NUS-ISS

Problem Solving Using Pattern Recognition

Deep learning: Before and After

Nicholas Ho
Lecturer & Consultant
Institute of System Science
National University of Singapore
nicholas.ho@nus.edu.sg

Topic for today's lesson:

What are the problems we can solve pertaining to pattern recognition using deep learning?

About Nicholas Ho

- nicholas.ho@nus.edu.sg
- Lecturer at NUS ISS; Courses covered include:
  - Intelligent Sensing and Sense Making
  - Autonomous Robots and Vehicles
  - Human-Robot System Engineering
- BEng and PhD degree from School of Mechanical Engineering, NUS
- Specialized in architecture, design & development
  - Artificial Intelligence
  - Augmented/Virtual Reality
  - Internet-of-Things (IoT) & Cyber-Physical System (CPS)

The Future Classrooms?

Source: https://www.youtube.com/watch?v=JMLsHI8aV0g

The Future Classrooms?

Deep Learning Examples

- Translates brain signals into an attention score
- Recognize faces to take attendance
- Recognize gestures (e.g. raising hands, chatting, using mobile phones)
- Quiz students via robots
- Generate and inform health status with sensory data

Source: https://www.wsj.com/articles/chinas-efforts-to-lead-the-way-in-ai-start-in-its-classrooms-11571958181

The Future Classrooms?

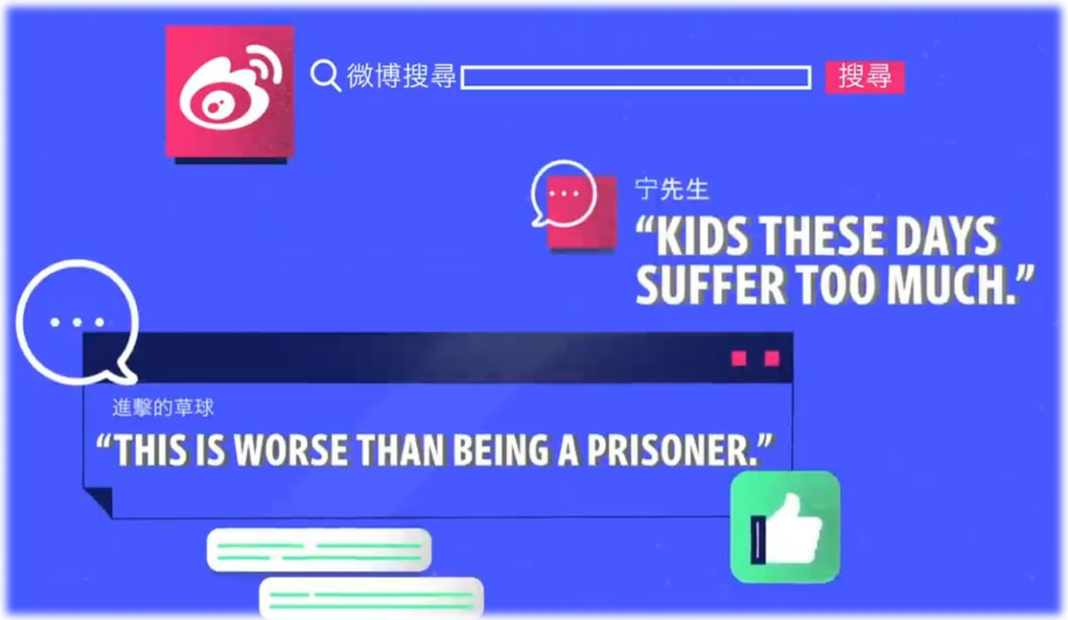
Satisfactory results

- Increased attention lead to improved grades
- Less work for teachers (automated attendance taking, less focus on observing and more focus on teaching)
- Monitoring by worried parents (some responsibility is shifted to the parents to discipline/ educate their own children instead of the teacher bearing full responsibility)

Source: https://www.wsj.com/articles/chinas-efforts-to-lead-the-way-in-ai-start-in-its-classrooms-11571958181

The Future Classrooms?

Issues of such applications



Source: https://www.wsj.com/articles/chinas-efforts-to-lead-the-way-in-ai-start-in-its-classrooms-11571958181

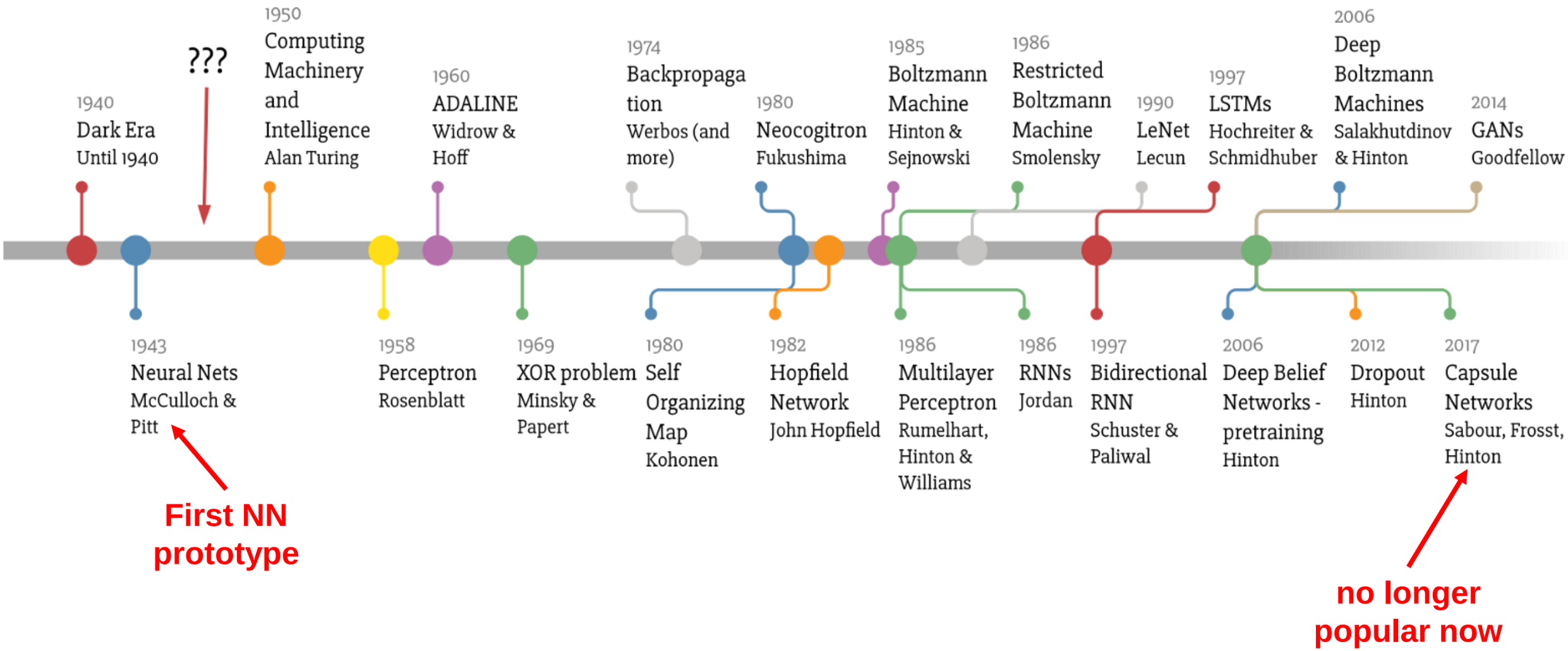
- No data privacy
- Chances of not being to obtain parent consent
- Technologies might not be accepted by society as a whole (undesirable social media feeds)
- Concerns that students will feel too pressured under surveillance

Deep learning: The Before

Time line

Of deep learning

Many methods will sink throughout until there is a convergence (i.e. standards produced)

