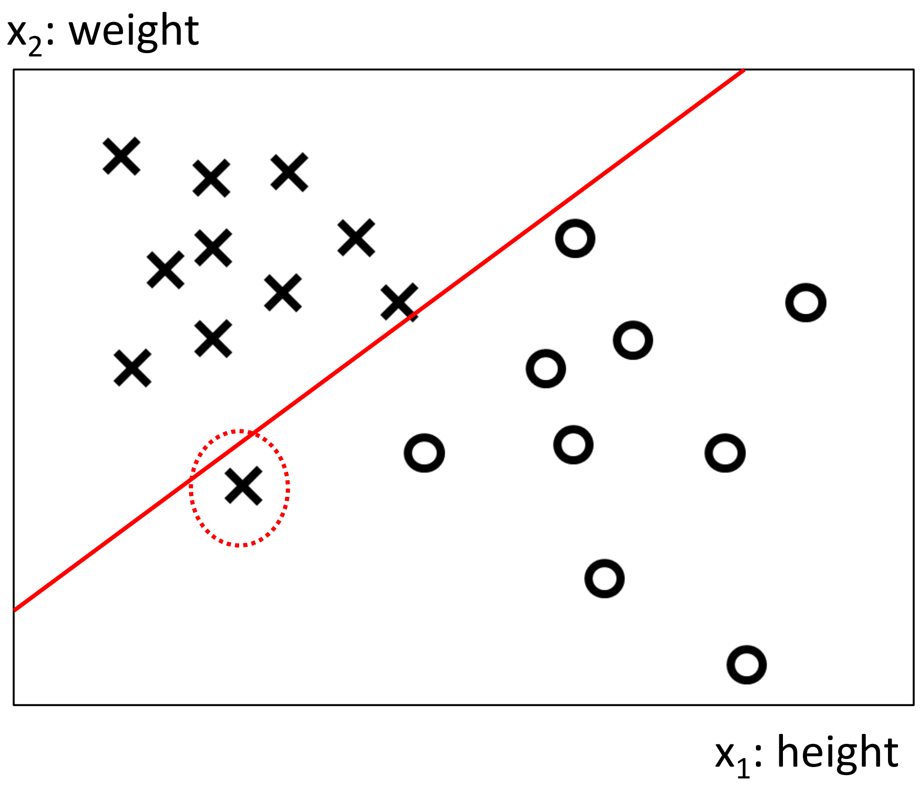
An example: training the hypothesis to produce less error



$$h(x) = sign\left(\mathrm{W}^{\mathrm{T}}\mathrm{X}\right)$$

- ❑ Random selection of *W*
- ❑ Misclassified data points are found
- ❑ Update *W* in order to correctly classify the misclassified data points.
  - How? : depending on learning algorithm
    - Neural network: backpropagation?
    - Linear algebra: perceptron algorithm

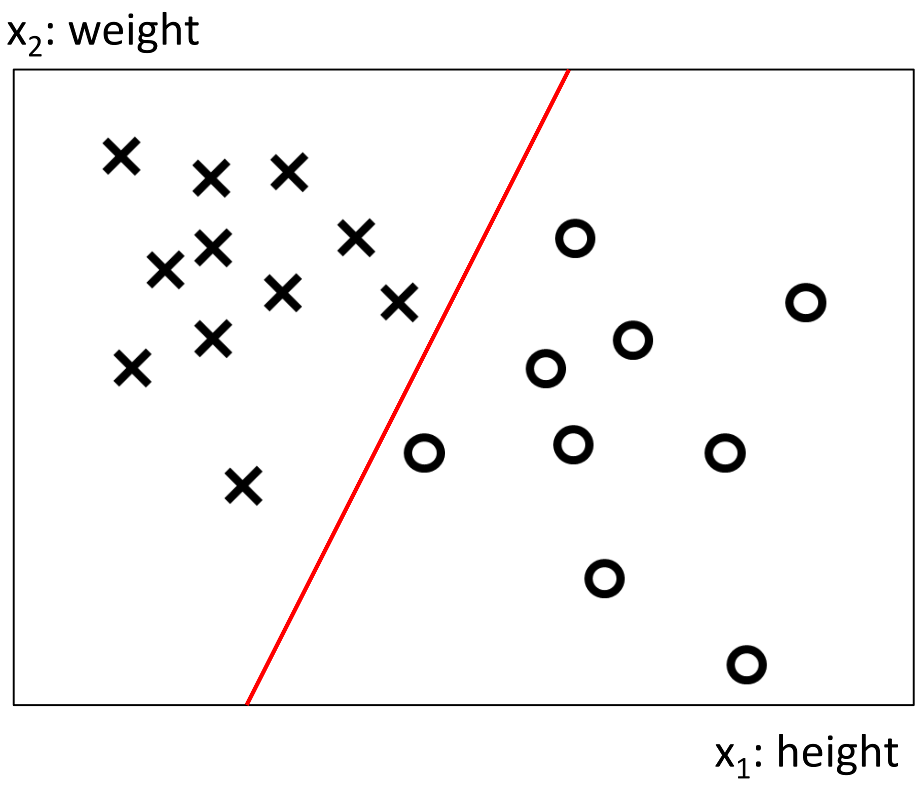
An example: training the hypothesis to produce less error



