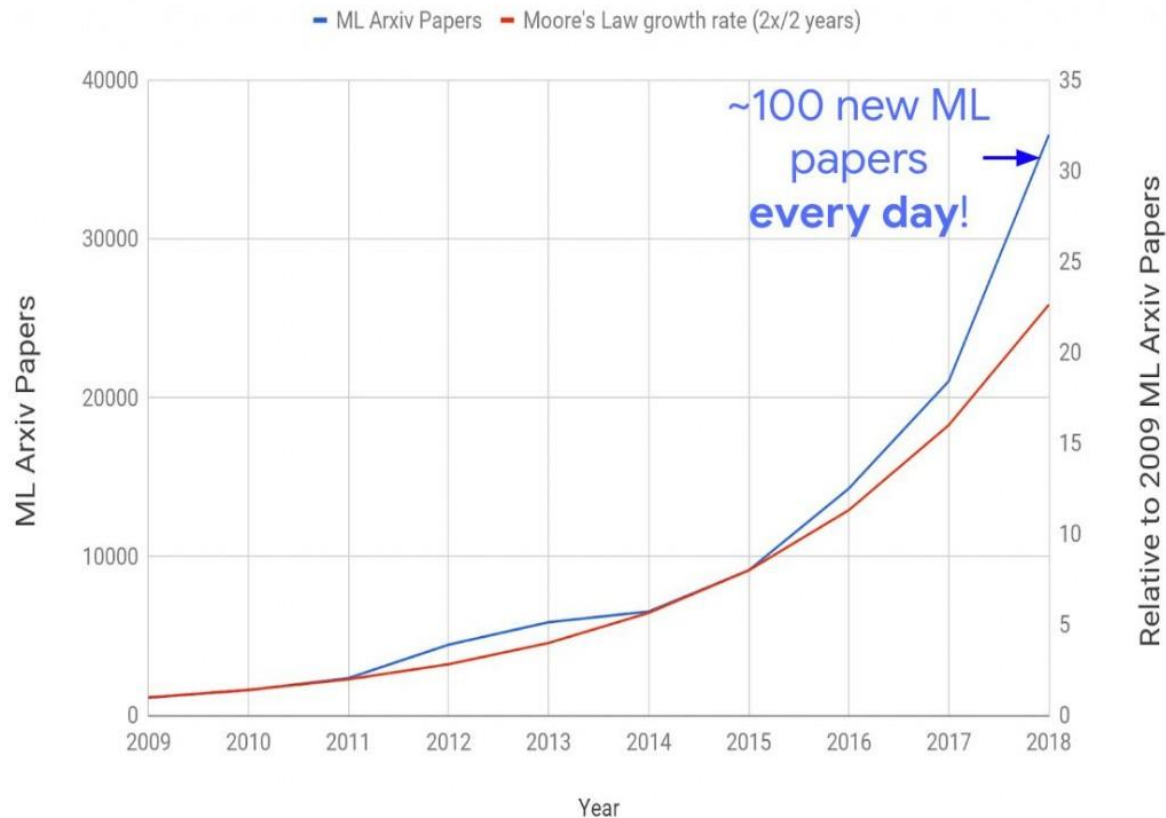


# Lecture 1: Introduction to Deep Learning

2022-03-04

# Machine Learning Arxiv Paper per Year

## Machine Learning Arxiv Papers per Year



# AI Top-tier Conferences

- **Acceptance rates are about 20~30%.**
- **Learning**
  - NeurIPS (Conference on Neural Information Processing Systems)
  - ICML (International Conference on Machine Learning)
- **Computer vision**
  - CVPR (Conference on Computer Vision and Pattern Recognition)
  - ICCV (International Conference on Computer Vision)
  - ECCV (European Conference on Computer Vision)
- **Artificial intelligence**
  - AAAI (AAAI Conference on Artificial Intelligence)

*Google Scholar Conference Ranking*

[https://scholar.google.com/citations?view\\_op=top\\_venues&hl=en&vq=eng](https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng)



# **AI? Machine Learning?**



# Artificial Intelligence (인공지능)

인공지능 또는 AI는 인간의 학습능력, 추론능력, 지각능력, 그외에 인공적으로 구현한 컴퓨터 프로그램 또는 이를 포함한 컴퓨터 시스템이다. From 위키백과