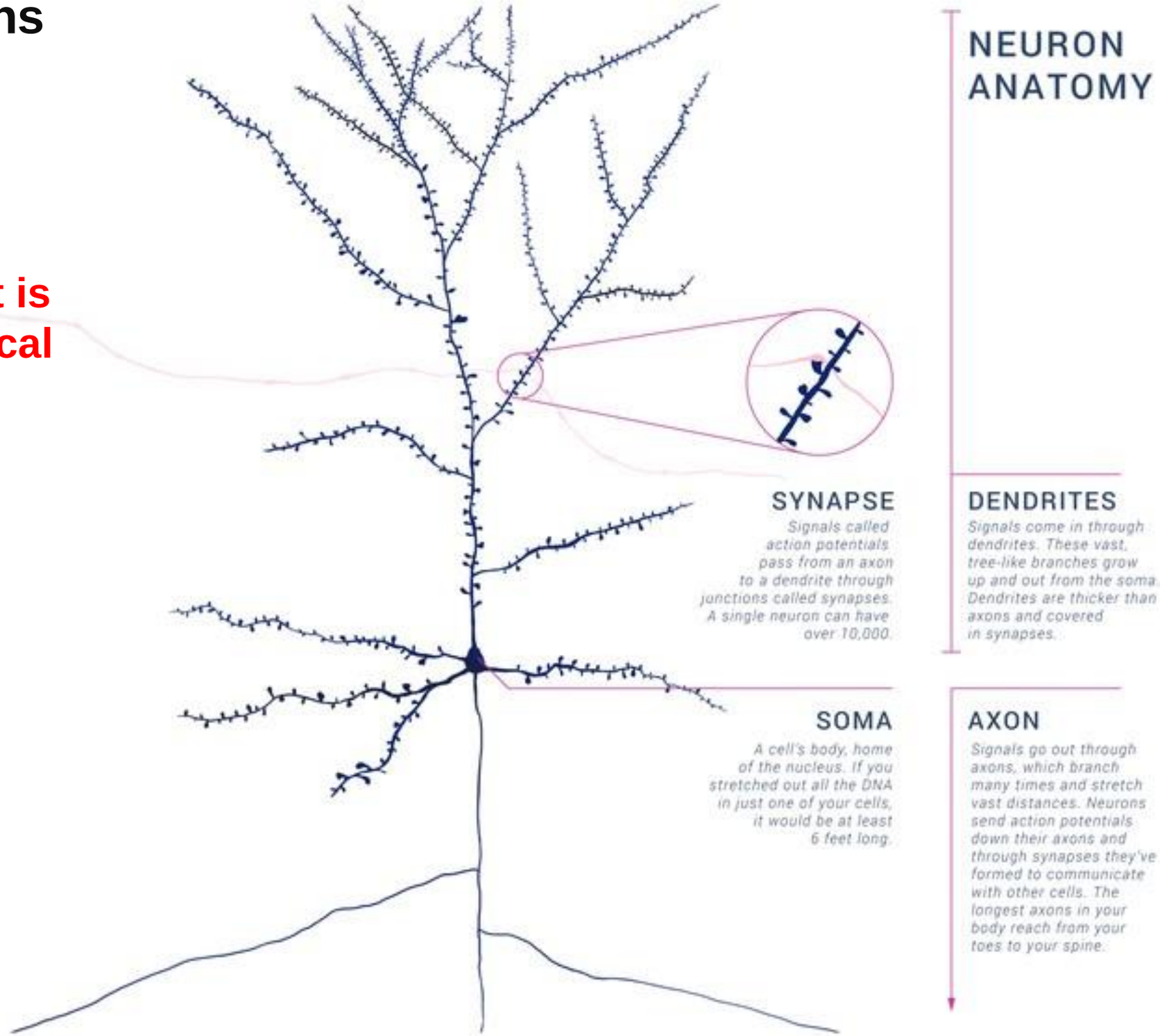


Source: https://medium.com/@faviovazquez

Biological neurons

Simplified illustration

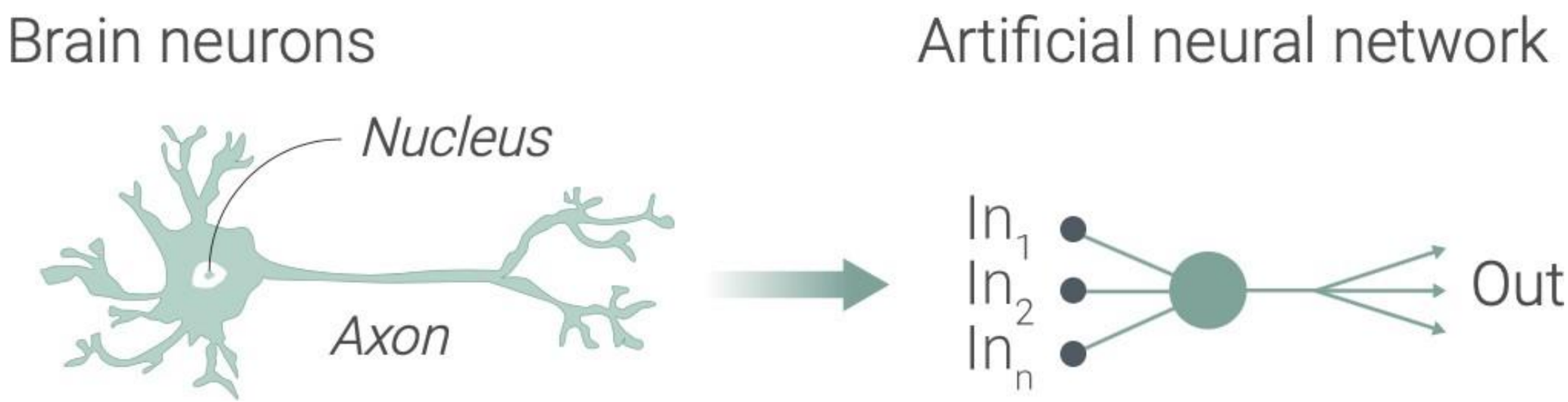
Neural Net concept is inspired by biological neurons!



Source: https://en.wikipedia.org/wiki/File:Anatomy_of_a_Neuron_with_Synapse.png

The evolution

From neurons to neural network



For ANN to be powerful, we will need more neurons, just like our brain!

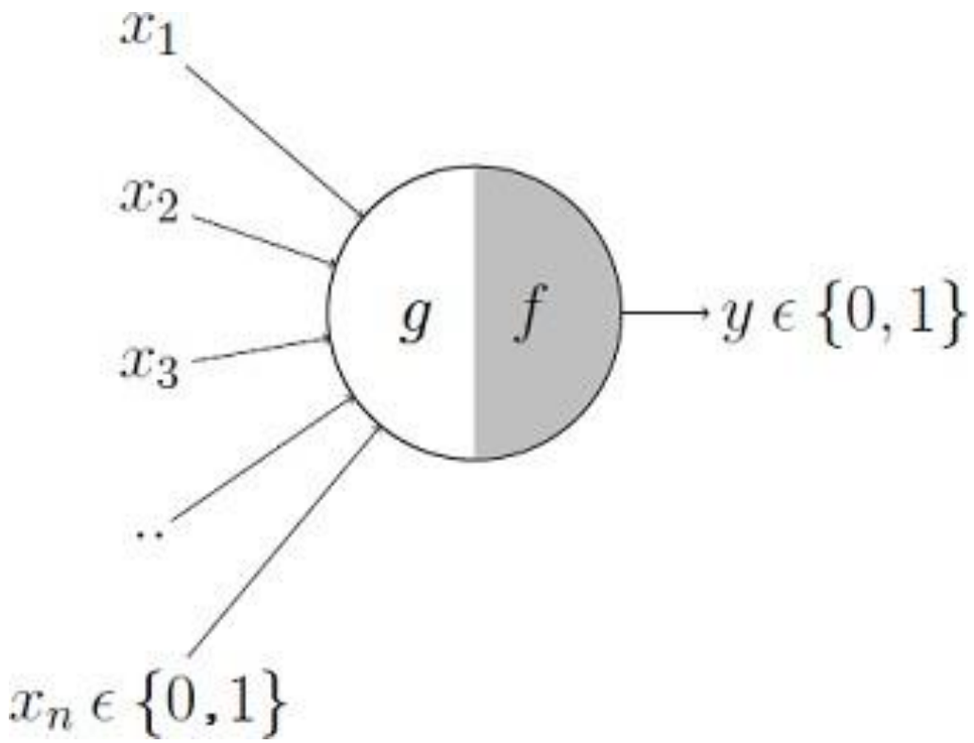
Source: https://multimedia.scmp.com/news/china/article/2166148/china-2025-artificial-intelligence/index.html?src=follow-chapter

The first artificial neuron

McCulloch-Pitts Neuron

- By Warren McCulloch (neuroscientist) and Walter Pitts (logician) in 1943

Note that the inputs (i.e. x_n) in here are only 0 or 1 (true or false)



$$g(x_1, x_2, x_3, ..., x_n) = g(\mathbf{x}) = \sum_{i=1}^{n} x_i$$

$$y = f(g(\mathbf{x})) = 1 \quad if \quad g(\mathbf{x}) \geq \theta$$
$$= 0 \quad if \quad g(\mathbf{x}) < \theta$$

g(x) = sum of all inputs
f(g(x)) = output defined by the function

Source: https://en.wikipedia.org/wiki/File:Anatomy_of_a_Neuron_with_Synapse.png

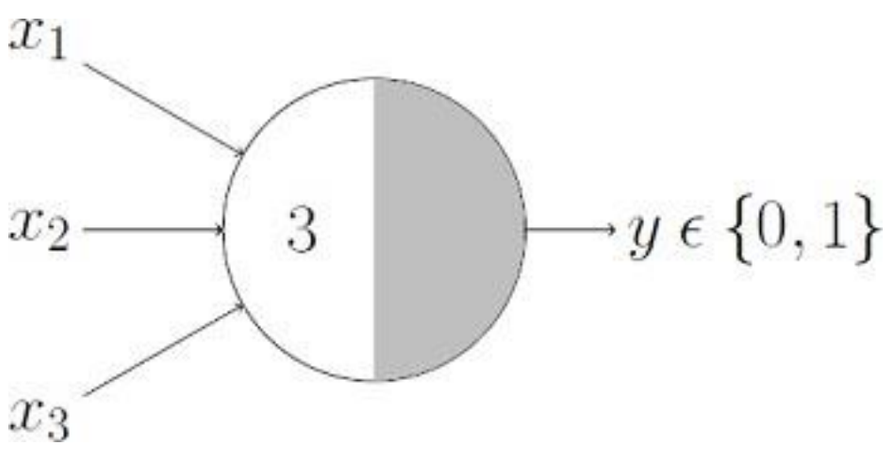
The first artificial neuron

McCulloch-Pitts Neuron

- Can be used to represent a few Boolean functions

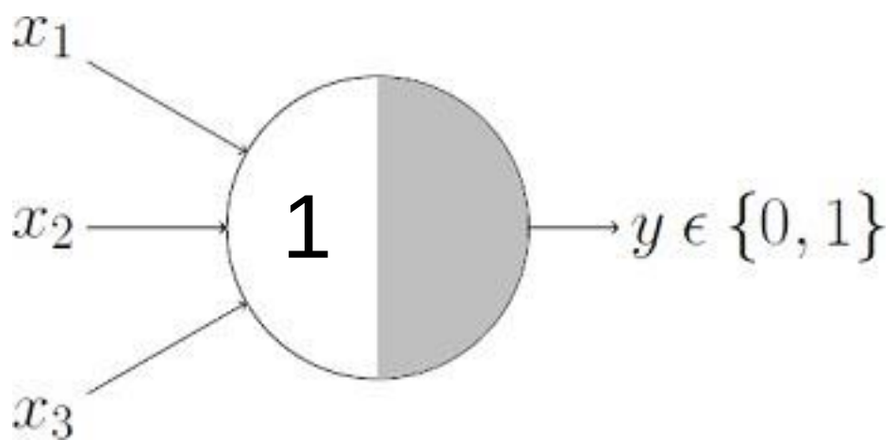
Interesting but NOT useful in real life application!