$$h(x) = sign\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}\right)$$

- ❑ Random selection of *W*
- ❑ Misclassified data points are found
- ❑ Update *W* in order to correctly classify the misclassified data points.
  - How? : depending on learning algorithm
    - Neural network: backpropagation?
    - Linear algebra: perceptron algorithm

An example: training the hypothesis to produce less error



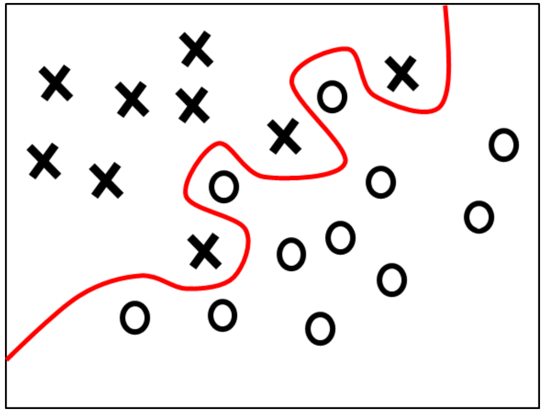
$$h(x) = sign\left(\mathrm{W}^{\mathrm{T}}\mathrm{X}\right)$$

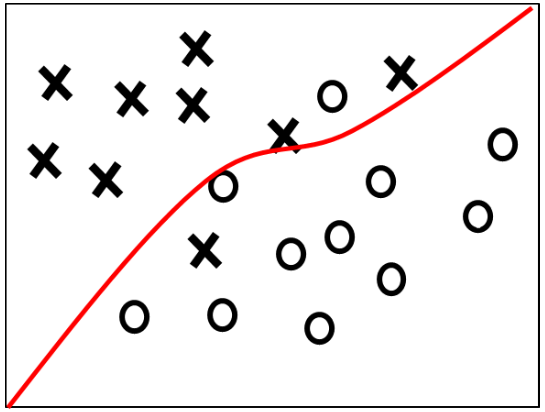
- ❑ Random selection of *W*
- ❑ Misclassified data points are found
- ❑ Update *W* in order to correctly classify the misclassified data points.
  - How? : depending on learning algorithm
    - Neural network: backpropagation?
    - Linear algebra: perceptron algorithm

Cross validation (CV)

- You are generally given one big training data set.
- How to verify goodness of your model?

