AMCS 394C Contemporary Topics in Applied Math

Mathematical Introduction to Deep Learning

Course Information

Jinchao Xu

KAUST

jinchao.xu@kaust.edu.sa

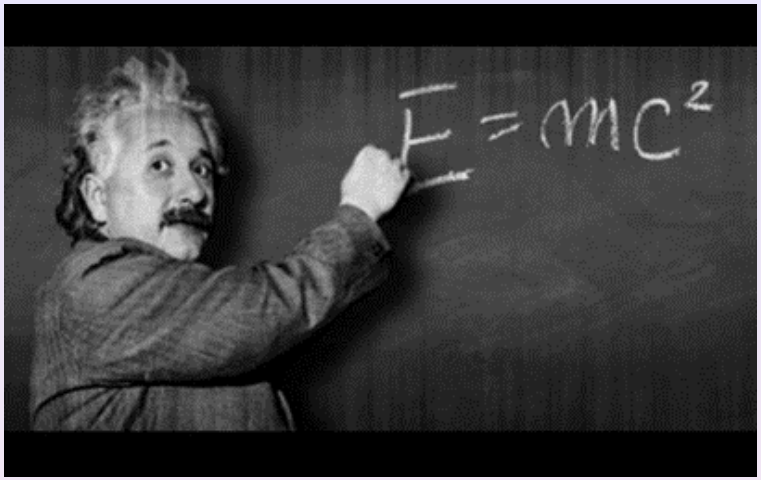
Spring 2023

Contents

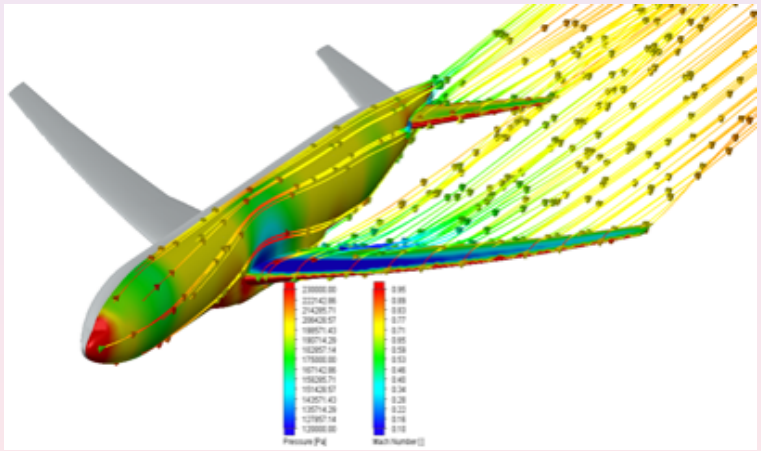
Outlines

Four major methods of scientific research

Experimental Science



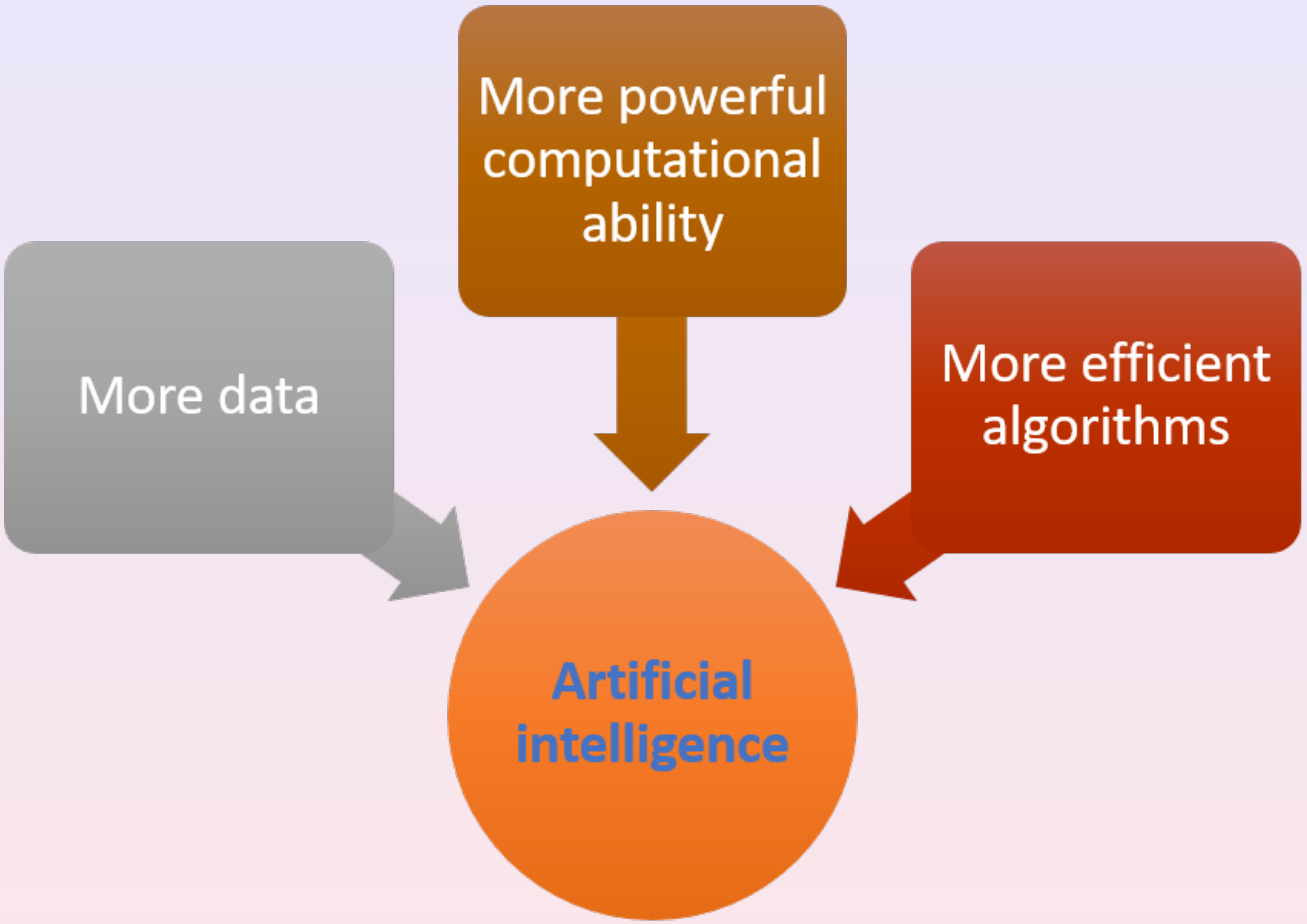
Theoretical Science

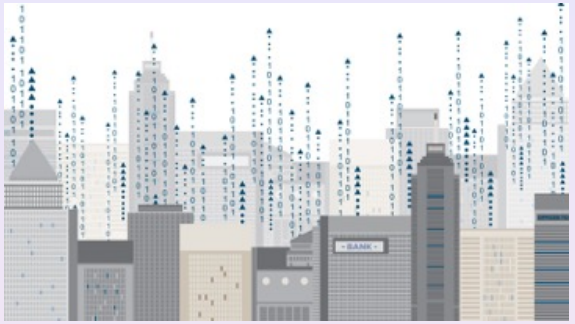


Computational Science

Data Science

The era of the artificial intelligence

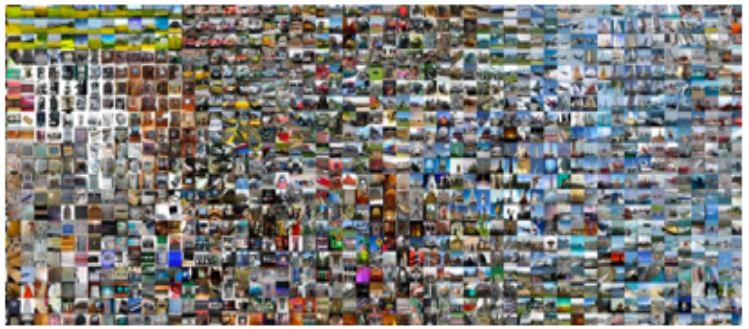




Success Stories - I: Image Classification

- In 1998, LeCun etc. proposed the neural network LeNet-5 based on the convolution and applied to the hand-written digits recognition successfully. Hence LeNet-5 is also called the first convolutional neural network(CNN) successfully applied.
- In 2012, Hinton and his student Alex joined the graph recognition competition by ImageNet and improved the accuracy significantly with the CNN AlexNet.
- In 2013 Google purchased a Canadian startup company on neural network, DNNResearch. It's set up by Geoffrey Hinton and his graduate students Alex Krizhevsky and Ilya Sutskever in 2012.
- In 2015, Kaiming He etc. proposed the ResNet CNN structure, which has become a state-of-art CNN structure and widely used in industry and discussed in academia.
- Afterwards CNN is widely used in the field of computer vision and breaks the records ceaselessly.