AND function



$y = 1$ if $x_1 + x_2 + x_3 \geq 3$

OR function



$y = 1$ if $x_1 + x_2 + x_3 \geq 1$

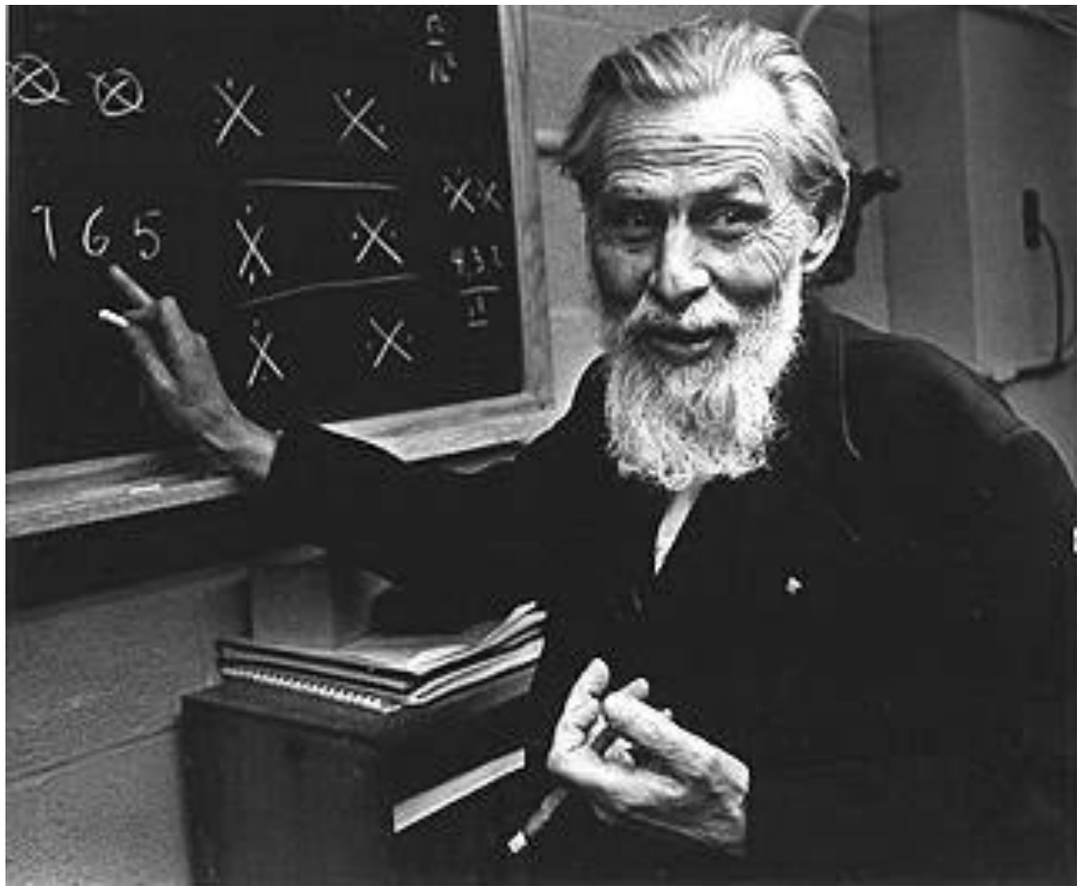
Source:https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1

The first artificial neuron

McCulloch-Pitts Neuron

- Inputs accepts only boolean values
- No learning algorithm to update the neurons

No learning = NOT true AI !!!

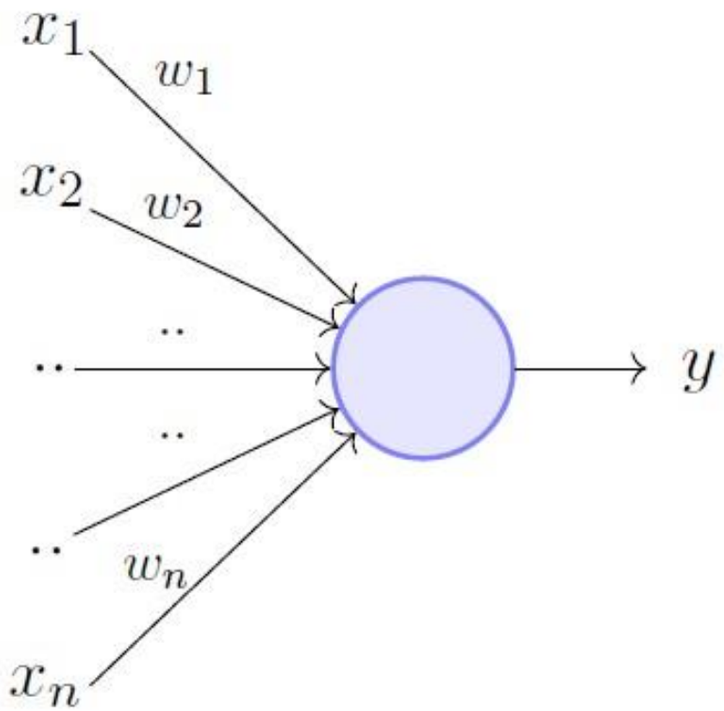


Left: Warren S. McCulloch. Right: Walter H. Pitts Jr.

The improved artificial neuron

Perceptron

w_n are the weights,
Θ is the threshold value



- By Frank Rosenblatt, refined by Minsky and Papert
- Support real inputs, not just boolean values

$$y = 1 \quad if \sum_{i=1}^{n} w_i * x_i \geq \theta$$

$$= 0 \quad if \sum_{i=1}^{n} w_i * x_i < \theta$$

Rewriting the above,

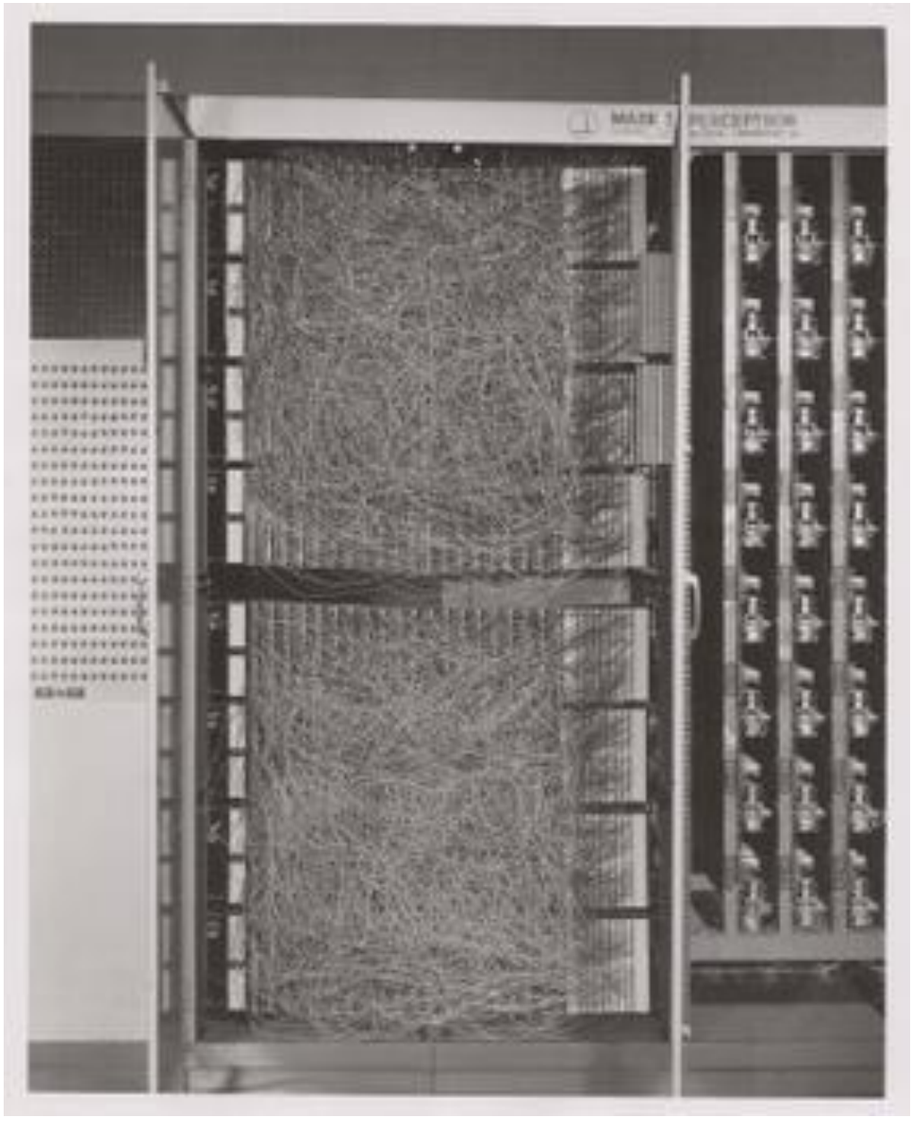
$$y = 1 \quad if \sum_{i=1}^{n} w_i * x_i - \theta \geq 0$$

$$= 0 \quad if \sum_{i=1}^{n} w_i * x_i - \theta < 0$$

Source: https://towardsdatascience.com/perceptron-the-artificial-neuron-4d8c70d5cc8d

