Introduction to Natural Language Processing

玖强

OUTLINE

- NLP发展现状
- 传统NLP方法面临的挑战
- Big Data和Deep Learning给NLP带来的变革和机遇
- NLP的发展趋势，以及和各行各业的结合应用

Recent Trends in Deep Learning Based NLP

What is Natural Language Processing ?

Natural Language Processing (NLP) is :

"ability of machines to understand and interpret human language the way it is written or spoken".

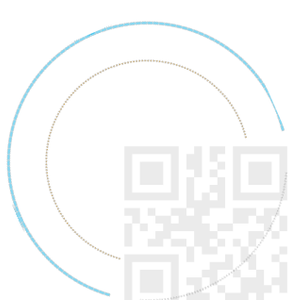
The objective of NLP is to make computer/machines as intelligent as human beings in understanding language.

1. Natural language processing is a field at the intersection of

- Computer science/计算机科学
- Artificial intelligence/人工智能
- Linguistics/语言学.

2. For computers to process or "*understand*" natural language in order to perform tasks that are useful, e.g.

- Question Answering/问答
- Perfect language understanding is AI-complete



DIFFERENCE BETWEEN NLP AND TEXT MINING OR TEXT ANALYTICS

- Natural language processing is responsible **for understanding meaning and structure of given text.**
- Text Mining or Text Analytics is a process of **extracting hidden information inside text data through pattern recognition.**

NLP (Natural Language Processing)	Text Mining or Text Analytics
Automated Speech	Automated Grouping (n grams approach)
Automated Writing	Automated Classification (bag of words)
Automated Translation	Pattern Discovery

NLP Applications

Applications range from simple to complex:

- Spell checking/语法检查, keyword search/关键词, finding synonyms/同义词
- Extracting information from websites such as: *product price, dates, location, people or company names, etc.*
- Classifying: *positive/negative* sentiment of longer documents
- Machine translation
- Spoken *dialog* systems
- Complex *question answering* /问答
- Conversation/对话. E.g., https://visualdialog.org/

NLP IN INDUSTRY … IS TAKING OFF / 已经起飞了

1. Applications:

- Search (written and spoken)
- Online advertisement matching
- Automated/assisted translation
- Sentiment analysis for marketing or finance/trading
- Speech recognition
- Chatbots / Dialog agents
  - Automating customer support
  - Controlling devices
  - Ordering goods

What's special about human language?

1. A human language is a system specifically constructed to convey the speaker/writer's meaning

- Not just an environmental signal, it's a deliberate communication/ 经过思考的交流
- Using an encoding which little kids can quickly learn (神奇!)

2. A human language is mostly a discrete/symbolic signaling system （离散/符号化）

- 火箭= 

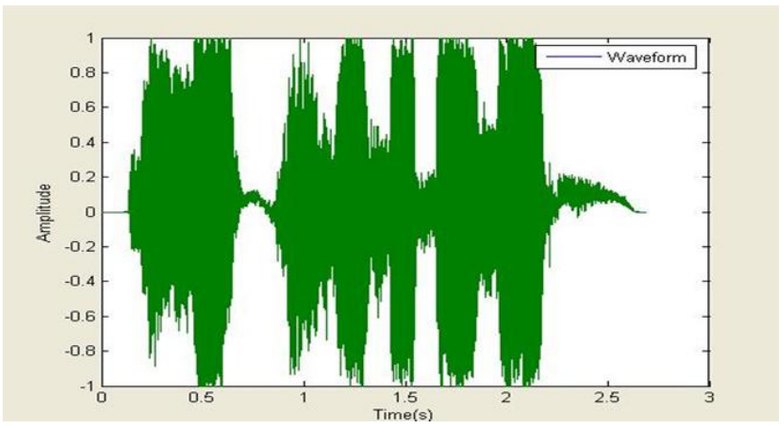
;小提琴 = 

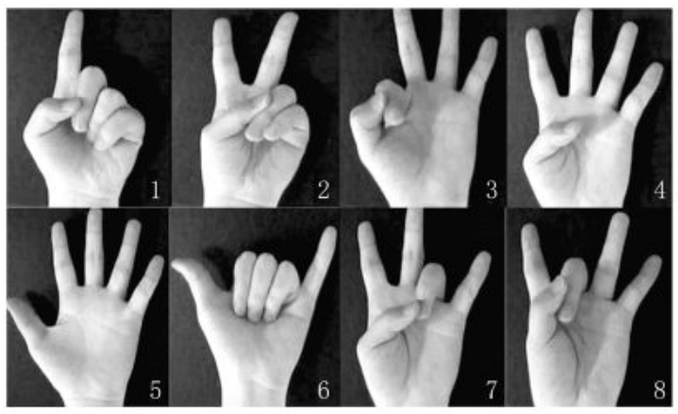
What's special about human language?

3. A The categorical symbols of a language can be encoded as a signal for communication in several ways:

- Sound
- Gesture
- Writing/Images

4. The symbol is invariant across different encodings!





What's special about human language?