IMAGENET

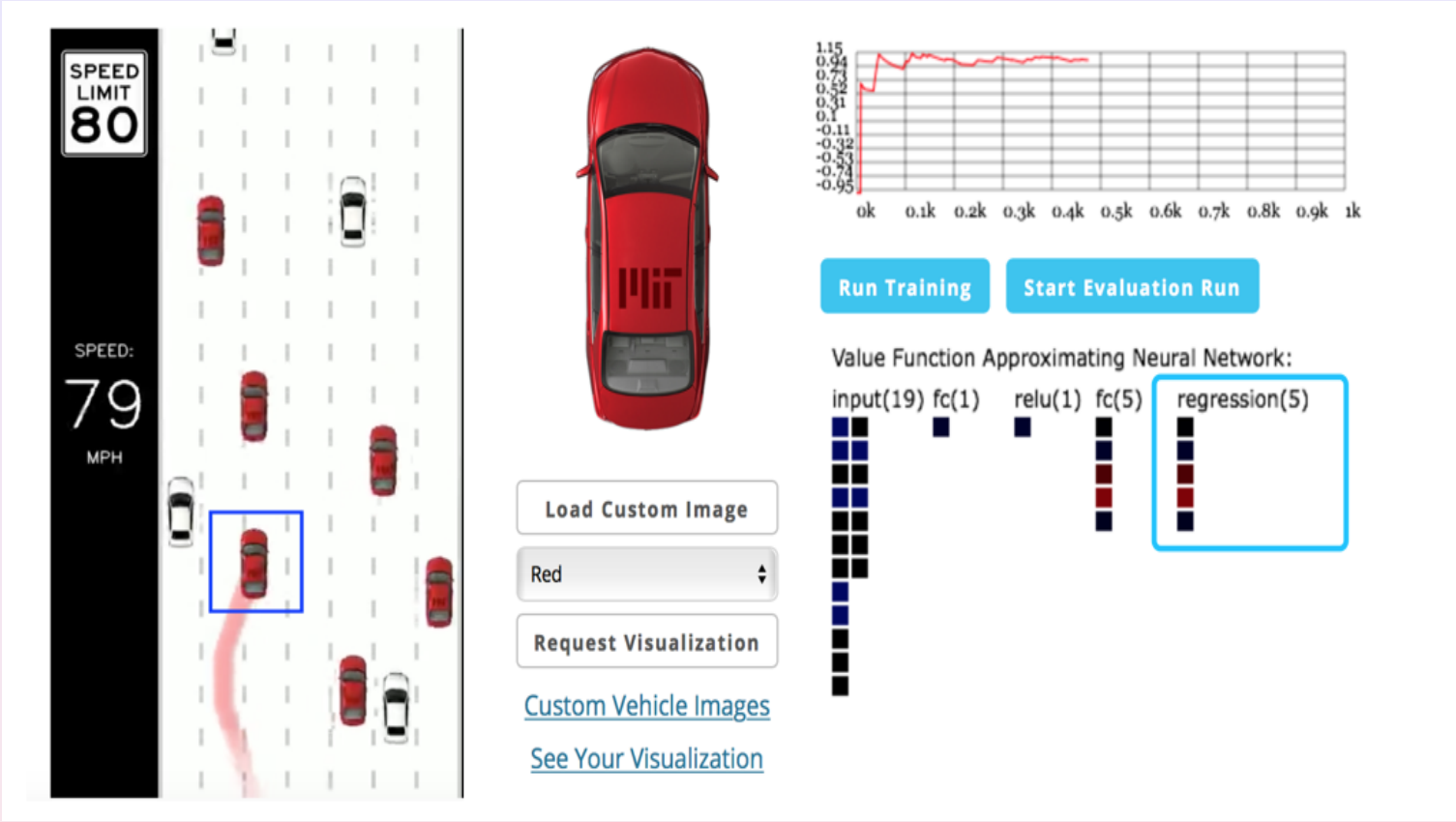


Success Stories - II: AlphaGo & AlphaZero

- From 2016 to 2017, the AI program AlphaGo developed by Google DeepMind beat down all the Go champions worldwide.
- 2018 AlphaZero gave a unified principle for many other board games.
- The CEO of Google DeepMind Demis Hassabis announced to integrate AlphaGo with medical, robots and so on. They can learn by themselves since they are artificial intelligence, and transfer learning can be done with enough data.

Auto-driving: a simple demo

DeepTraffic: https://selfdrivingcars.mit.edu/deeptraffic/



- Goal: drive safe and fast
- Input: situation around the car
- Output: among five decisions (stay, accelerate, decelerate, right turn and left turn)

Classification problems with five classes !

Success Stories - III: Auto-driving

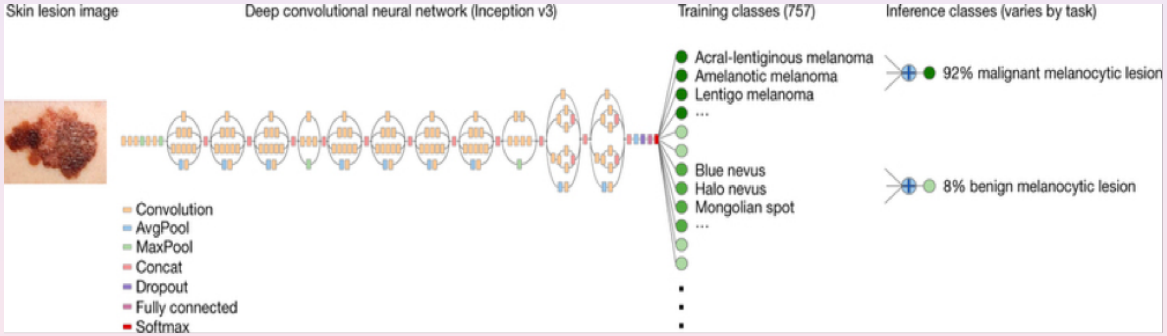
Auto-driving (the next competition in AI)

- In 2009, Google proposed the plan to replace human driving with softwares. Afterwards, many large technology companies such as Tesla, Google, Uber and Benz devoted a lot on investigating the technology of automated driving.
- There are about dozens of companies focus on automatic drive techniques from L2 to L4.
- In many countries, road examination autonomous vehicles are allowed with applications.
- Tesla – Autopilot; Google – Waymo; Baidu – Apollo; GM – Cruise; Volkswagen – DAS Autonomy...