The improved artificial neuron

Perceptron



Source: 'Mark I Perceptron at the Cornell Aeronautical Laboratory', hardware implementation of the first Perceptron (Source: Wikipedia / Cornell Library)

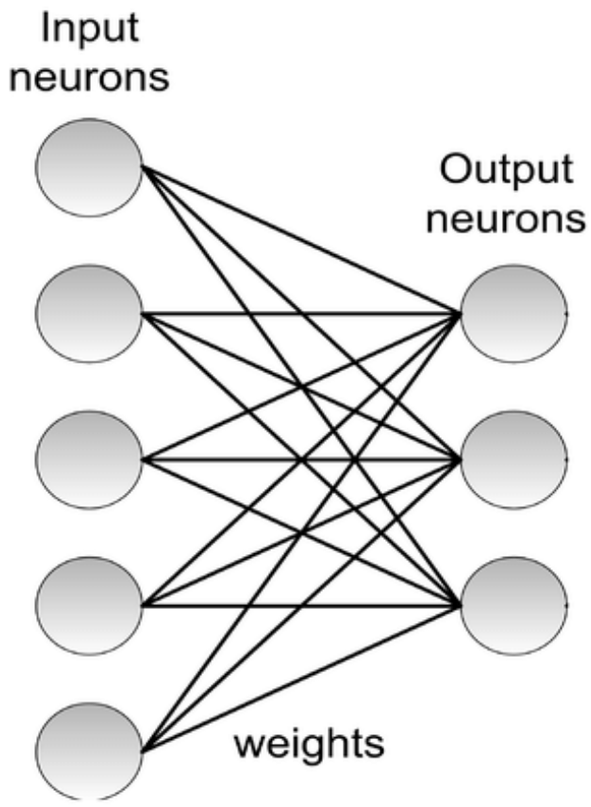
- Rosenblatt's achievement: artificial neurons could actually learn from data
- He came up a supervised learning algorithm!
- He implemented Perceptron in custom hardware, which can learn to classify simple shapes correctly with 20x20 pixel-like inputs

The perceptron method can learn from past data!

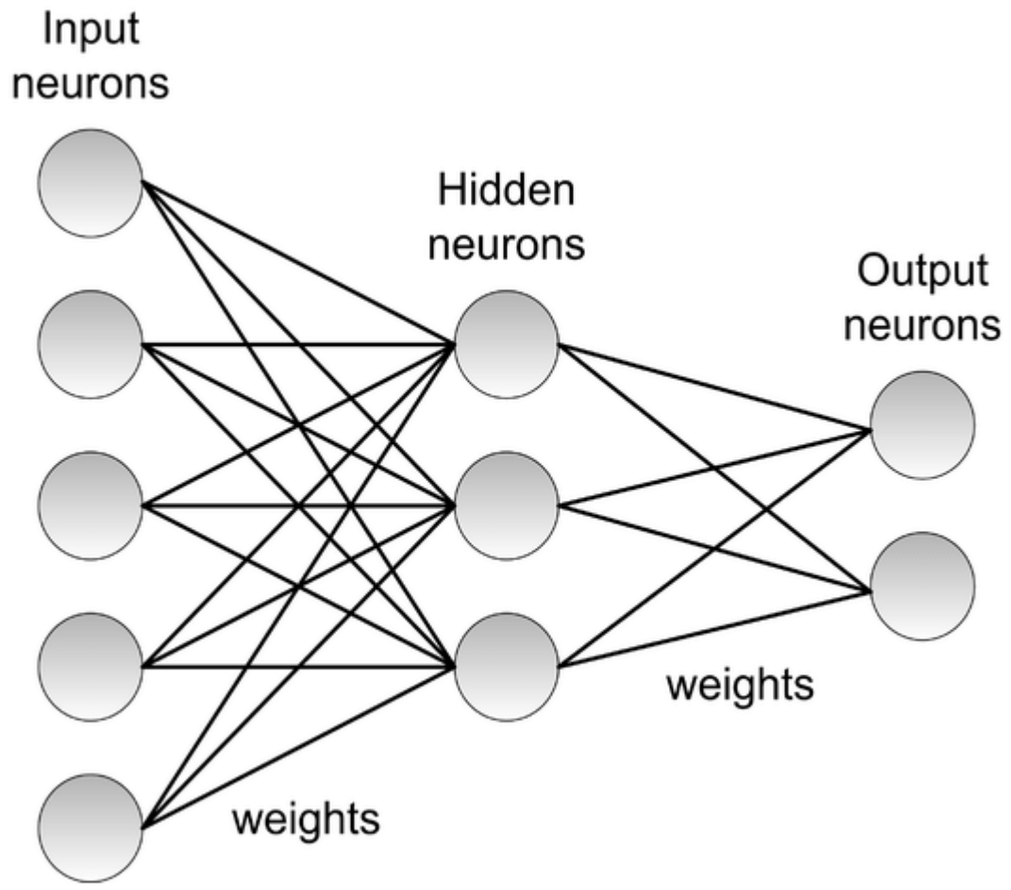
The improved artificial neuron

Perceptron

Note that the neurons are aligned into a single line for each layer (can be represented with a 1D array)



Single layer perceptron



Multi-layer perceptron

Source: 'Mark I Perceptron at the Cornell Aeronautical Laboratory', hardware implementation of the first Perceptron (Source: Wikipedia / Cornell Library)

The first AI winter

XOR affair

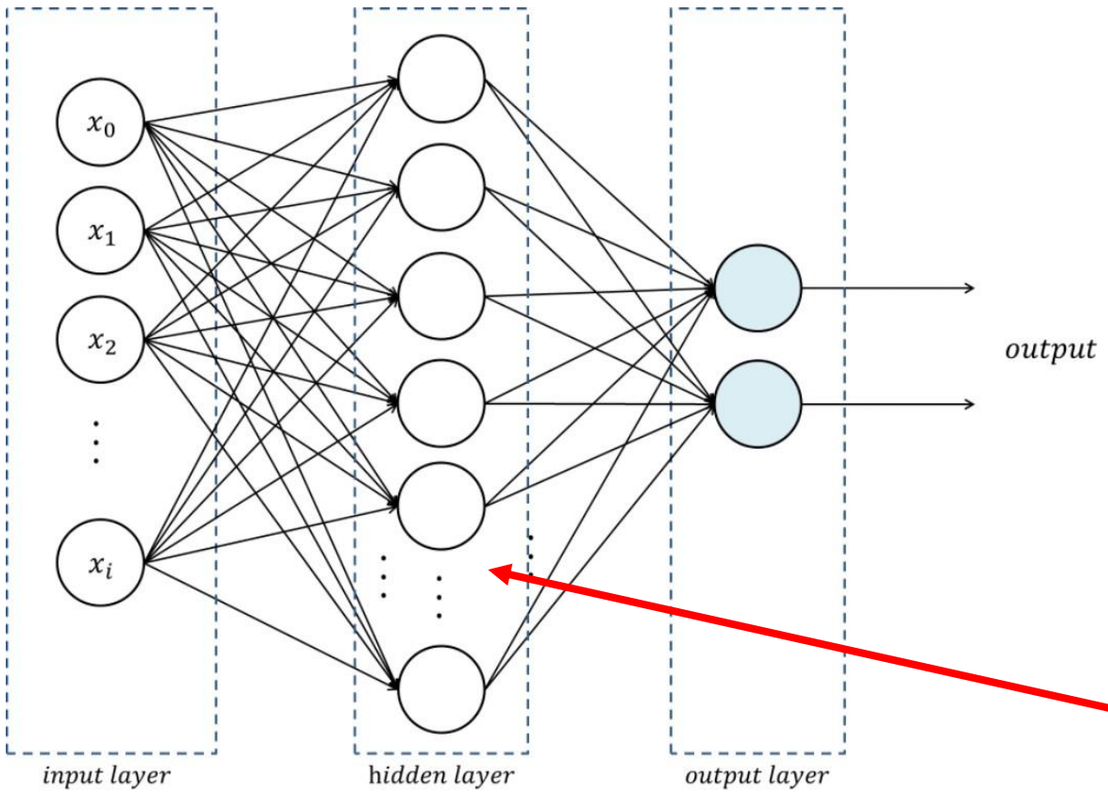
Source: https://amethix.com/2018/06/ai-winter-is-coming/

- Marvin Minsky, founder of MIT AI lab, and Seymour Papert, director of the lab in 1969 published a book named 'Perceptrons'
- They showed that a single perceptron cannot do XOR
- Multiple layers of Perceptron (aka MLP) can do XOR, but the proposed learning algorithm does not work for that!
- Here comes the winter ...

Where is the learning algorithm?

The thaw of AI winter

Backpropagation = a learning algorithm that makes the Neural Network (NN) learn!