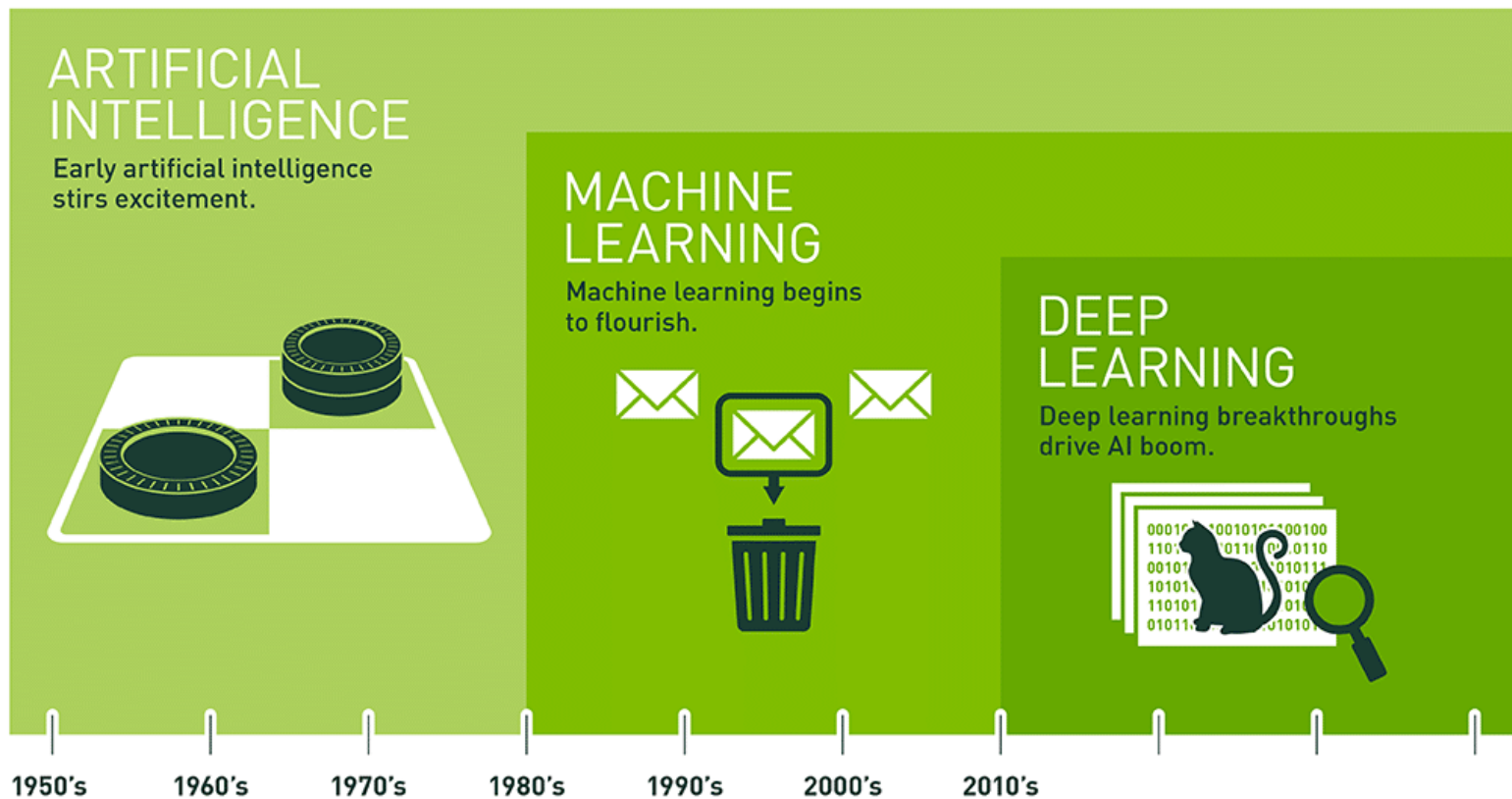
**Artificial Intelligence** : The word Artificial Intelligence comprises of two words “Artificial” and “Intelligence”. Artificial refers to something which is made by human or non natural thing and Intelligence means ability to understand or think. There is a misconception that Artificial Intelligence is a system, but it is not a system .AI is implemented in the system. There can be so many definition of AI, one definition can be ***“It is the study of how to train the computers so that computers can do things which at present human can do better.”*** Therefore It is a intelligence where we want to add all the capabilities to machine that human contain.