Success Stories - IV: Diagnosis and Classification of Cancer

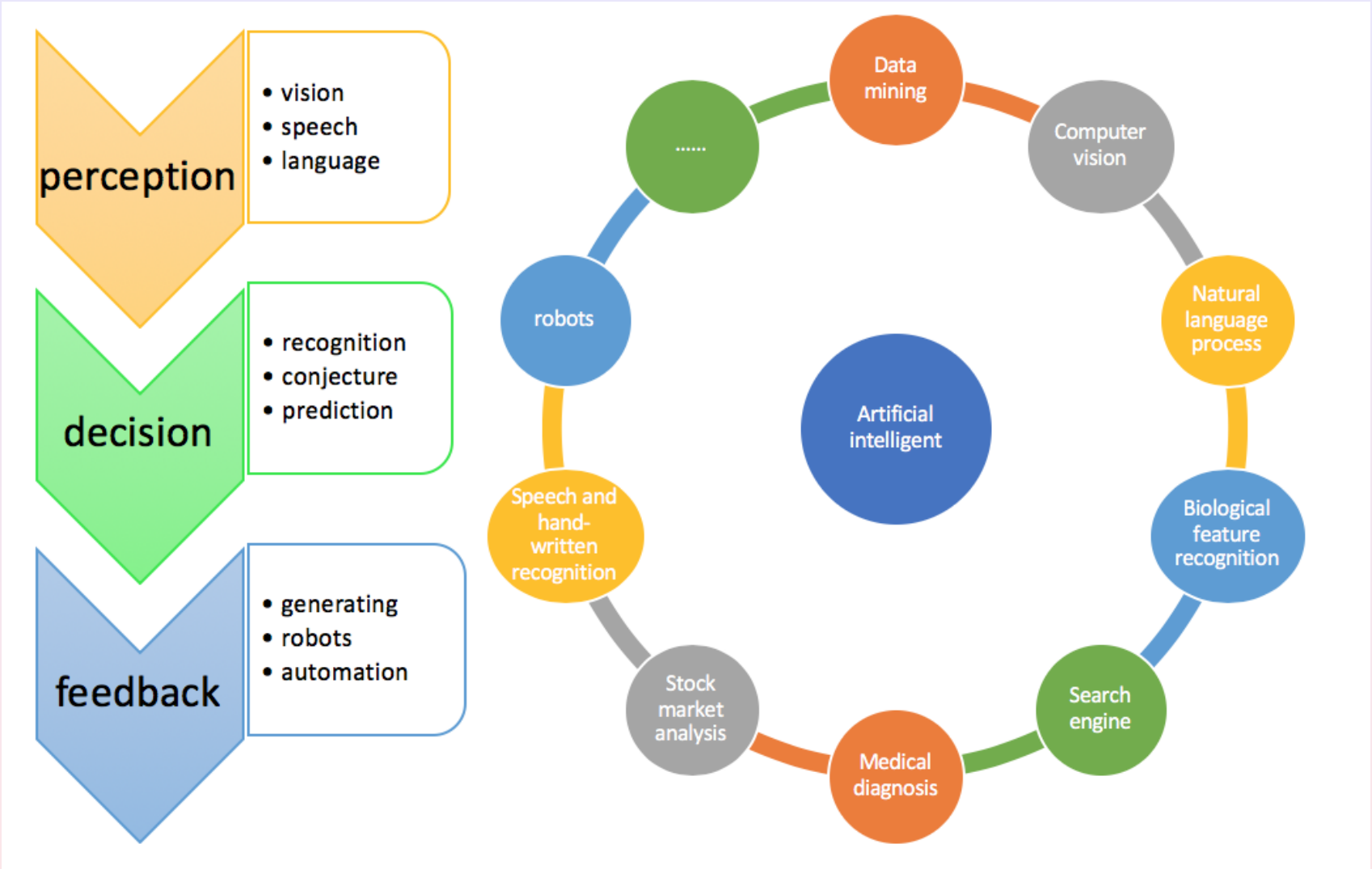
AI can automatically diagnose the cancer

- In 2017, a team in Stanford University achieved AI automatic diagnosis of the skin cancer with the CNN, the accuracy of which is as high as a human expert.
- This model is trained based on a public model by Google, while the original model is only used to classify the cats and dogs in photos.

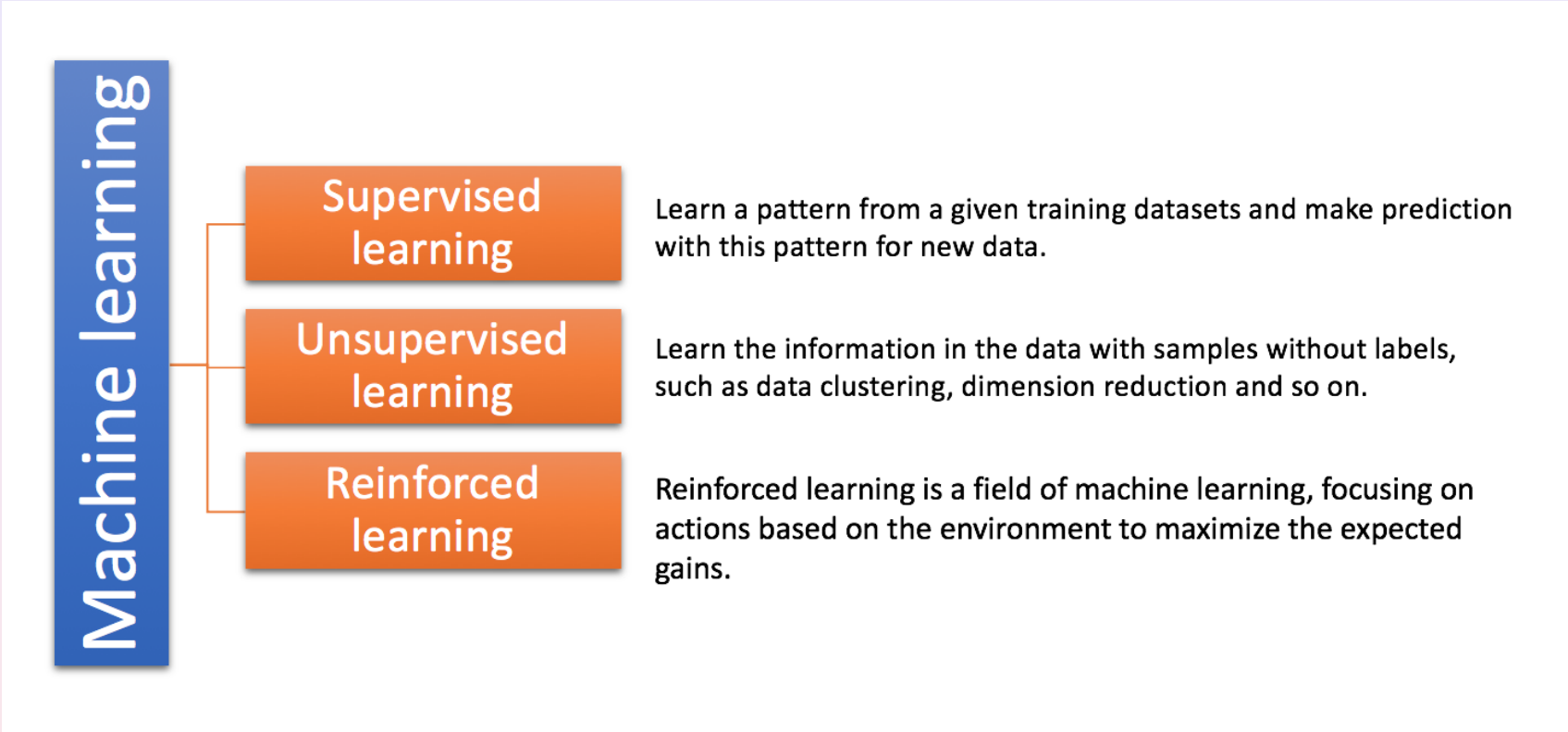


Ref: *Dermatologist-level classification of skin cancer with deep neural networks*, Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun, *Nature* 542.7639 (2017): 115-118.