Source: https://www.cc.gatech.edu/~san37/post/dlhc-fnn/

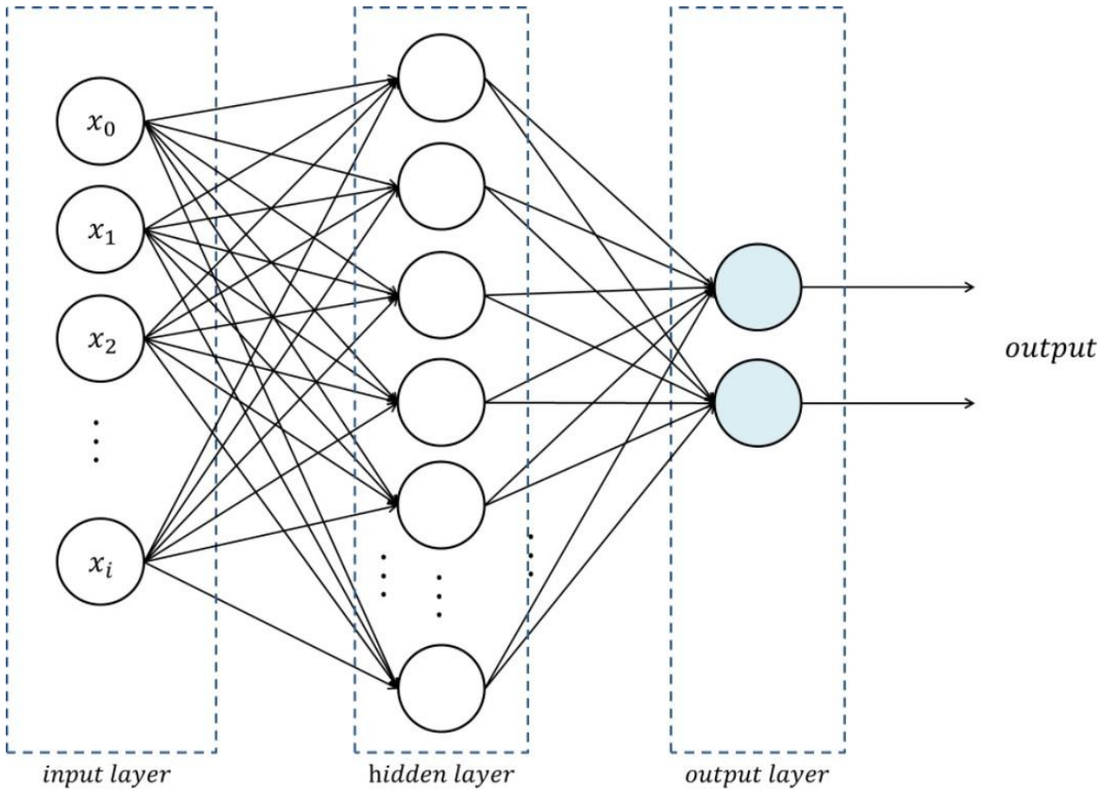
- Multilayer layers of perceptron should work, but need learning algorithm
- Between 60s and 80s, several researchers separately derived the solution, but few people knew
- In 1986, Rumelhart, Hinton Williams published a method in Nature
- They called the learning procedure "backpropagation"

Can have more than 1 hidden layer; more layers = deep NN

Where is the learning algorithm?

The thaw of AI winter

Question:
How can we fit an image (square matrix) into a NN?



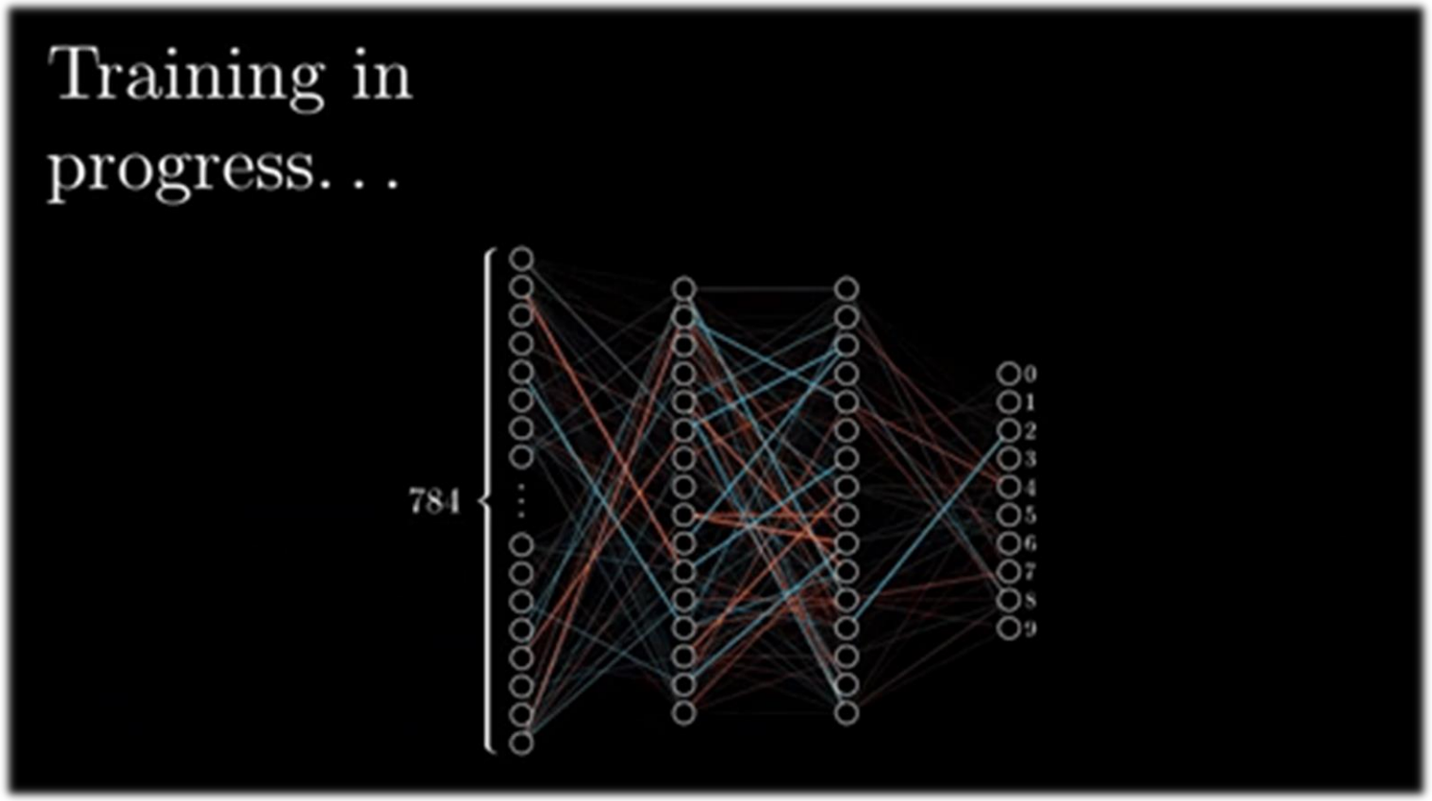
Source: https://www.cc.gatech.edu/~san37/post/dlhc-fnn/

- Multilayer layers of perceptron should work, but need learning algorithm
- Between 60s and 80s, several researchers separately derived the solution, but few people knew
- In 1986, Rumelhart, Hinton Williams published a method in Nature
- They called the learning procedure "backpropagation"

Where is the learning algorithm?

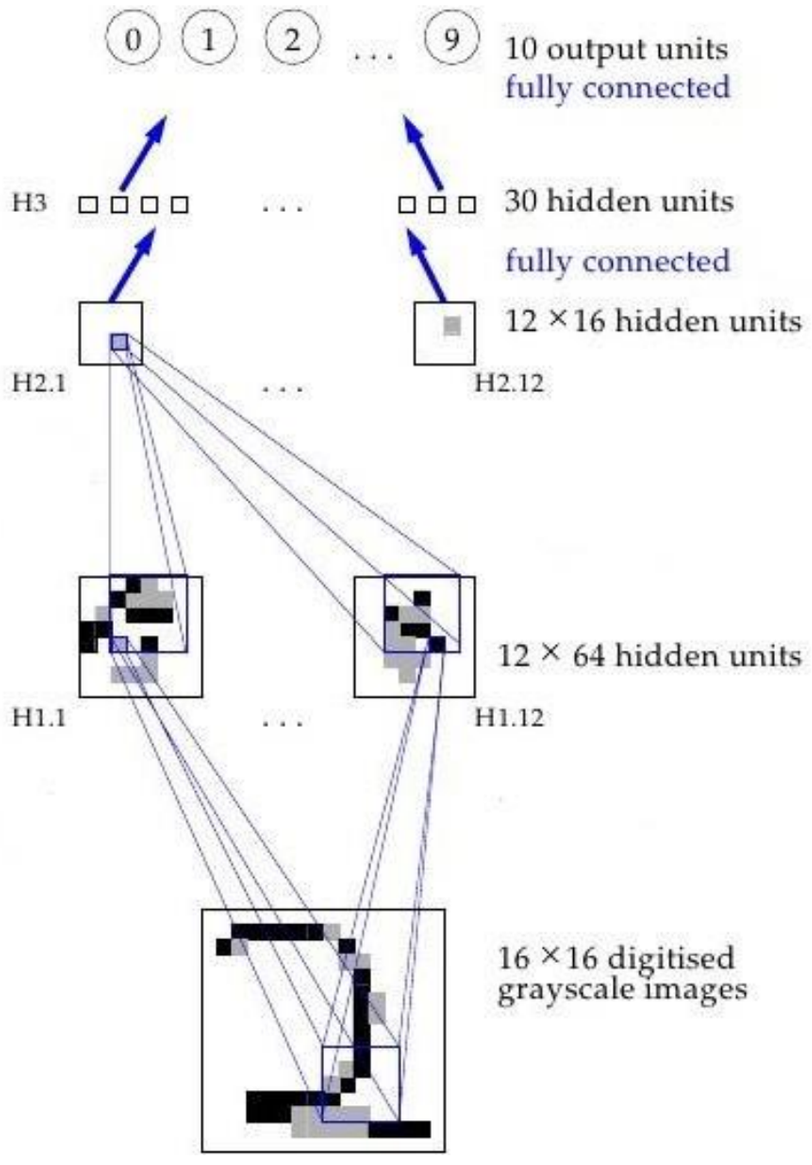
The thaw of AI winter

Visual Illustration of Backpropagation Process:



Year 1989

Annus mirabillis, sort of



Source: http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning-part-2/

- Multilayer feedforward networks are proved to be universal approximators (i.e. able to approximate any functions)
- LeNet was proposed and put into actual significant use: recognizing numbers
- It is a neural net + convolutional layers (weight sharing)

Convolutions = performing filtering

Another winter dawns

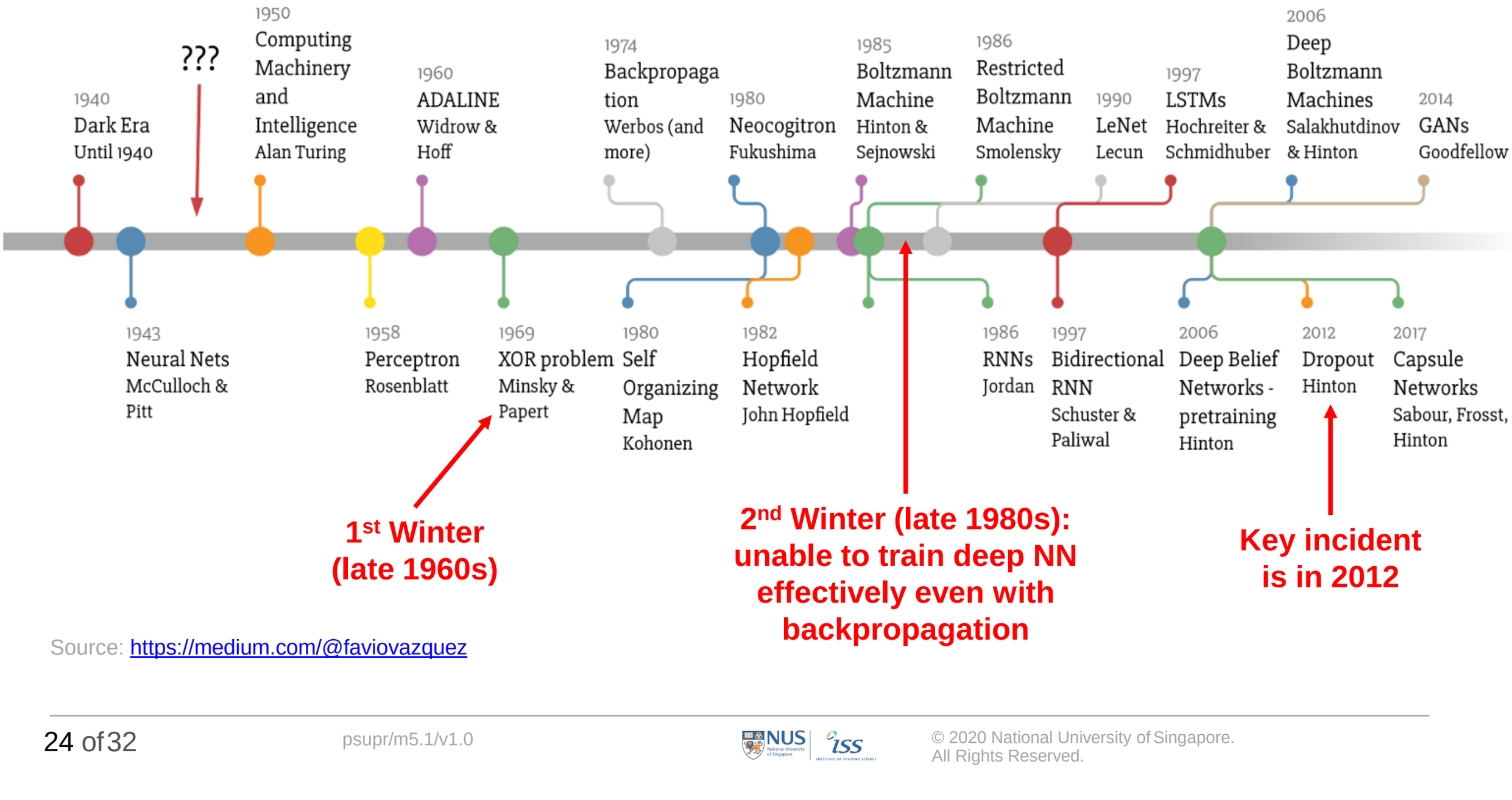
by backpropagation

Source: https://amethix.com/2018/06/ai-winter-is-coming/

- By late 80s, already knew that deep neural net was hard to train (i.e. net unable to coverage / no improvement in the classification results / the net cannot conclude anything)
- Deep neural nets trained with backpropagation did not work very well, not as well as nets with fewer layers
- Support vector machine (SVM) came into fashion
- Random forests, with lovely mathematical theory behind proved to be effective

Time line

Of deep learning



Source: https://medium.com/@faviovazquez

ImageNet

by Stanford Vision Lab

1.2 million images, 1000 object categories

Source: https://gluon-cv.mxnet.io/build/examples_datasets/imagenet.html

Competition
on ImageNet

1.2 million images:
- **1 million for training**
- **150k for testing**
- **50k for validation (final testing)**

- ILSVRC: ImageNet Large Scale Visual Recognition Challenge
- Started from 2010; teams evaluate their algorithms on given data set, compete to achieve highest accuracy on visual recognition tasks.
- ILSVRC training dataset: 1000 object categories, 1.2 million images

IMAGENET

Source: http://www.image-net.org

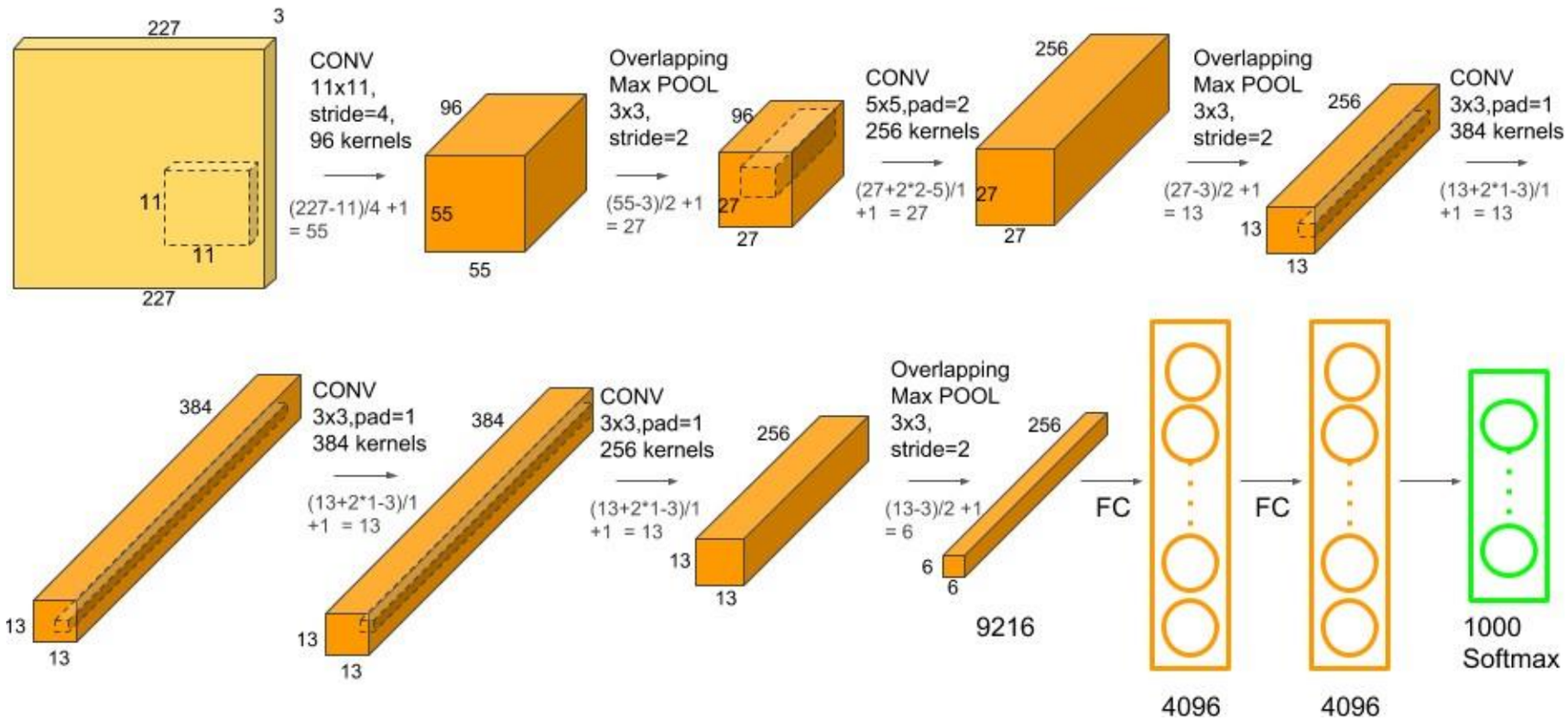
ILSVRC2012

AlexNet

- Team Hinton entered the competition

AlexNet is a Convolutional Neural Network (CNN) model; similar to LeNet

- They achieved an error rate of 15.3%, far far better than the next closest: 26.2%



Where are the input, CNN, hidden and output layers?

Source: https://neurohive.io/en/popular-networks/alexnet-imagenet-classification-with-deep-convolutional-neural-networks/

AlexNet

Key of success?

- The net structure is not the most important point
- Use of Rectified Linear Unit (ReLU) activation function
- Use of dropout
- GPU implementation (through CUDA); 20x time difference!