4. A human language is a symbolic/categorical signaling system

- However, a brain encoding appears to be a continuous pattern of activation, and the symbols are transmitted via continuous signals of sound/vision

5. The large vocabulary/字典, symbolic encoding of words creates a problem for machine learning – sparsity/ 稀疏!

- We will explore a continuous encoding pattern of thought

$$\begin{pmatrix} 1.0 & 0 & 5.0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3.0 & 0 & 0 & 0 & 0 & 11.0 & 0 \\ 0 & 0 & 0 & 0 & 9.0 & 0 & 0 & 0 \\ 0 & 0 & 6.0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7.0 & 0 & 0 & 0 & 0 \\ 2.0 & 0 & 0 & 0 & 0 & 10.0 & 0 & 0 \\ 0 & 0 & 0 & 8.0 & 0 & 0 & 0 & 0 \\ 0 & 4.0 & 0 & 0 & 0 & 0 & 0 & 12.0 \end{pmatrix}$$

Challenges

Ambiguity

- **Natural language is highly ambiguous/ 模糊不清 and must be disambiguated/ 消除了歧义.**
  - 词法模糊： 冬天能穿多少穿多少, 夏天能穿多少穿多少
  - 句法模糊： 是(老男人)和女人 呢, 还是年老的(男人和女人)?
  - **Substitute A for B？**
    - 可能是A把B换下了，也可能是B把A换下了

Humor and Ambiguity

- **Many jokes rely on the ambiguity of language:**
  - 谐音误解
    - 比如在中国南北曲艺界，谐音误解是极极常见的表现手法，也正是因为过多倚重方言出效果，才导致南北笑星到彼此的地盘上水土不服。赵本山卖拐以后近十年的小品作品里，抛去模仿残疾人的戏码，谐音误解的创作方法比比皆是
  - 预期违背
    - *听者的内心OS是这样的：这种情景下发生这样的事儿，太不科学了！而讲者的内心OS一定是：呵呵，我哪能让你们猜到结尾呢。*
  - 同文异读
    - 陈佩斯的小品《警察与小偷》

Co-Reference Resolution

- **Determine which phrases in a document refer to the same underlying entity.**
  - **StanfordCoreNlp, FudanNlp, OpenNlp, and LTP**
  - 奥巴马出生在夏威夷。他是美国总统。他在2008年当选

– John put the carrot on the plate and ate it.

– Bush started the war in Iraq. But the president needed the consent of Congress.

Computers are no better than your dog.

- But we can teach them "how-to" by coding our knowledge of the language comprehension process

Big Data + Deep Learning + NLP

WHAT IS BIG DATA?

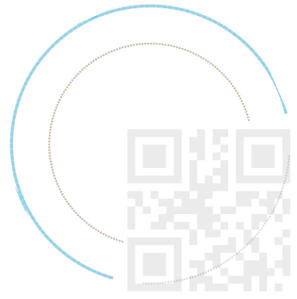
- According to the Author Dr. Kirk Borne, Principal Data Scientist, Big Data Definition is described as *big data is everything, quantified, and tracked*. For More Details on Big Data, Please Read – Ingestion And Processing of Data For Big Data and IoT Solutions