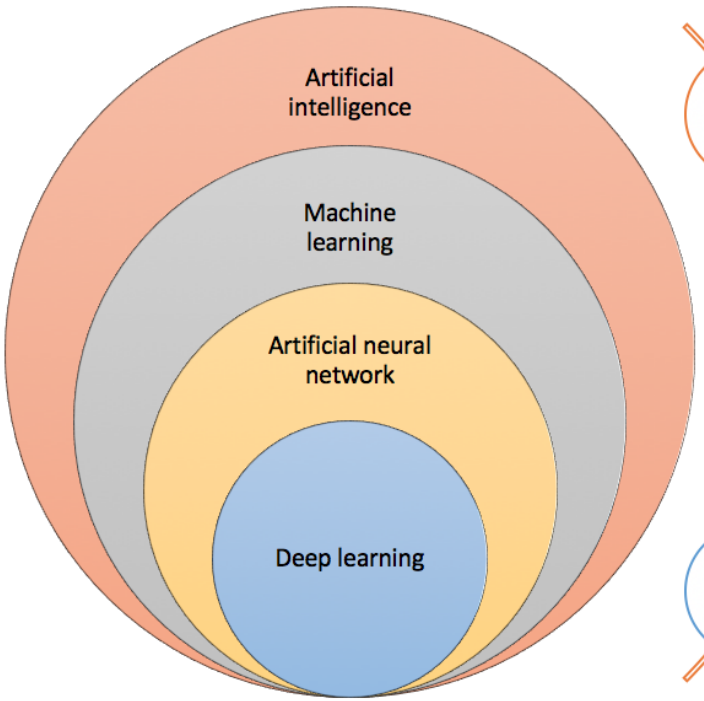
Artificial intelligent



Machine learning



Deep learning



The object of the artificial intelligence is to have the machines to model the human intelligence

The core of the AI is the machine learning, which is the fundamental way to equip the computers with intelligence

Artificial neural network is an algorithm in early machine learning, consisting of perception, multilayer perception and deep learning

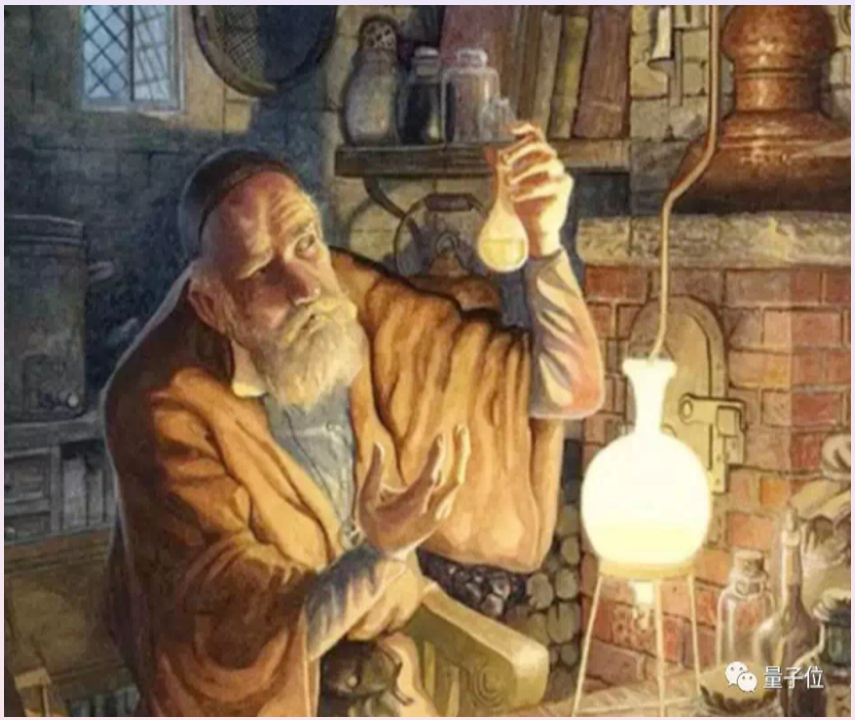
The motivation of the deep learning is to establish and model the human brain in the process of analyzing and learning

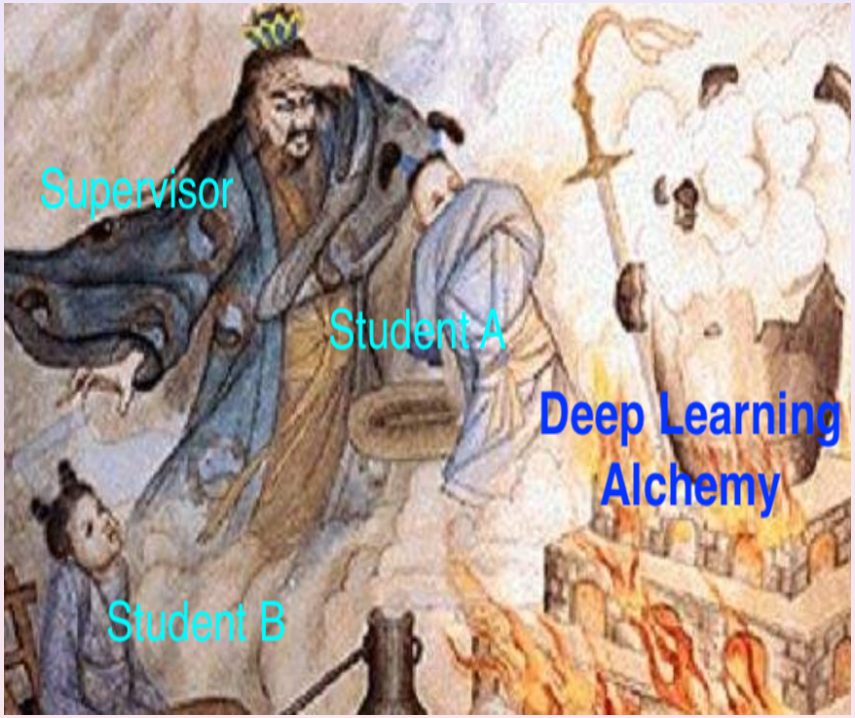
Outlines

Deep Concern

Deep learning is "alchemy"

– Ali Rahimi, NIPS 2017





Deep Concern

Quotes(Researchers from U Wash, Princeton, MIT, . . .)

- “Deep learning is killing image processing, natural language processing...”
- “Graduate students only work on deep learning”
- “One time extinction event – graduate students won’t know the fundamental tools”

Source: D. Donoho/ H. Monajemi/ V. Papyan “Stats 38” at Stanford and B. Dong at PKU

The objectives of this course

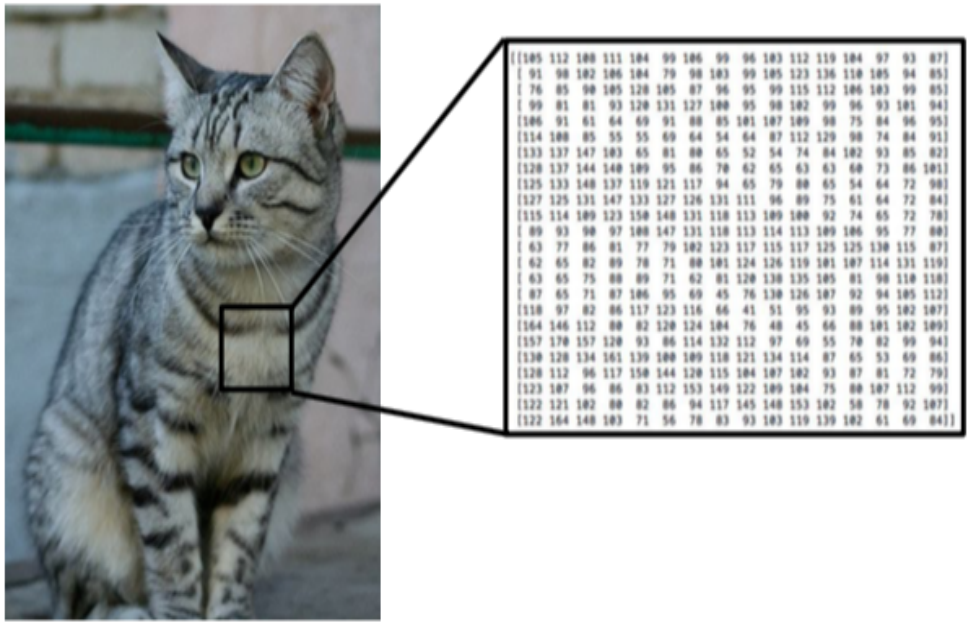
1. provide mathematical understanding of how and why deep learning works
2. hands-on programing experiences on popular and core deep learning model:
   - convolutional neural networks

Outlines

A basic AI problem: classification

- Can a machine (function) tell the difference ?