Activation function decides, whether a neuron should be activated or not, by calculating the weighted sum and further adding bias with it

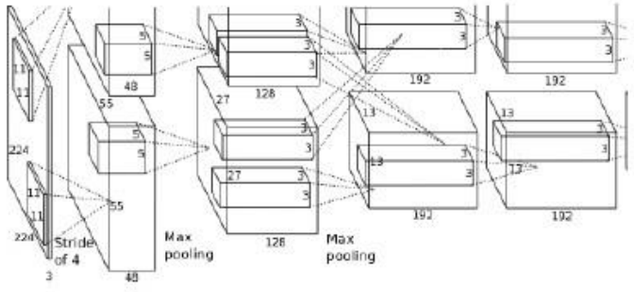
The main purpose of the activation function is to introduce non-linearity into the output of a neuron

Why need non-linear activation function ???

Deep learning computer vision recipe


Lots of data

+

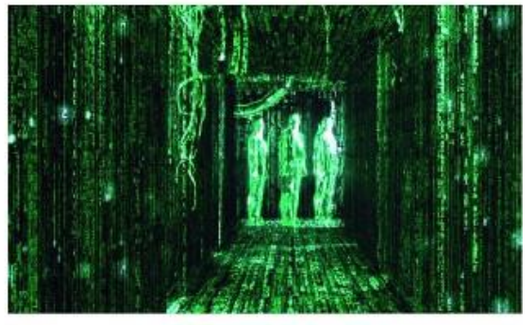


Deep convolutional neural network

+

Backprop on GPU

=


Vision

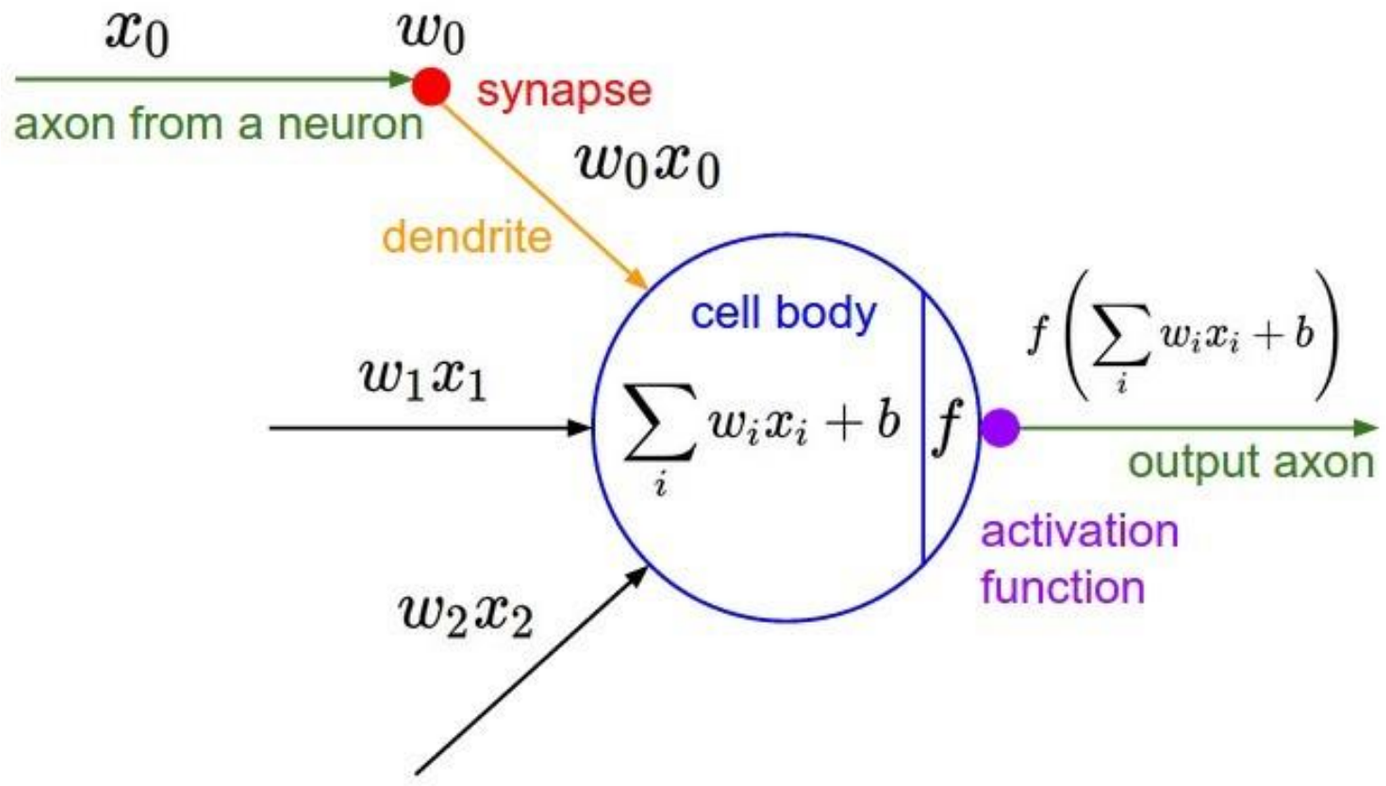
Source: http://www.computervisionblog.com/2015/05/deep-learning-vs-big-data-who-owns-what.html

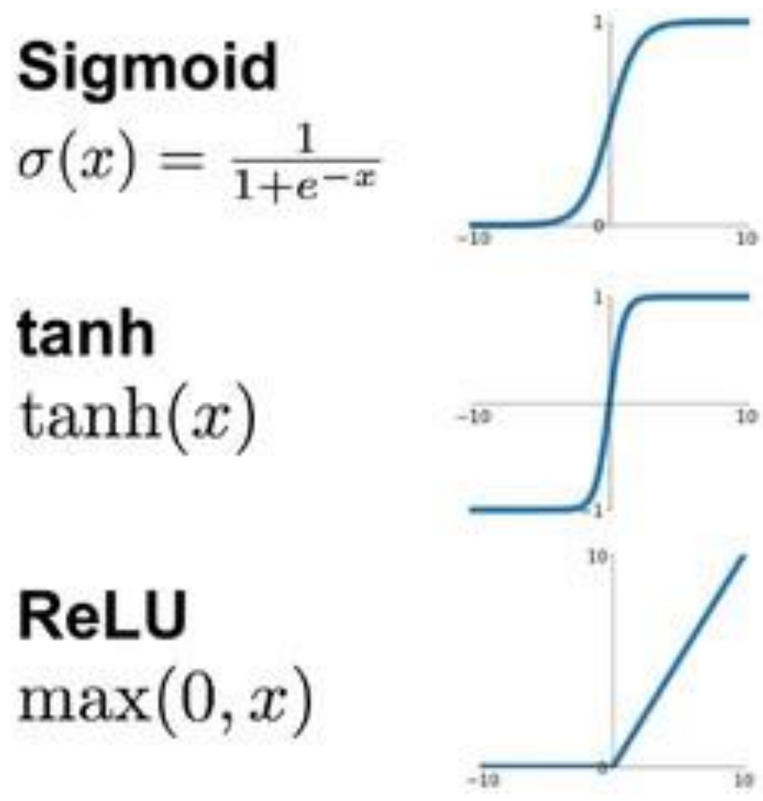
ReLU, as compared to others

Simple is better

ReLU is the fastest, simplest and most efficient as compared to the other 2 methods; now the default for activation function

- The zero value output from ReLU introduces sparsity representation (more zeros in each layer, only the important neurons contribute)
- ReLU is easier to calculate

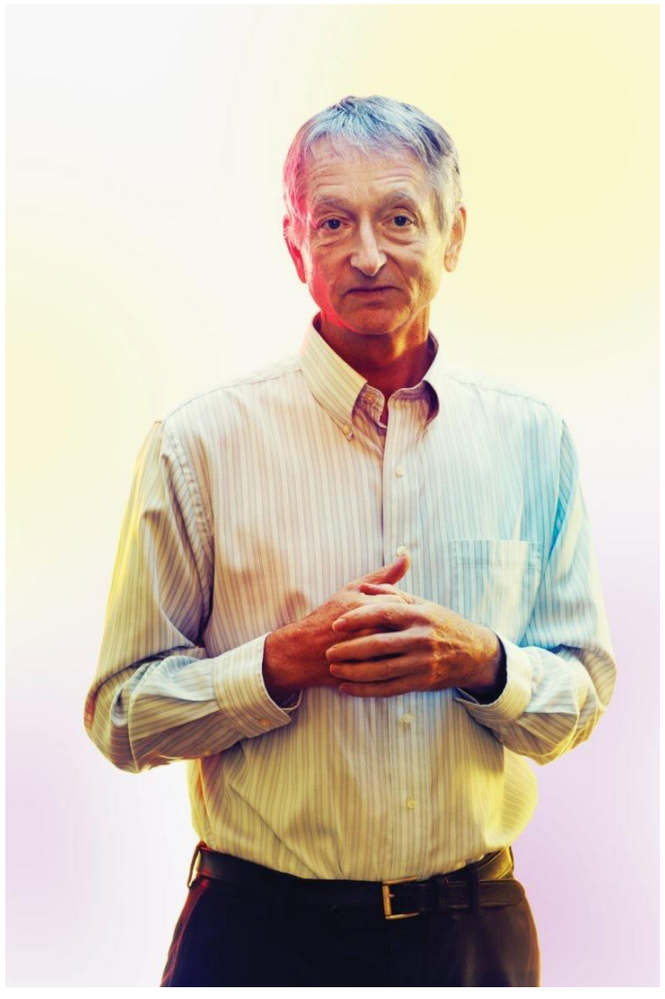




Source: http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/

The big question

Why did backpropagation fail in the past?