Occam's razor



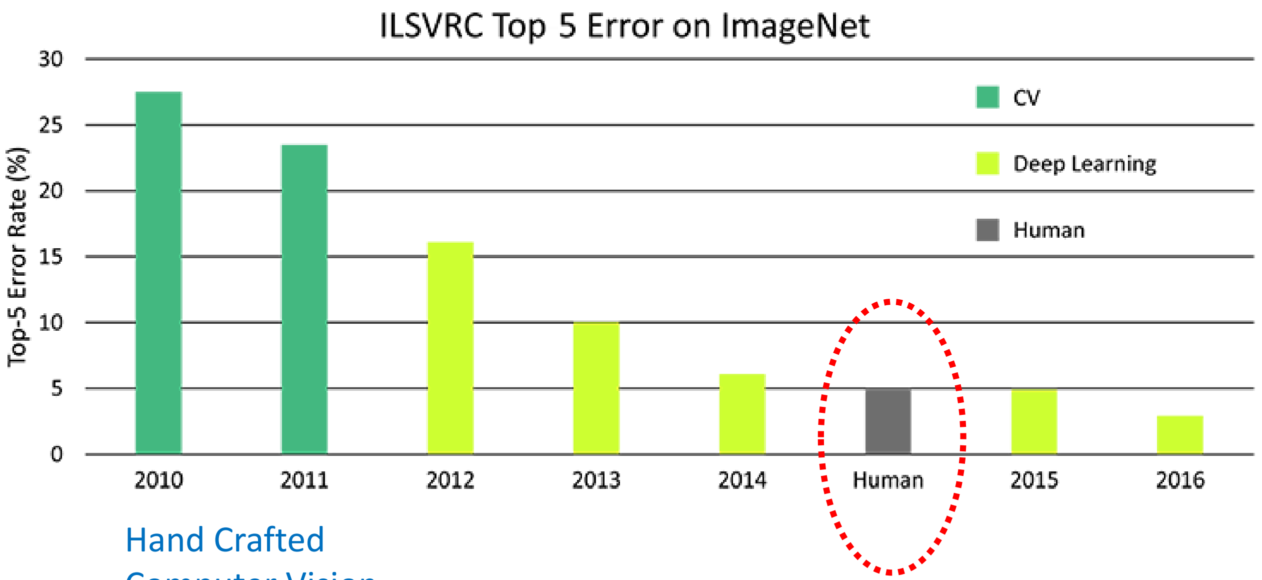


$$y(x, w) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

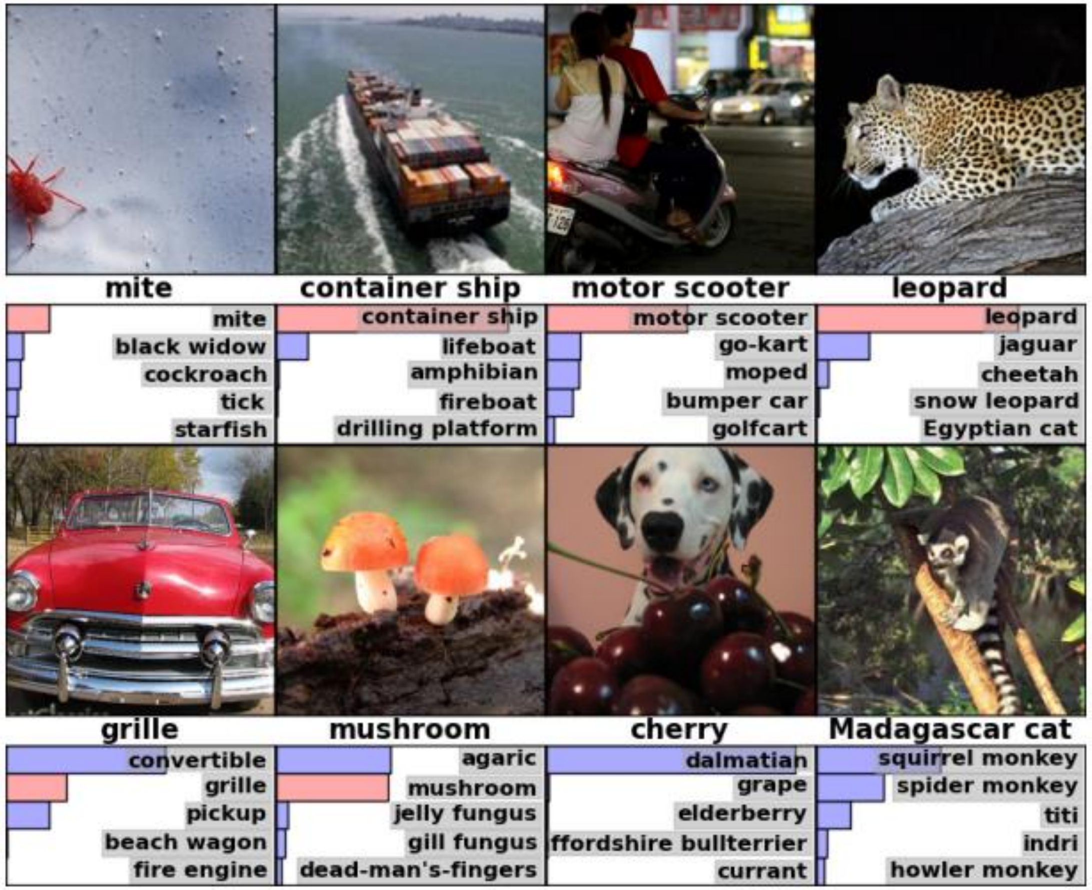
What you can do after this course

Image classification

Large Scale Visual Recognition Challenge (ILSVRC)
- 1000 class objects