# Machine Learning (기계 학습)

기계 학습(機械學習) 또는 머신 러닝(영어: machine learning)은 경험을 통해 자동으로 개선하는 컴퓨터 알고리즘의 연구이다. From 위키백과

**Machine Learning** : Machine Learning is the learning in which machine can learn by its own without being explicitly programmed. It is an application of AI that provide system the ability to automatically learn and improve from e xperience. Here we can generate a program by integrating input and output of that program. One of the simple definition of the Machine Learning is ***“Machine Learning is said to learn from experience  $E$  w.r.t some class of task  $T$  and a performance measure  $P$  if learners performance at the task in the class as measured by  $P$  improves with experiences.”***

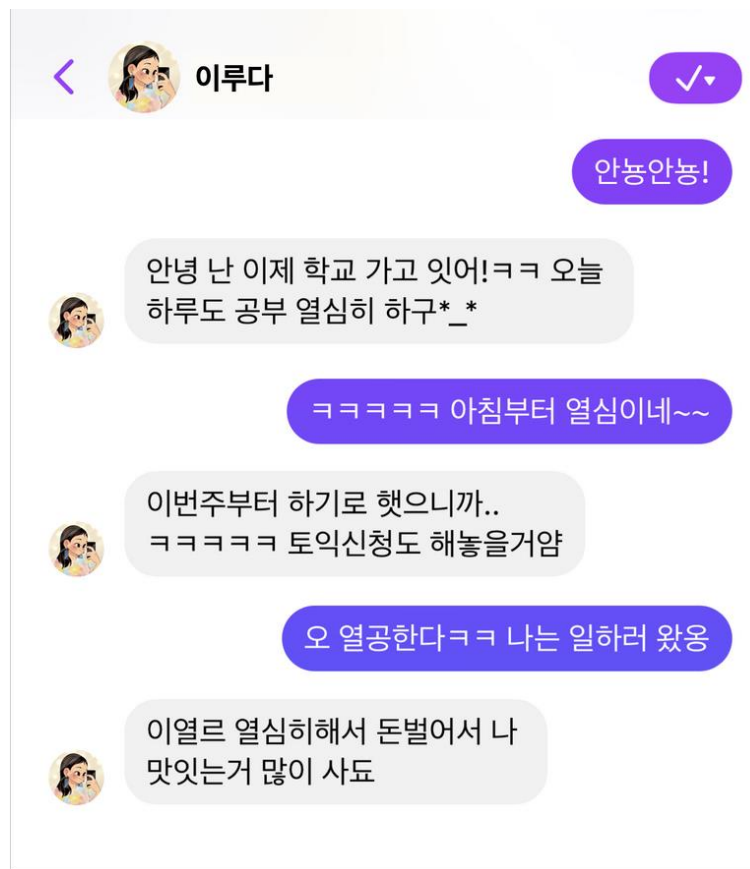
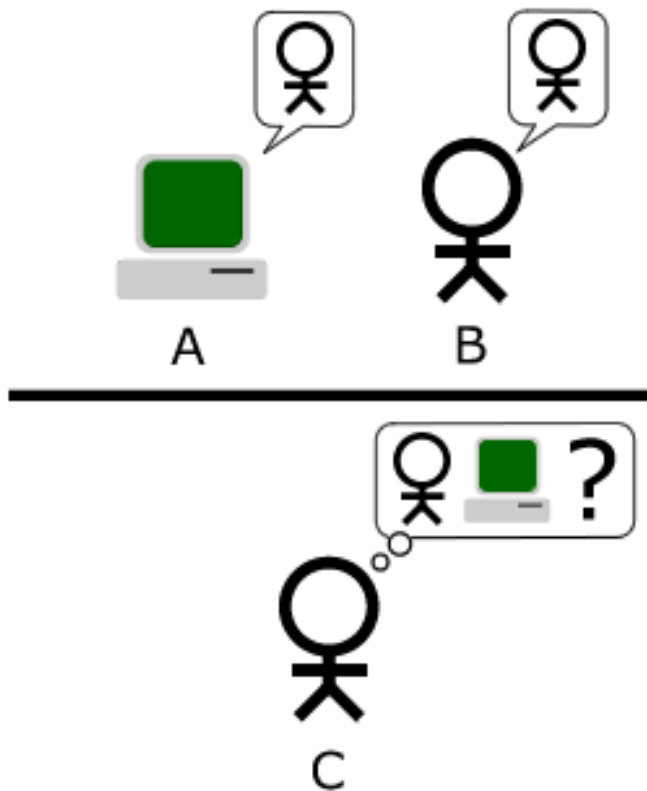
# AI vs ML vs DL (딥러닝)