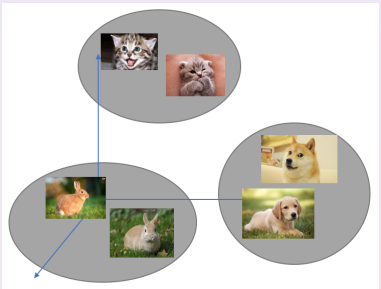
Human vision and computer representation

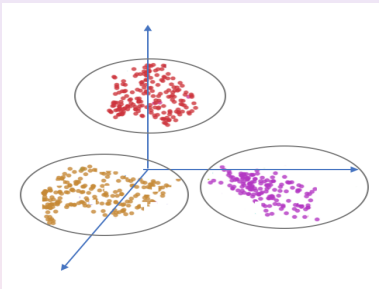


- An image is just a big grid of numbers between [0, 255]
  - e.g. $800 \times 600 \times 3$ (3 channels RGB)

Supervised learning

- Function interpolation (data fitting)
  - Each image = a big vector of pixel values
    - ★ $d = 1280 \times 720 \times 3$ (width$\times$ height $\times$ RGB channel) $\approx$ 3M.
  - 3 different sets of points in $\mathbb{R}^d$, are they separable?





  - Mathematical problem: Find $f(\cdot;\Theta) : \mathbb{R}^d \to \mathbb{R}^3$ such that:

$$f(\text{  };\Theta) \approx \begin{pmatrix}1\\0\\0\end{pmatrix} \quad f(\text{  };\Theta) \approx \begin{pmatrix}0\\1\\0\end{pmatrix} \quad f(\text{  };\Theta) \approx \begin{pmatrix}0\\0\\1\end{pmatrix}$$

How to formulate "learning"?

- Data: $\{x_i, y_i\}_{i=1}^m$
- Find f^* in some function class such that $f^*(x_i) \approx y_i$.
- Mathematically: solve the optimization problem by parameterizing the abstract function class

$$\min_{\Theta} L(x, y, \Theta) \tag{1}$$

where

$$L(x, y, \Theta) = \mathbb{E}_{(x,y)\sim\mathcal{D}}[L(f(x;\Theta), y)] \approx \frac{1}{N}\sum_{i=1}^{N} \|f(x_i;\Theta) - y_i\|^2$$

  - Or combine the feature map f with the logistic regression model to obtain cross-entropy loss function.

- Application: image classification:

$$f(\boxed{?};\Theta) = \begin{pmatrix} 0.7 \\ 0.2 \\ 0.1 \end{pmatrix} \quad \Longrightarrow \quad \boxed{?} = \text{cat}$$

Critical parts of deep learning

Data sets

- image: MNIST, CIFAR-10, CIFAR-100, ImageNet...
- video: UCF101, Kinetics, Sports-1M, YouTube-8M...
- text: Wikitext-2, Wikitext-103, Amazon reviews - Full...

Models: function classes to approximate the ground truth f^*

- numerical analysis: polynomials, finite element, multigrid methods...
- classical machine learning: logistic regression, support vector machine, decision tree...
- deep learning: deep neural networks, convolutional neural networks, recurrent neural networks...

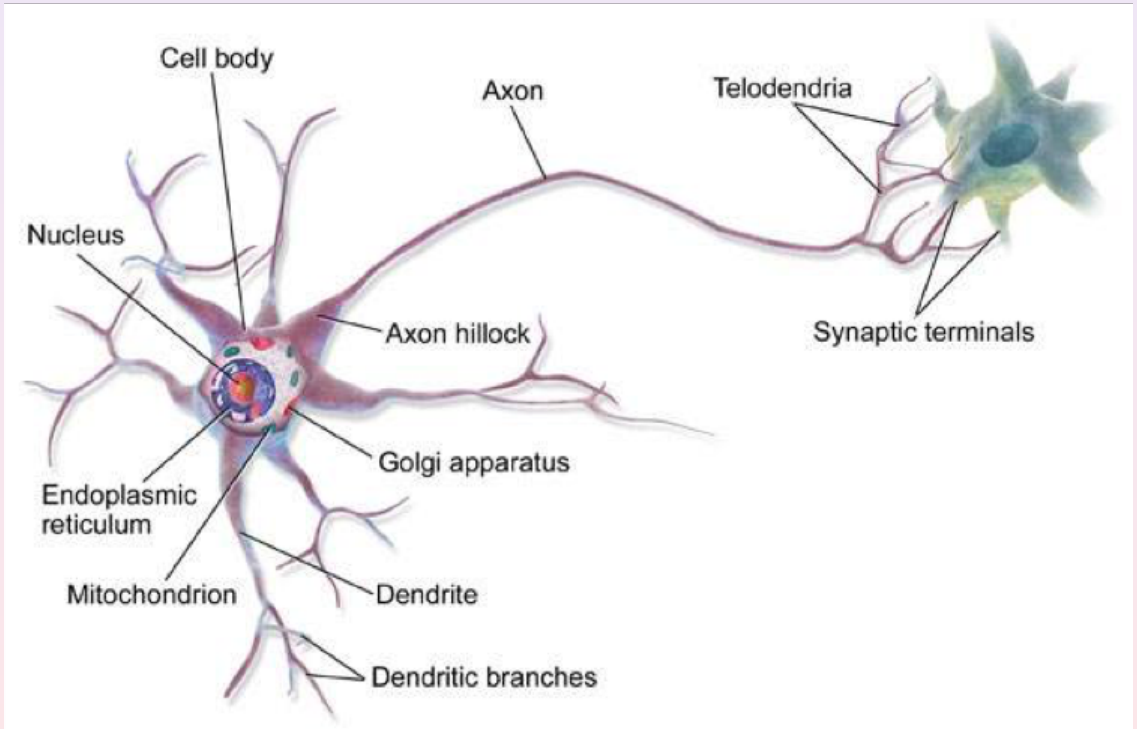
Training algorithms: optimization to solve the best approximation

- gradient descent,
- stochastic gradient descent

Example: deep neural networks

Neurons of animals

- Dendrite: receive the electricsignal from other neurons
- Cell body: process the electricsignal from other neurons
- Axon: send the electricsignal to other neurons
- Activation: the intensity of the signal is larger than the threshold