- around 1.4 million images



Hand Crafted
Computer Vision
approaches



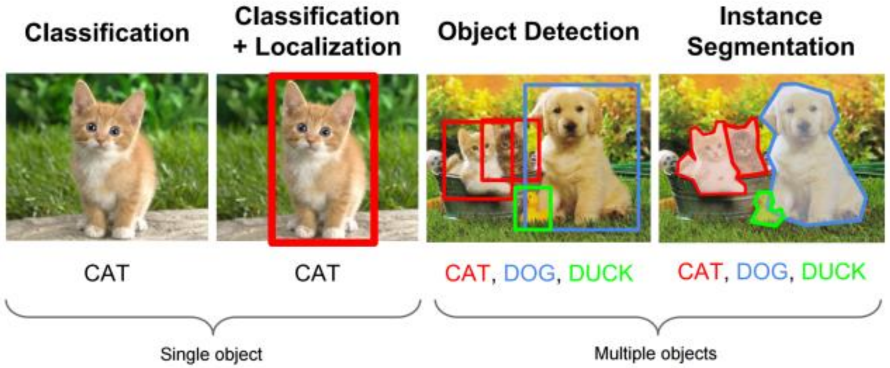
https://www.dsiac.org/resources/journals/dsiac/winter-2017-volume-4-number-1/real-time-situ-intelligent-video-analytics
http://www.cs.toronto.edu/~fritz/absps/imagenet.pdf