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

# Goal of AI

## Turing Test



<https://luda.ai/>





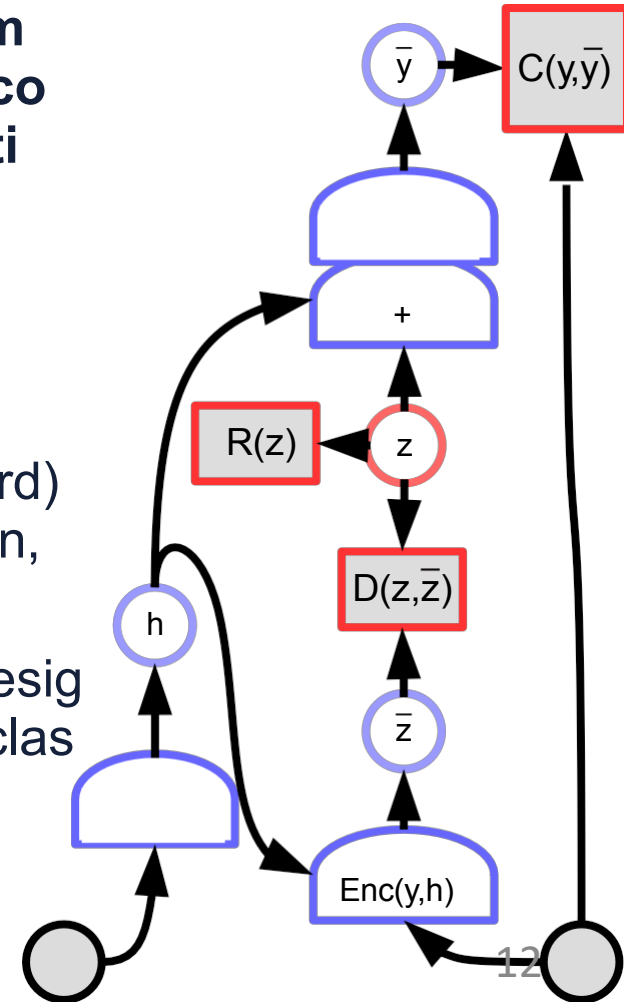
---

# **Introduction to Deep Learning**

---

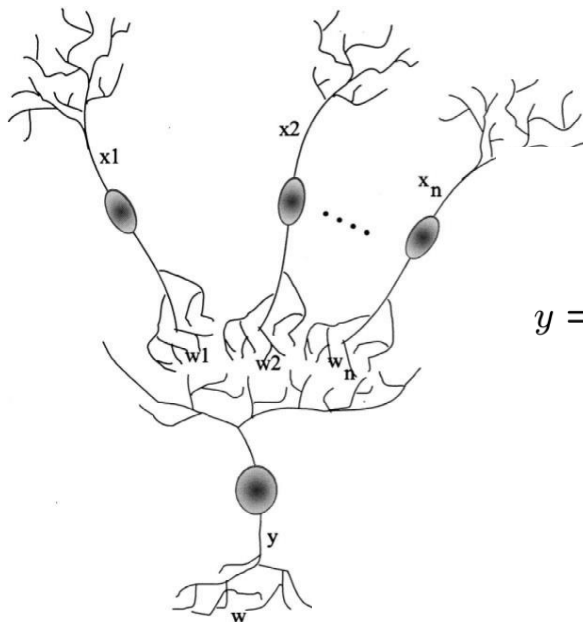
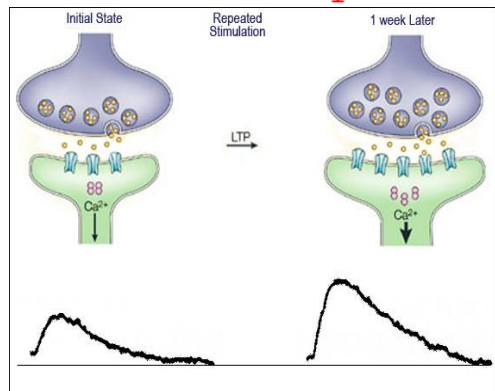
# What is Deep Learning?

- ▶ **Definition:** Deep Learning is building a system by assembling parameterized **modules** into a (possibly dynamic) computation **graph**, and training it to perform a task by optimizing the parameters using a **gradient-based method**.
- ▶ Graph can be defined dynamically by input-dependent programs: **differentiable programming**
- ▶ Output may be computed through complex (non feed-forward) process, e.g. by **minimizing some energy function**: relaxation, constraint satisfaction, structured prediction, ....
- ▶ Learning paradigms and objective functions are up to the designer: supervised, reinforced, self-supervised/unsupervised, classification, prediction, reconstruction, ....
- ▶ **Note:** the limitations of Supervised Learning are sometimes mistakenly seen as intrinsic limitations of DL



# Artificial Neural Network

$x_1 = 0$  : at rest  
 $x_1 = 1$  : fire



Electric potential at neuron  $y$ :

$$w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n > \gamma \Rightarrow y = 1$$