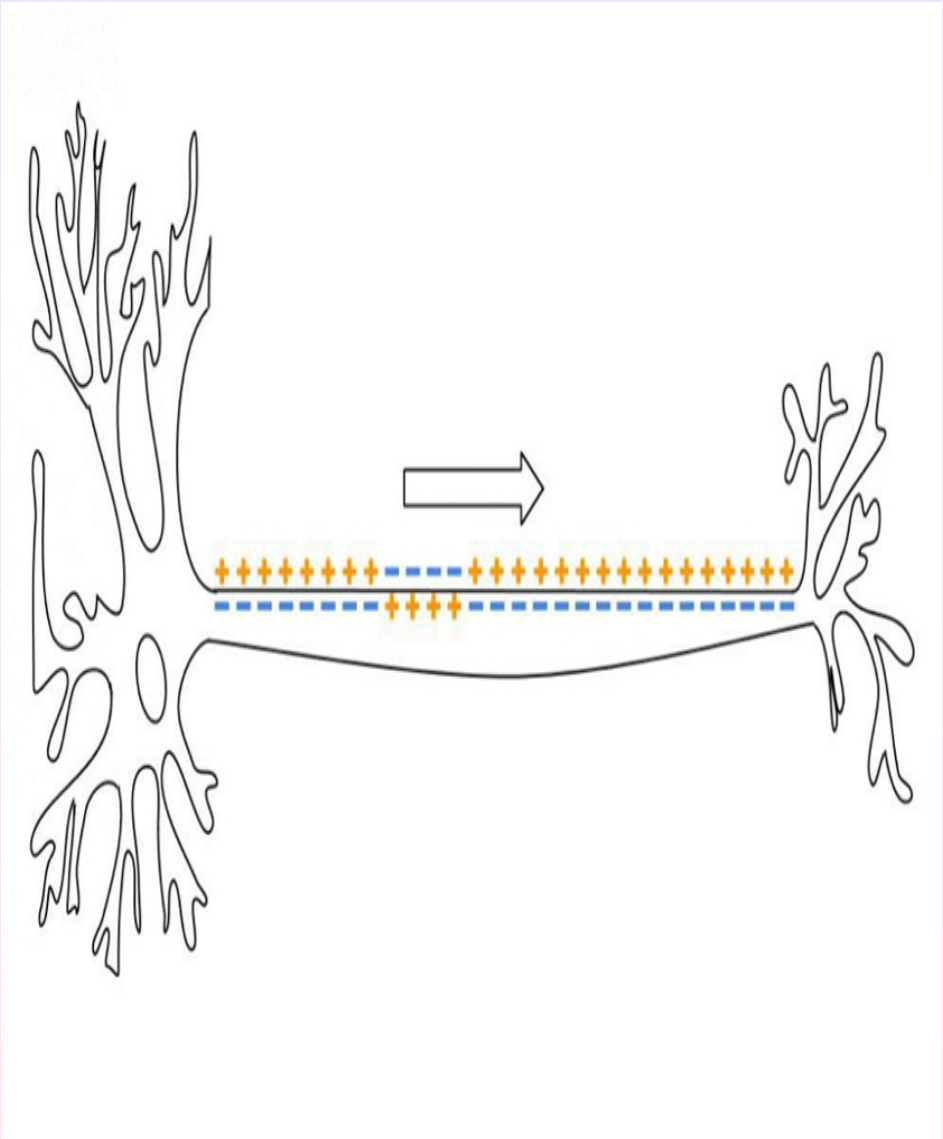
Linear functions

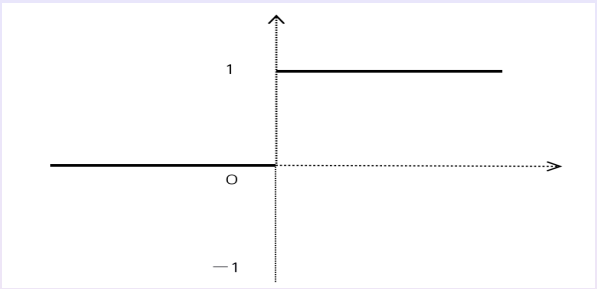
$$Wx + b = \begin{pmatrix} W_1 x + b_1 \\ \dots \\ W_n x + b_n \end{pmatrix}$$

where

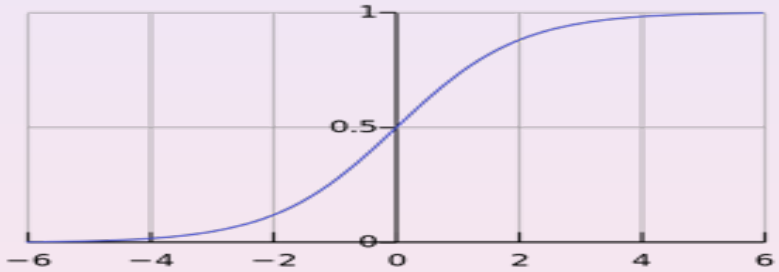
$$W_k x + b_k = \sum_{i=1}^{m} w_{k,i} x_i + b_k$$

Activation Function

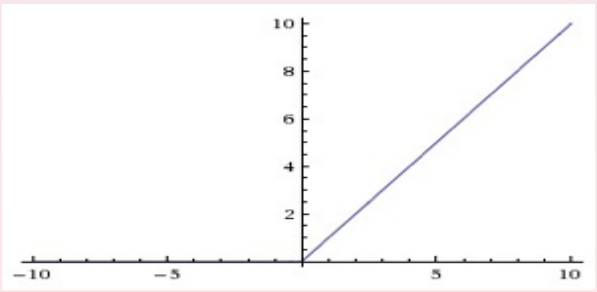




$$\sigma = \mathrm{H}(x)$$



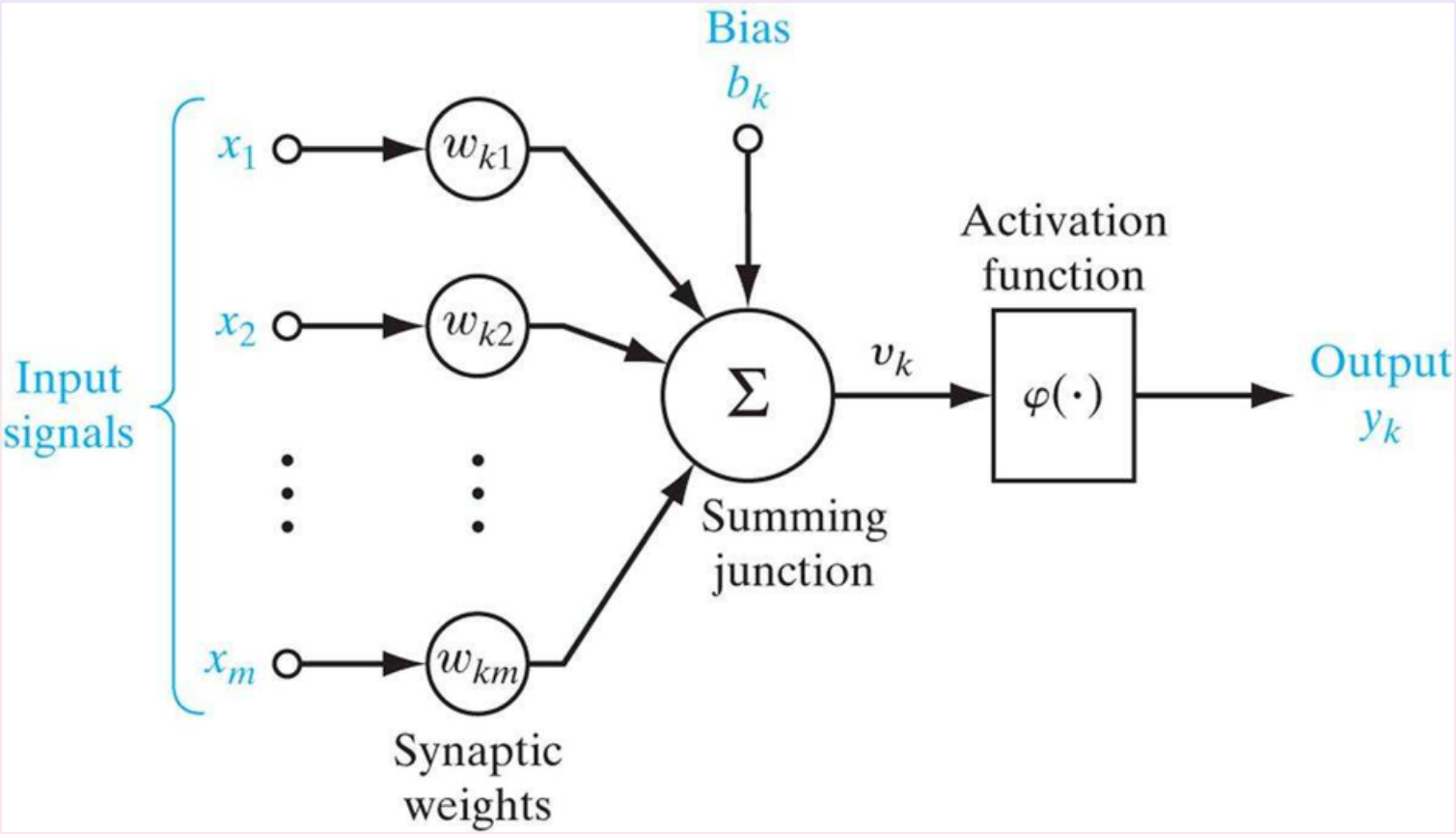
$$\sigma = \frac{1}{1 + e^{-x}}$$



$$\sigma = \mathrm{ReLU}(x) := max(0, x)$$

Artificial Neuron

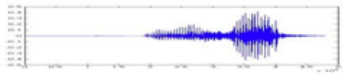
- An artificial neuron consists of
  - An affine transformation: $\theta(x) = Wx + b$
  - A nonlinear activation function: σ
- The output of the neuron is $y = \sigma(\theta(x))$



Model of the neural network

- Model consists of multiple artificial neurons in some way $y = f(x; W)$
  - x—input data
  - y—result of prediction
  - W—parameters
- Learning: find the optimized function $f^*(\cdot) = f(\cdot; W^*)$ to minimize the given loss function