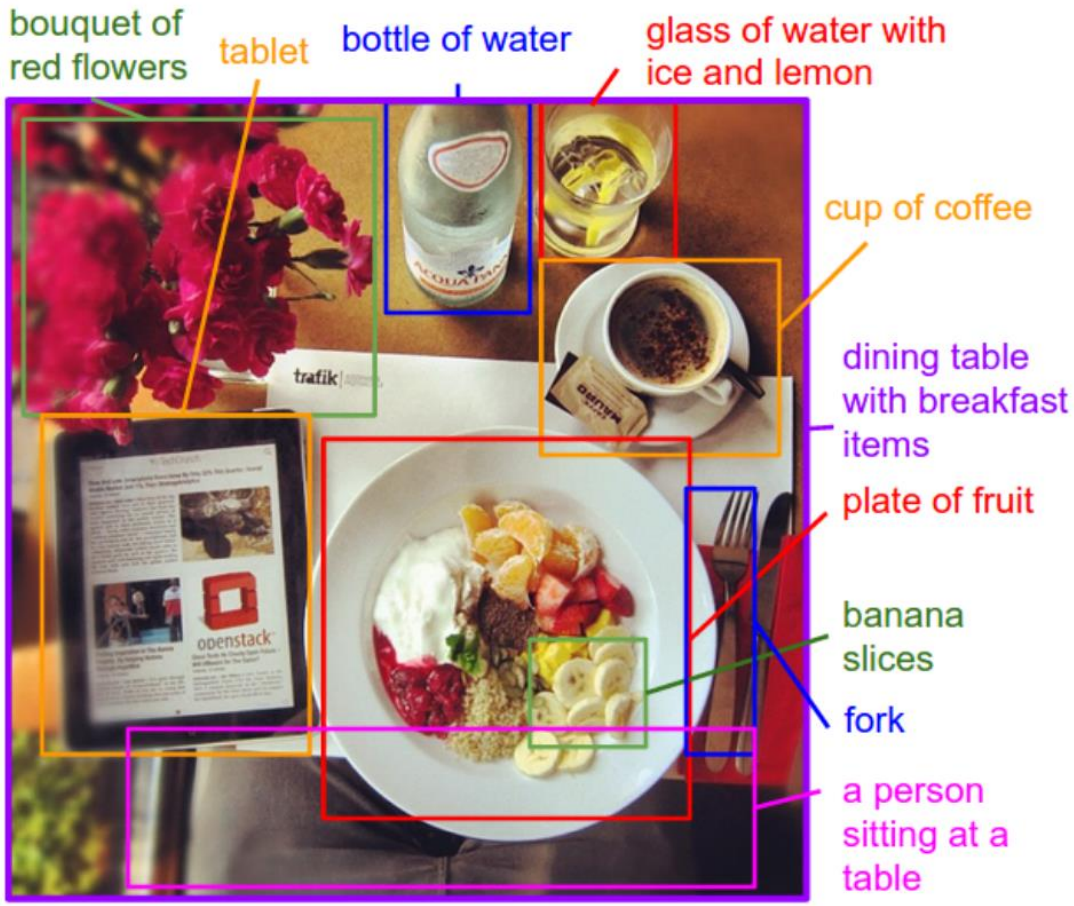
Image detection

2016

Mean Average Precision

Real-Time Detectors	Train	mAP
100Hz DPM [31]	2007	16.0
30Hz DPM [31]	2007	26.1
Fast YOLO	2007+2012	52.7
YOLO	2007+2012	**63.4**
Less Than Real-Time		
Fastest DPM [38]	2007	30.4
R-CNN Minus R [20]	2007	53.5
Fast R-CNN [14]	2007+2012	70.0
Faster R-CNN VGG-16[28]	2007+2012	73.2
Faster R-CNN ZF [28]	2007+2012	62.1
YOLO VGG-16	2007+2012	66.4





- Classification
- Localization
- Detection
- Segmentation

https://arxiv.org/pdf/1506.02640.pdf
https://arxiv.org/pdf/1412.2306v2.pdf

Image caption

a man is riding a motorcycle on a street
logprob: -8.65

a bus is parked on the side of the road
logprob: -7.19

a zebra standing in a field of grass
logprob: -7.88

a woman is standing in front of a store
logprob: -11.40

a woman holding a teddy bear in front of a mirror
logprob: -9.65

a baby laying on a bed with a stuffed bear
logprob: -8.85

https://cs.stanford.edu/people/karpathy/deepimagesent/generationdemo/

Creation: writing

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

Writing new episodes of Friends is easy if you use a neural network

"Chandler: Well, I proposed to my shoe..."

By James Vincent | @jjvincent | Jan 21, 2016, 4:03am EST

SHARE

For $\bigoplus_{n=1,\ldots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, ?? and the fact that any U affine, see Morphisms, Lemma ??. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma ?? we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. □

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\mathrm{Arrows} = (Sch/S)^{opp}_{fppf}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \mathrm{Spec}(A))$$

is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

Proof. See discussion of sheaves of sets. □

The result for prove any open covering follows from the less of Example ??. It may replace S by $X_{spaces,\acute{e}tale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma ??. Namely, by Lemma ?? we see that R is geometrically regular over S.

http://karpathy.github.io/2015/05/21/rnn-effectiveness/
https://www.theverge.com/2016/1/21/10805398/friends-neural-network-scripts