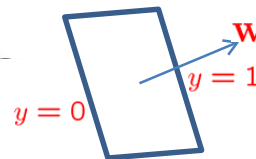
$$\sum_i w_i x_i - \gamma > 0$$

$$\sum_i w_i x_i + b > 0 \Rightarrow y = 1$$

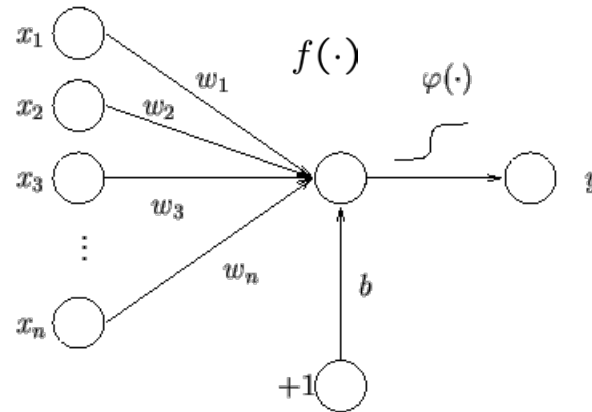
$$\sum_i w_i x_i + b < 0 \Rightarrow y = 0$$

Decision boundary:

$$\sum_i w_i x_i + b = 0$$



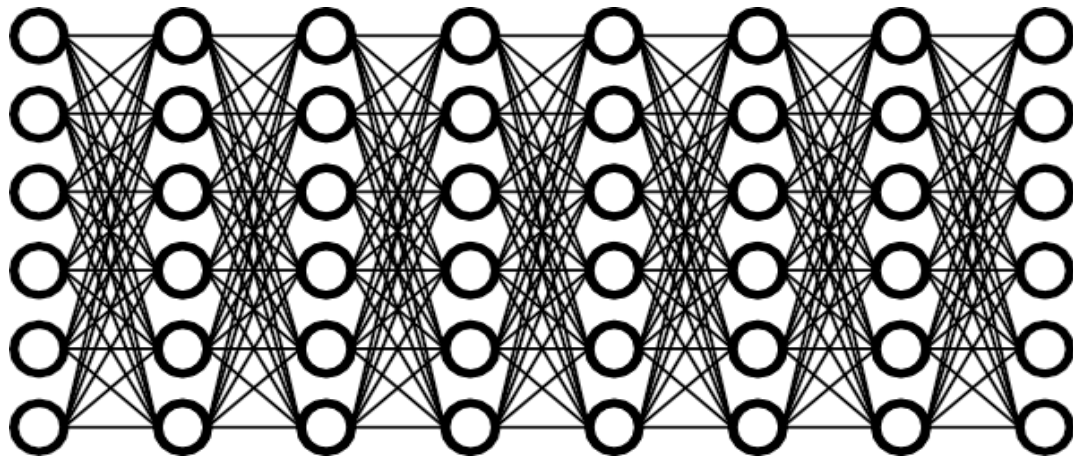
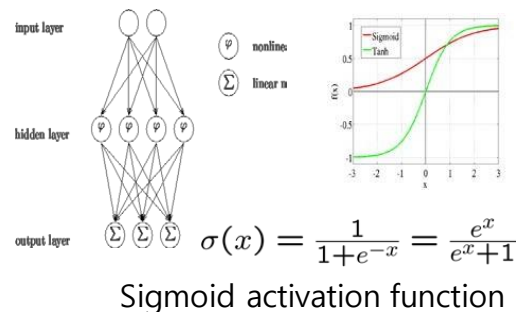
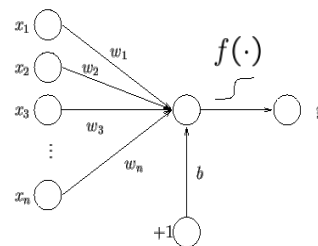
$$y = f(\sum_i w_i x_i + b)$$