- Speech Recognition

  $f^*($  $)=$ “你好”
- Handwritten Recognition

  $f^*($  $)=$ “2”
- Playing Go

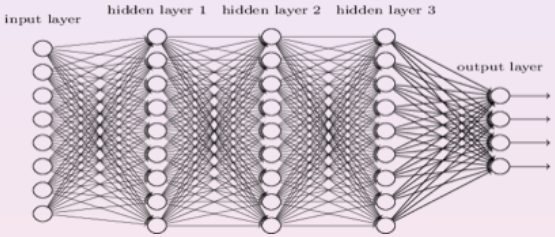
  $f^*($  $)=$ “5-5” (step)
- Dialogue System

  $f^*($ “Hi” $)=$ “Hello”

  (what the user said) (system response)

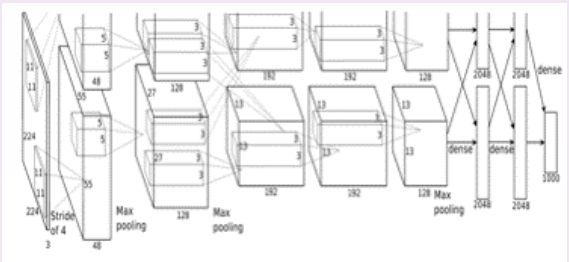
Three major models (function classes) in deep learning

Deep neural network



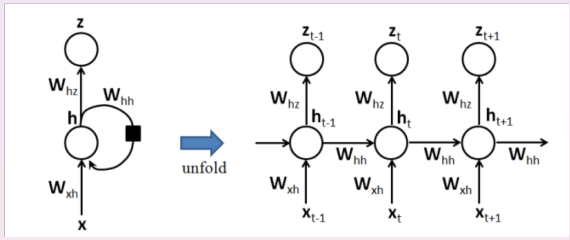
- Model with nonlinear statistical data
- Data classification
- Feature extraction

Convolutional neural network



- Graph recognition
- Graph segmentation
- Video analysis
- Natural language processing

Recurrent/recursion neural network



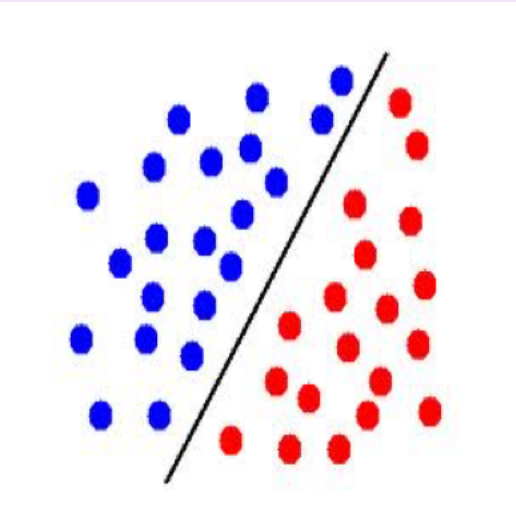
- Speech recognition
- Natural language processing
- Intelligent dialogue
- Machine translation

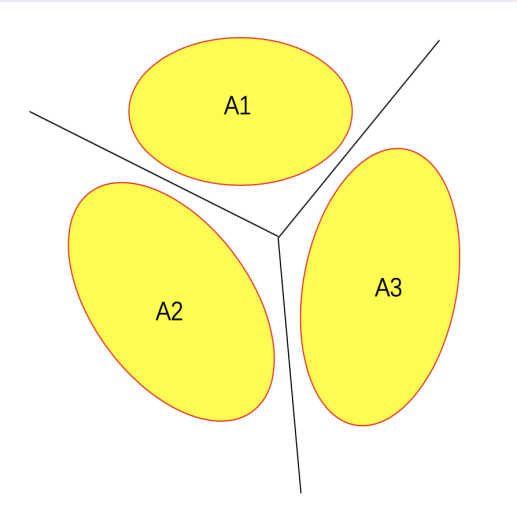
Outlines

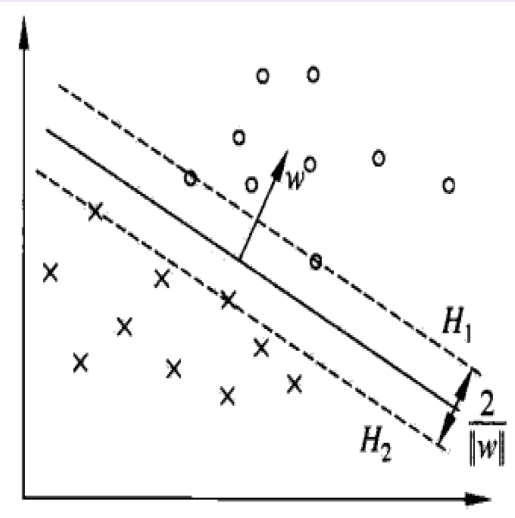
Sketch of this course in theoretical perspective

Basic machine learning models

- logistic regression from numerical viewpoint
- logistic regression from probability viewpoint
- support vector machine



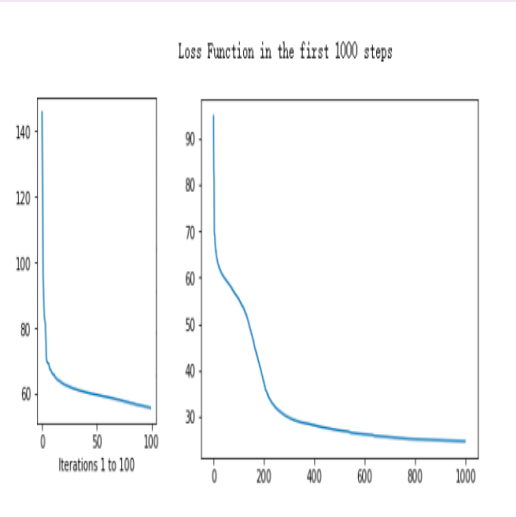


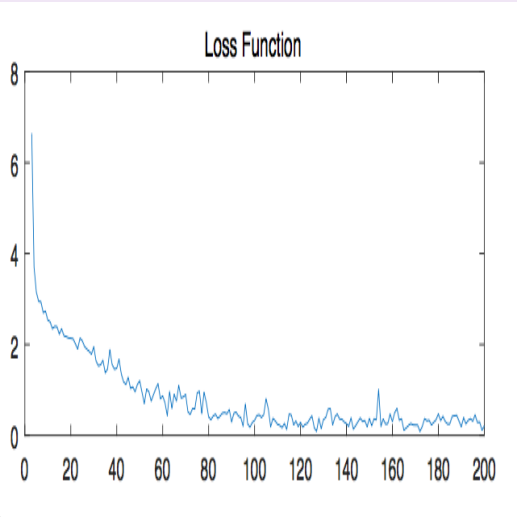


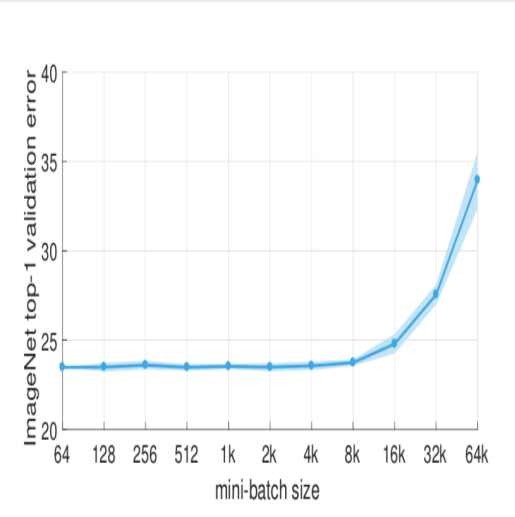
Sketch of this course in theoretical perspective

Training algorithms

- multi-variable calculus
- gradient descent method
- stochastic gradient descent