Source: https://torontolife.com/tech/ai-superstars-google-facebook-apple-studied-guy/

- Labelled datasets were thousands times too small; deep learning applications generally require a lot of datasets!

- Computing power was millions times slower back then; computing is now much faster as compared to the past

- Initialized the weight in stupid ways; weights are all set to 0 or 1 in the past

- Used the wrong type of non-linearity for activation function

Since then ...

Neural network Library

theano

Keras

Caffe

DEEPLEARNING4J

TensorFlow ← Machine Learning Libraries → PYTORCH

mxnet

The Microsoft Cognitive Toolkit

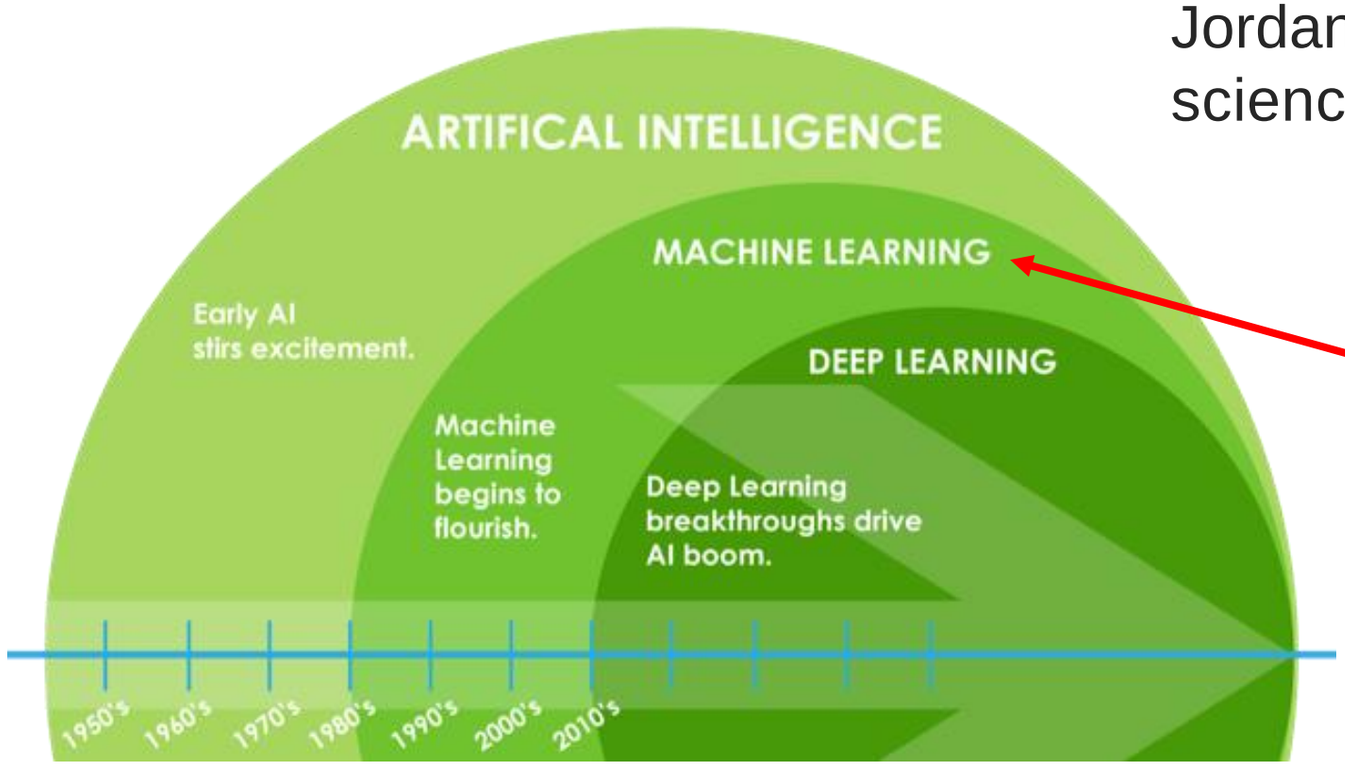
Chainer

Deep learning: The After

Confusion?

The differences among the few terms

- Arthur Samuel coined the term 'machine learning' in 1959 while at IBM
- Rina Dechter introduced the term deep learning in 1986
- Machine learning and statistics are closely related, thus Michael I. Jordan suggested the term 'data science' to refer to the overall field



Covered in past few modules!

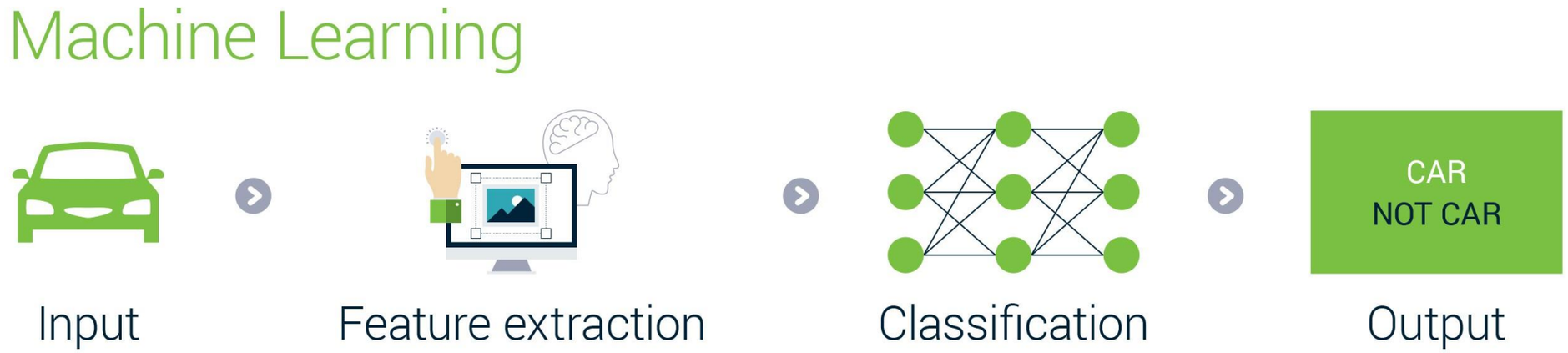
Source: https://buzzrobot.com/difference-between-artificial-intelligence-machine-learning-and-deep-learning-ccfd779eca7b

Confusion?

The differences among the few terms

- Feature: a number or a vector that describes something about the input

What is the main difference between machine learning and deep learning ???

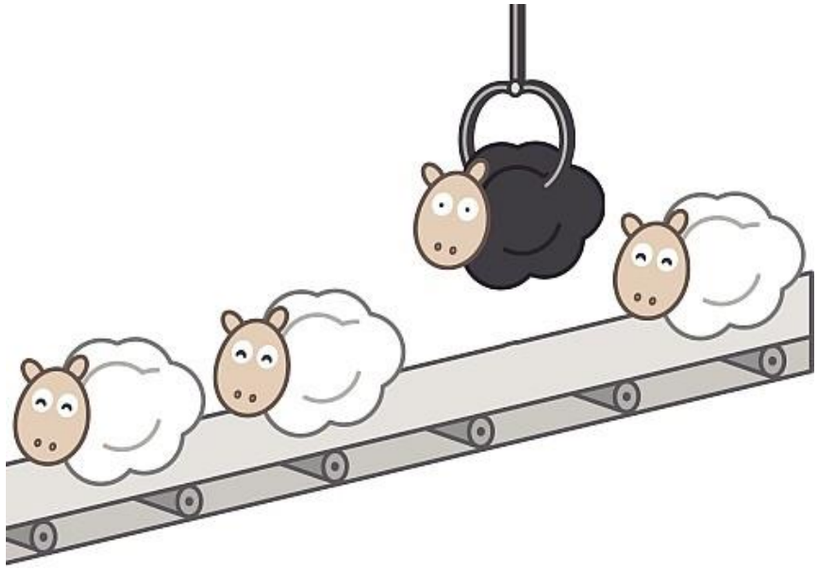


Source: https://verhaert.com/difference-machine-learning-deep-learning/

Application

Three main categories

Identify

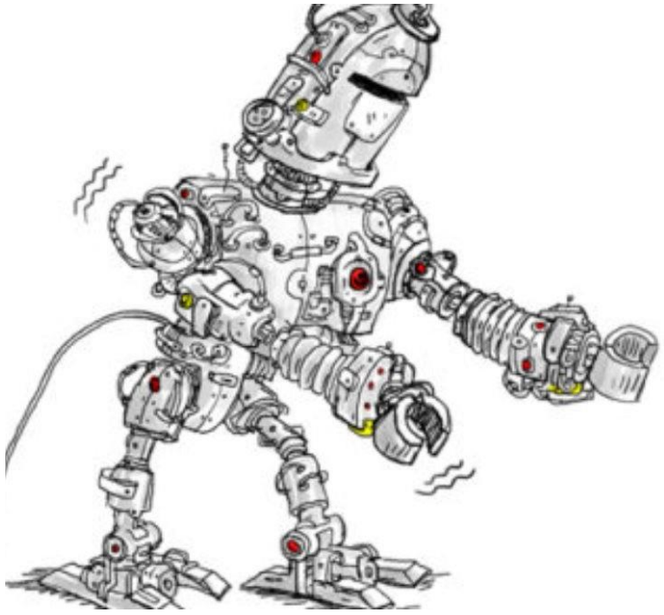


Source: http://blog.ss8.com

Create

Source: http://www.williammalone.com/articles/create-html5-canvas-javascript-drawing-app/

Act



Source: https://bitsandatoms.co/tag/reinforcement-learning/