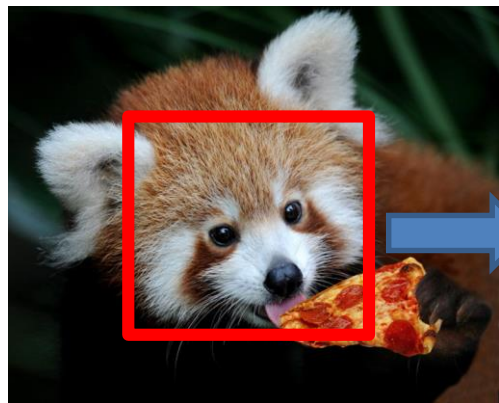
Perceptron

# History of Neural Networks

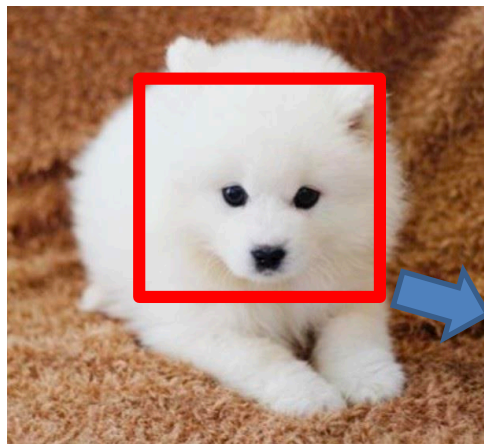
- First generation (1958~): perceptrons (F. Rosenblatt, 1958)
  - Criticized by Marvin Minsky about XOR problem
- Second generation (1986~) : multilayer perceptrons
  - Trained by back-propagating error signal (1986)
  - Mostly used shallow network with 1 hidden layer
- Third generation (2006~ ): deep learning
  - Deep belief nets (Hinton, 2006)
  - Deep neural network (DNN), convolutional neural network (CNN)