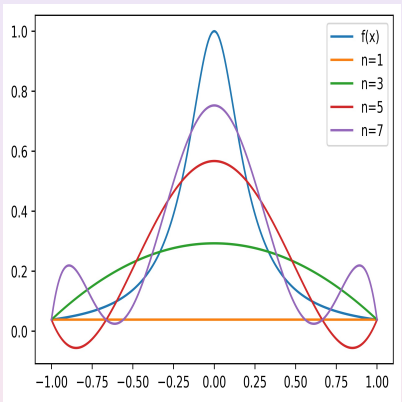


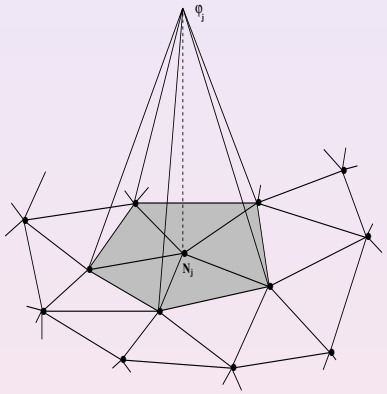


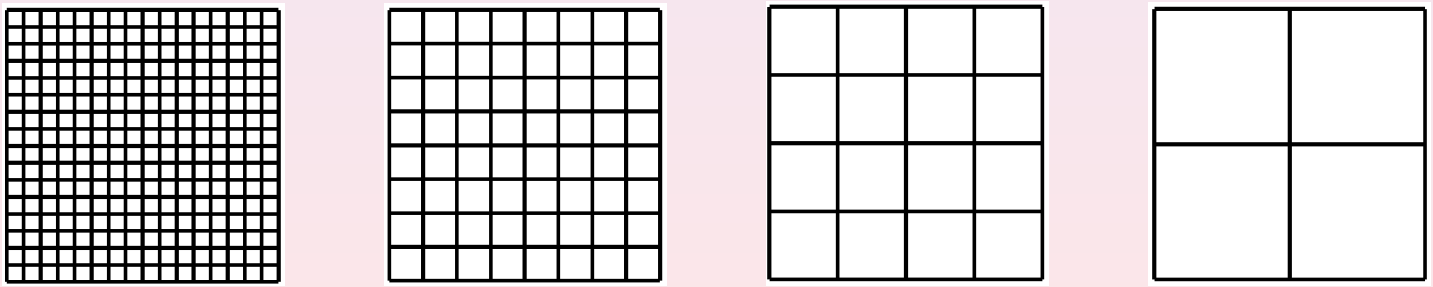
Sketch of this course in theoretical perspective

Numerical analysis techniques

- function approximation
- finite element methods
- multigrid methods



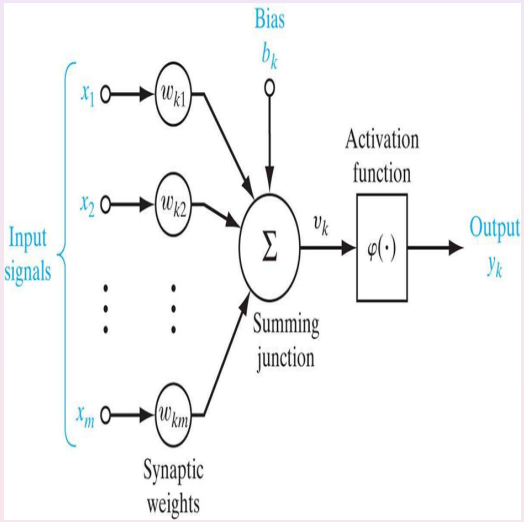


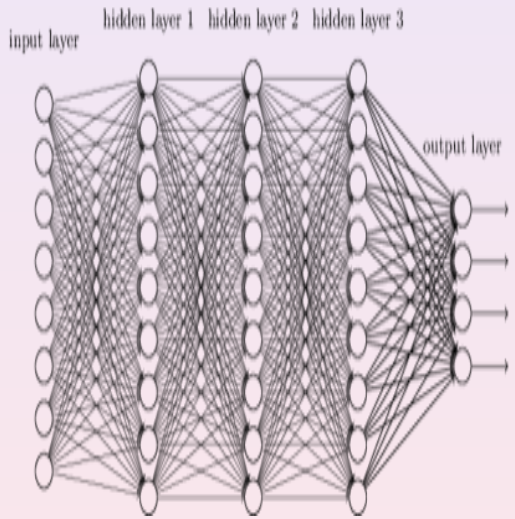


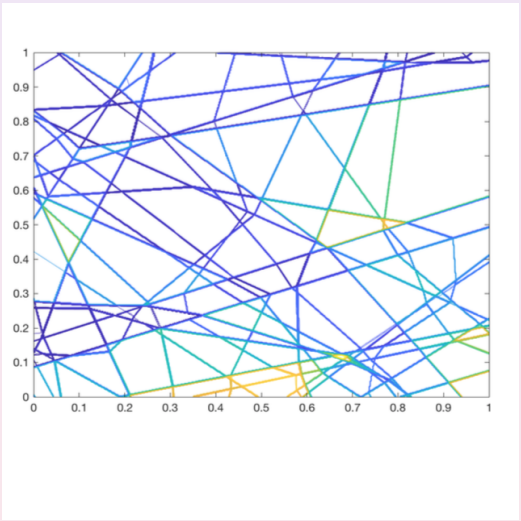
Sketch of this course in theoretical perspective

Deep neural networks

- model structures
- approximation properties
- relation to linear finite element



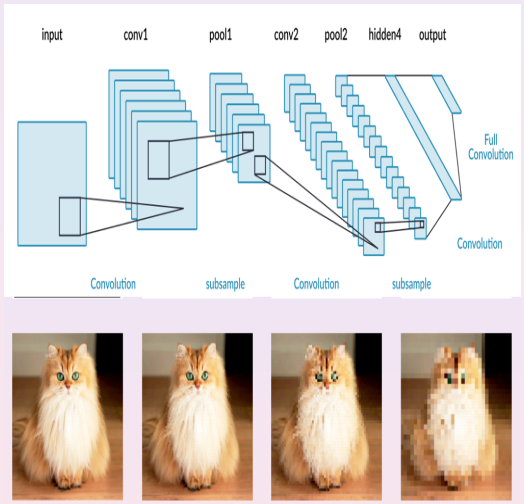


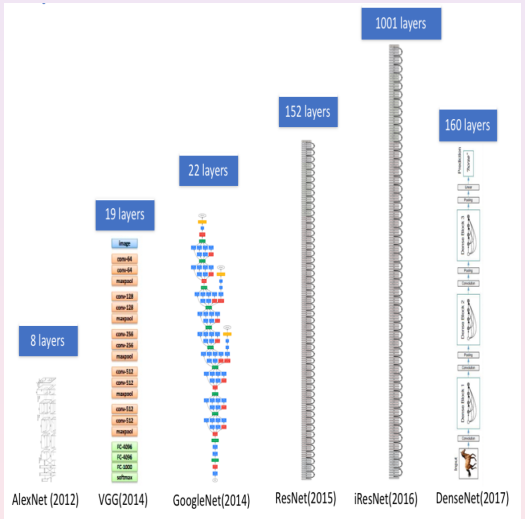


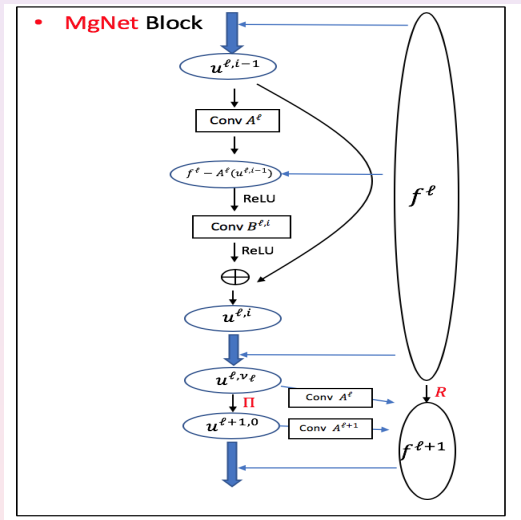
Sketch of this course in theoretical perspective

Convolutional neural networks

- classical networks architecture
- some key techniques in succeeding CNNs
- new understanding of CNN from multigrid
- MgNet







Main topics in this course

- Basic machine learning models
- Training algorithms
- Numerical analysis techniques
- Deep neural networks
- Convolutional neural networks
- . . .