Creation: music composition

https://magenta.tensorflow.org/performance-rnn

Generation: image generation

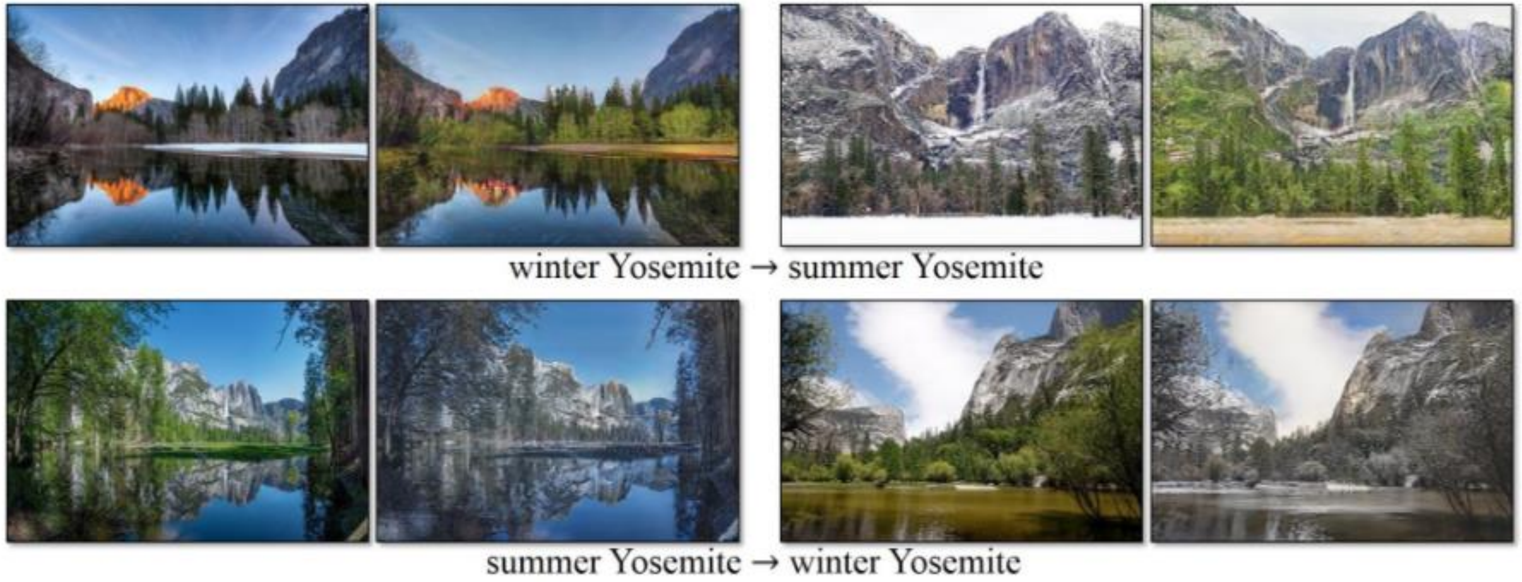
Generated by a machine

Generated by a machine based on given text

https://arxiv.org/pdf/1703.10717.pdf
https://arxiv.org/pdf/1710.10916.pdf

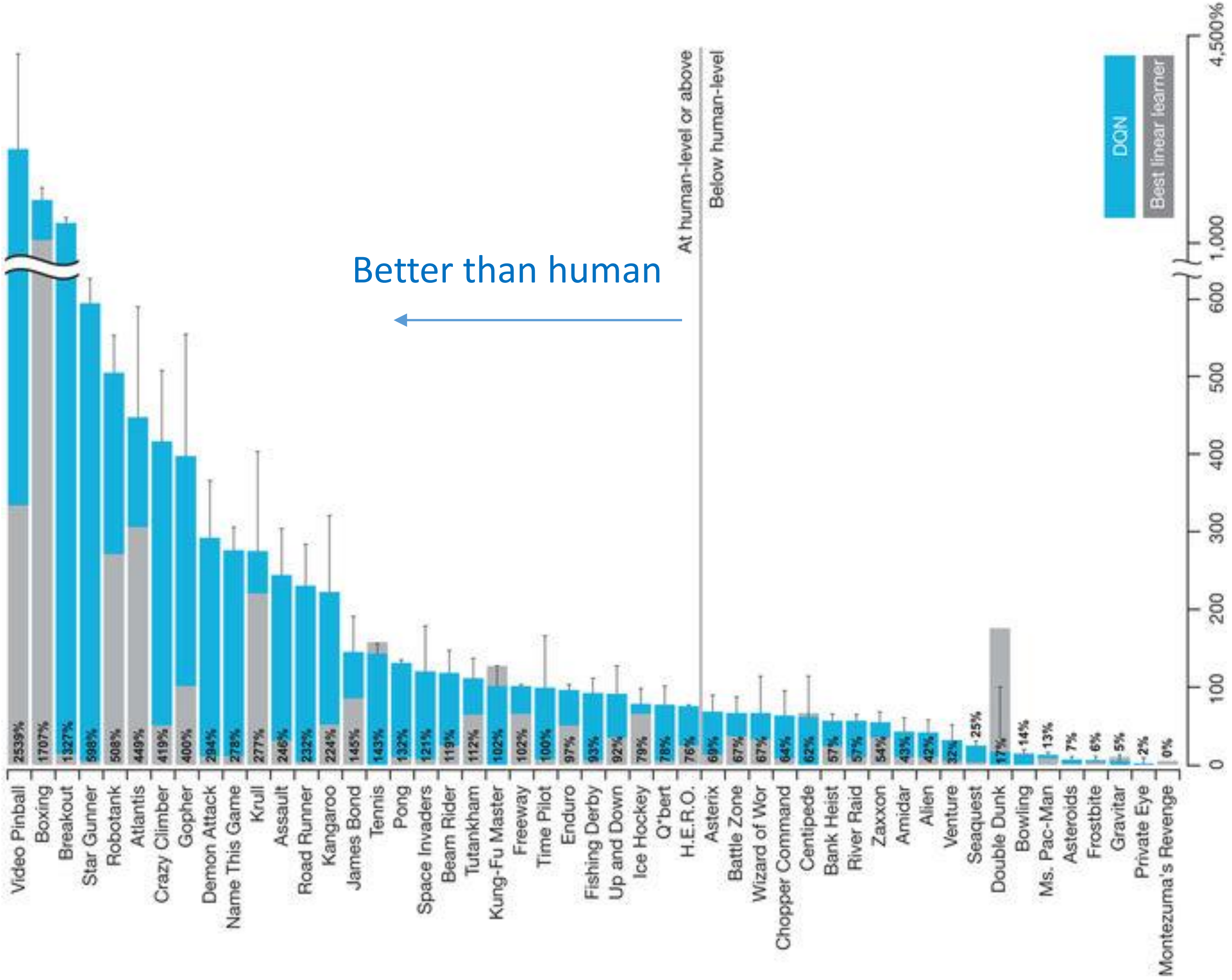