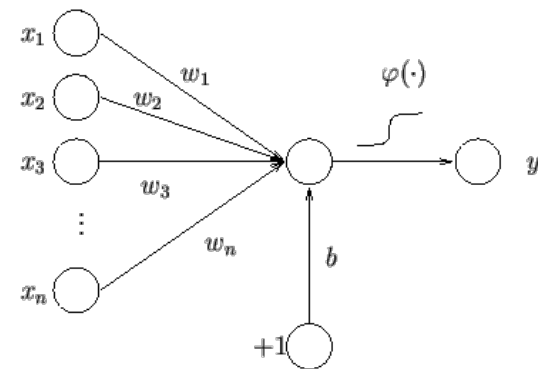
# Distinguish between Red Panda and Dog



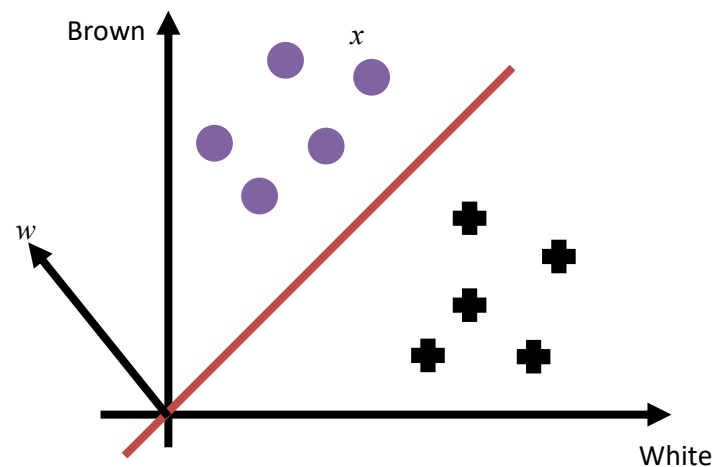
→ Brown 70%, white 50%



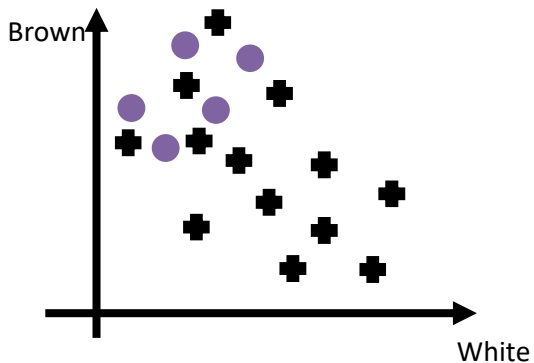
→ White 90%, Brown 5%



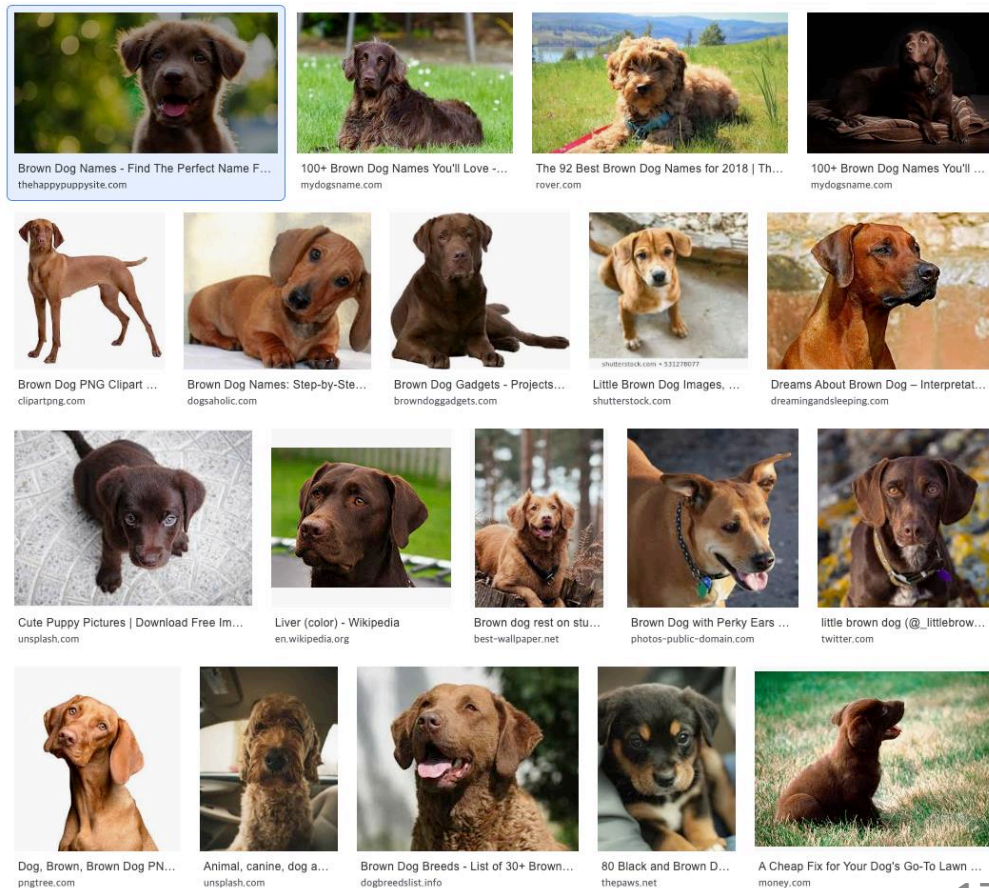
$w^T x > 0 \Rightarrow \text{Red Panda}$   
 $w^T x < 0 \Rightarrow \text{Dog}$



# Deep Learning Can Solve Hard Problem

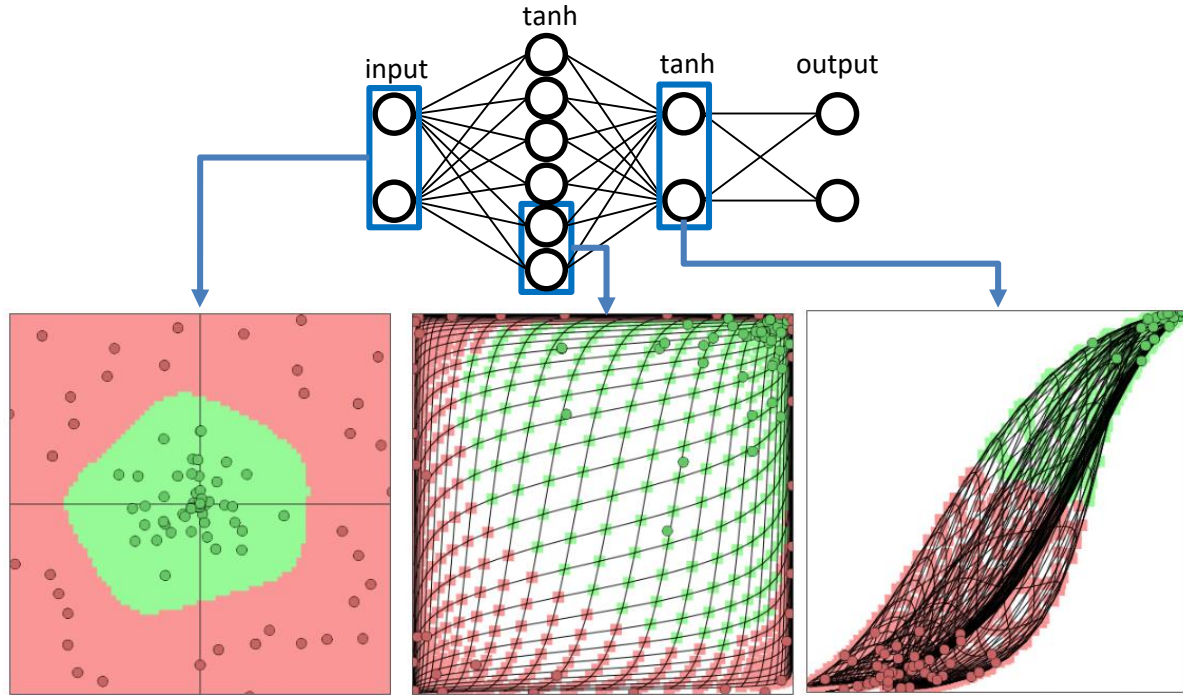


VS

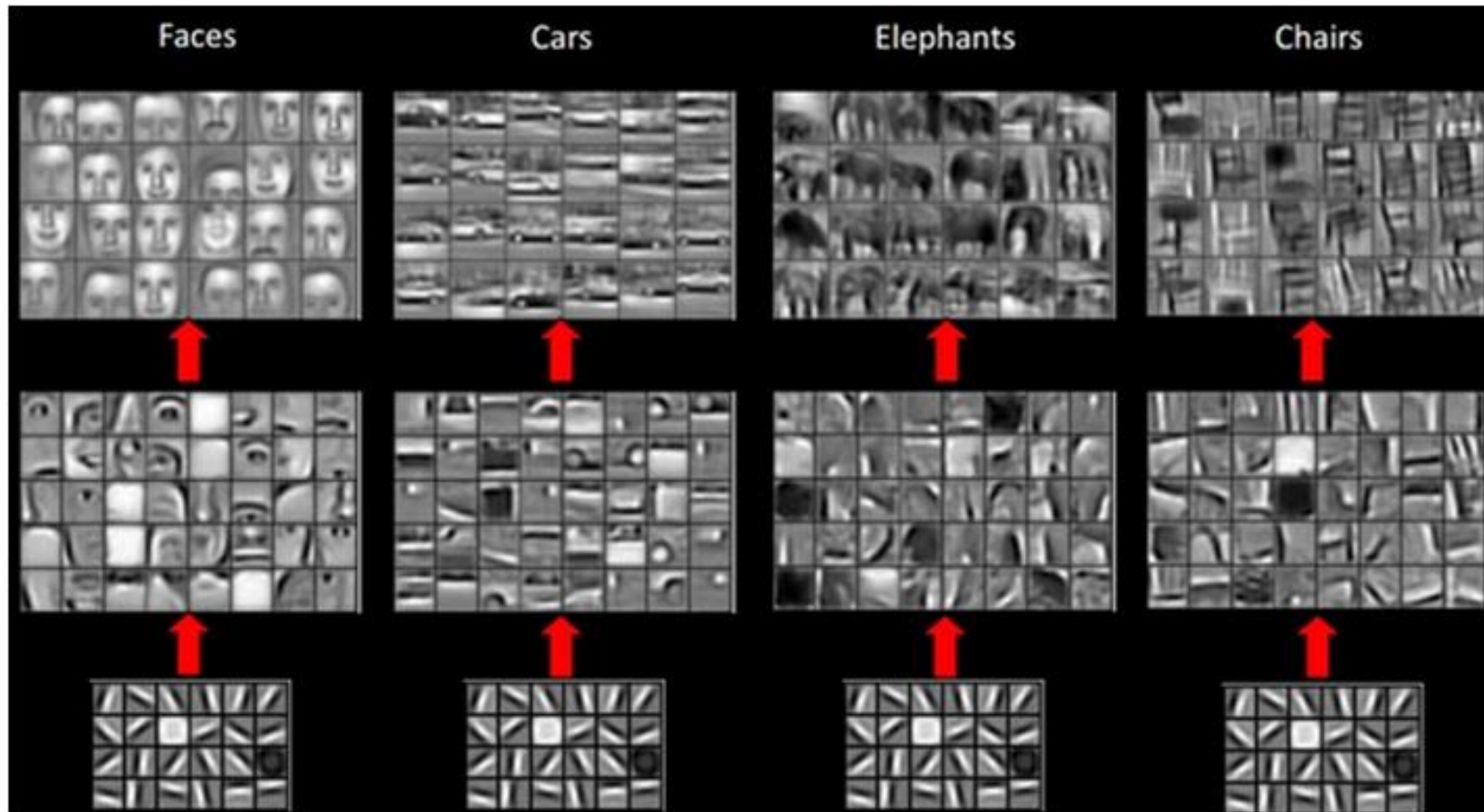




# Effectiveness of Deep Neural Network



# Convolutional Layer



[From recent Yann  
LeCun slides]



# Difficulty of Training Deep Neural Network

- **Vanishing gradient problem:** problems with non-linear activation
  - Solved by a new non-linear activation : rectified linear unit (ReLU)
- Typically requires **lots of labeled data**
  - Solved by big data & crowd sourcing, GPU
- **Overfitting problem:** Given limited amounts of labeled data, training via back-propagation does not work well
  - Solved by a new regularization method : dropout, dropconnect, etc.
- Get stuck in **local minima**
  - Local minima are all similar in high-dimensional non-convex optimization
  - Local minima are good and close to global minimum



---

# **Thank You!**

---