Generation: style transfer

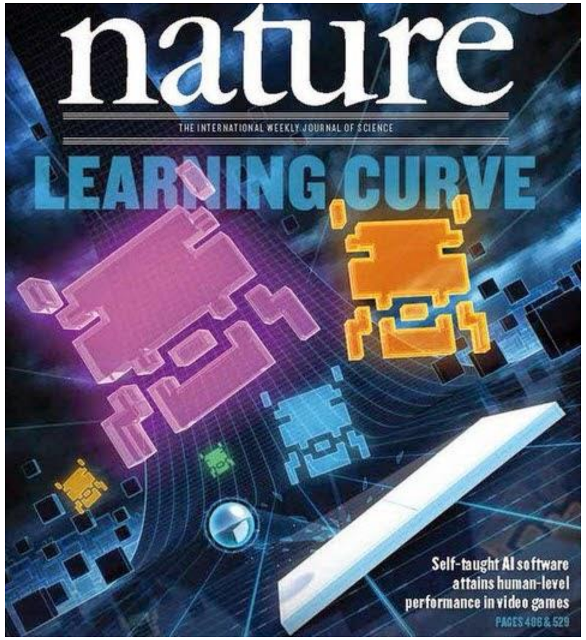
Season Transfer



https://github.com/junyanz/CycleGAN

Playing Atari game

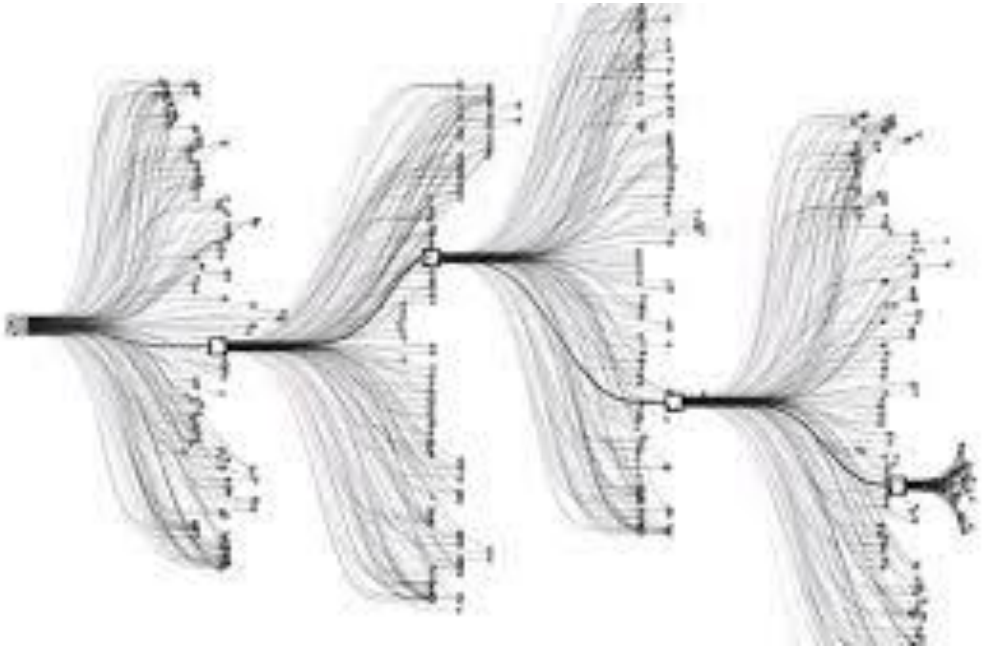




https://www.nature.com/articles/nature14236 nature, 2015

Alpha GO (2015)

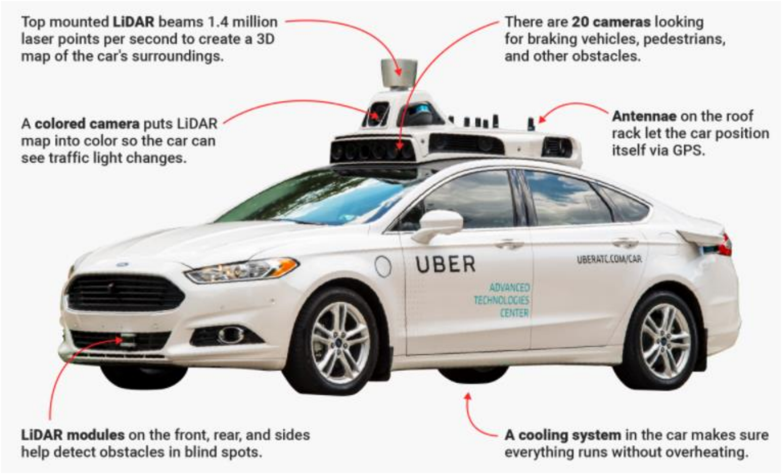
- ❑ Game Go: 10^{170} state space
- ❑ Beat European Champion: October 2015
- ❑ Beat World Champion: March 2016



https://deepmind.com/research/alphago/

Automatic Driving

- ❑ 1.3 million people die every year in car accidents.
- ❑ 94% of those accidents involve human error.
- ❑ 70% of the manned Taxis is related to labor cost.



https://newsroom.uber.com/us-pennsylvania/new-wheels/