- 大数据=大+数据
- 大： Volume（数据量）, Velocity（数据速度）还有 variety（数据类别）

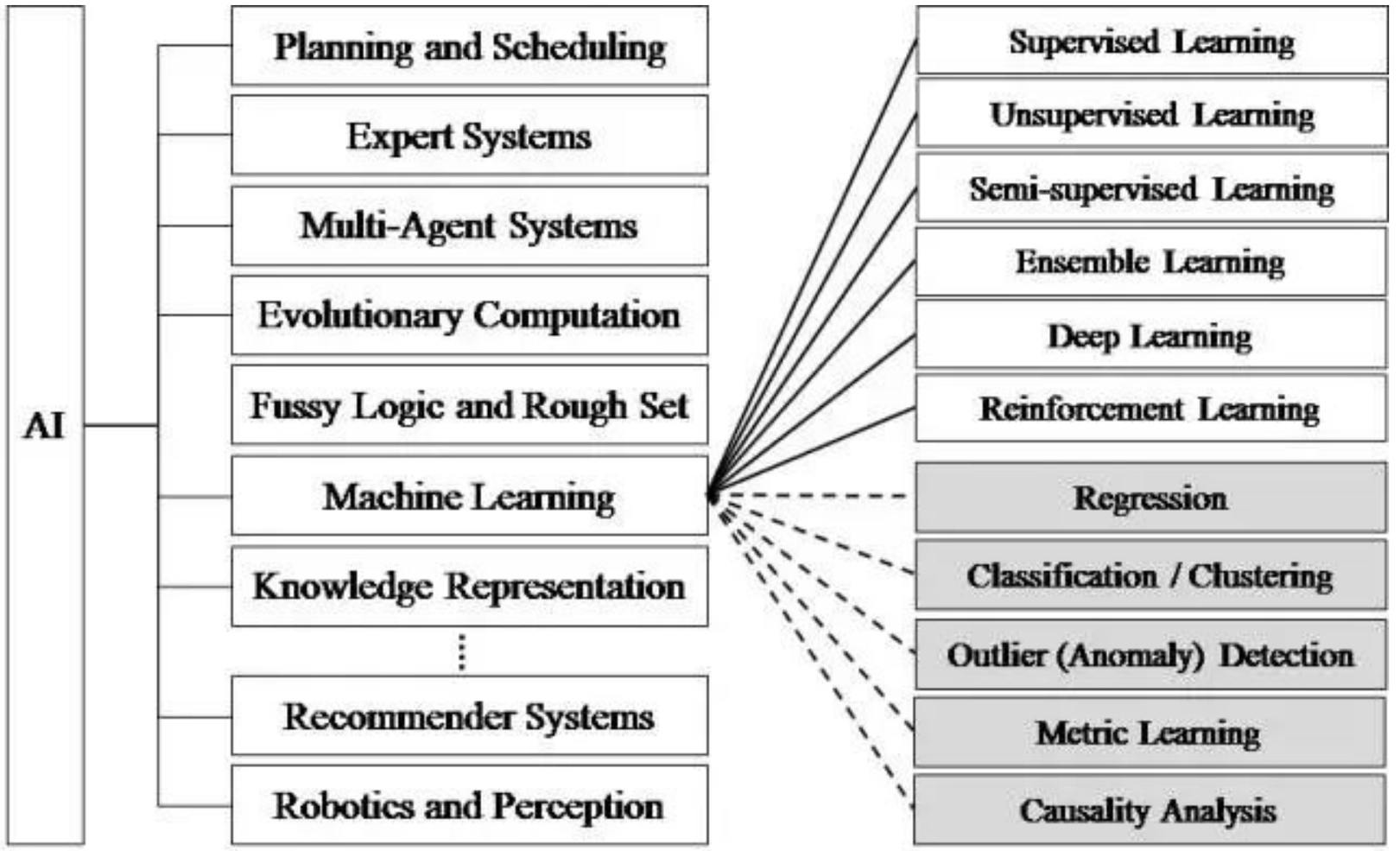
- Volume
  - 一般也可以认为是 Large-Scale data *or Big data?*
- Velocity
  - 数据到达的速度。要求：高效（Efficiency），即时（real-time），动态（dynamic），还有有预测性（predictive）等等
- Variety
  - 数据的类别。Multi-Modal data

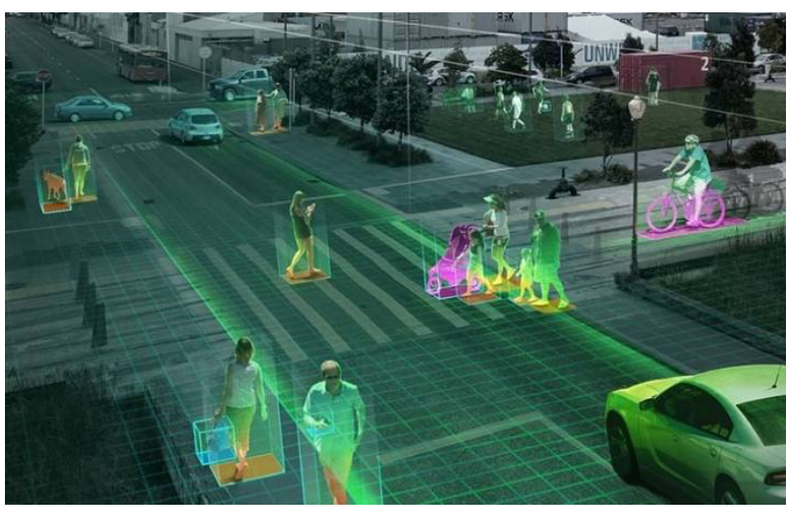
Deep NLP = Deep Learning + NLP

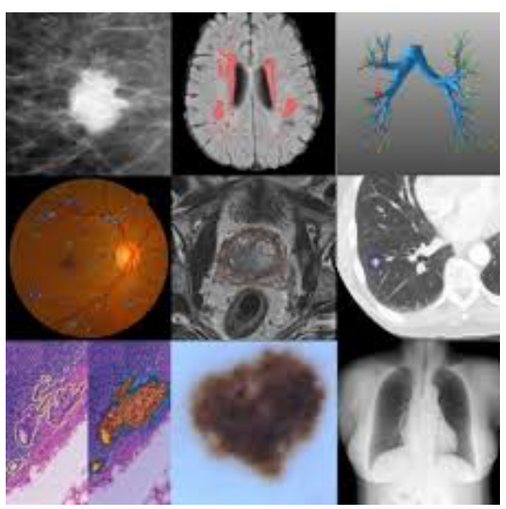
- **Combine ideas and goals of NLP with using representation learning and deep learning methods to solve them**
- **Several big improvements in recent years in NLP**
  - Linguistic levels: (speech), words, syntax, semantics
  - Intermediate tasks/tools: parts-of-speech, entities, parsing
  - Full applications:
    - *文本分类: sentiment analysis/ 情感分析;*
    - *question answering/问答;*
    - *dialogue agents/对话;*
    - *machine translation/机器翻译;*
    - *image captioning/图像描述生成;*
    - *visual question answering/图像问答*
    - *visual dialog/基于图像的对话*



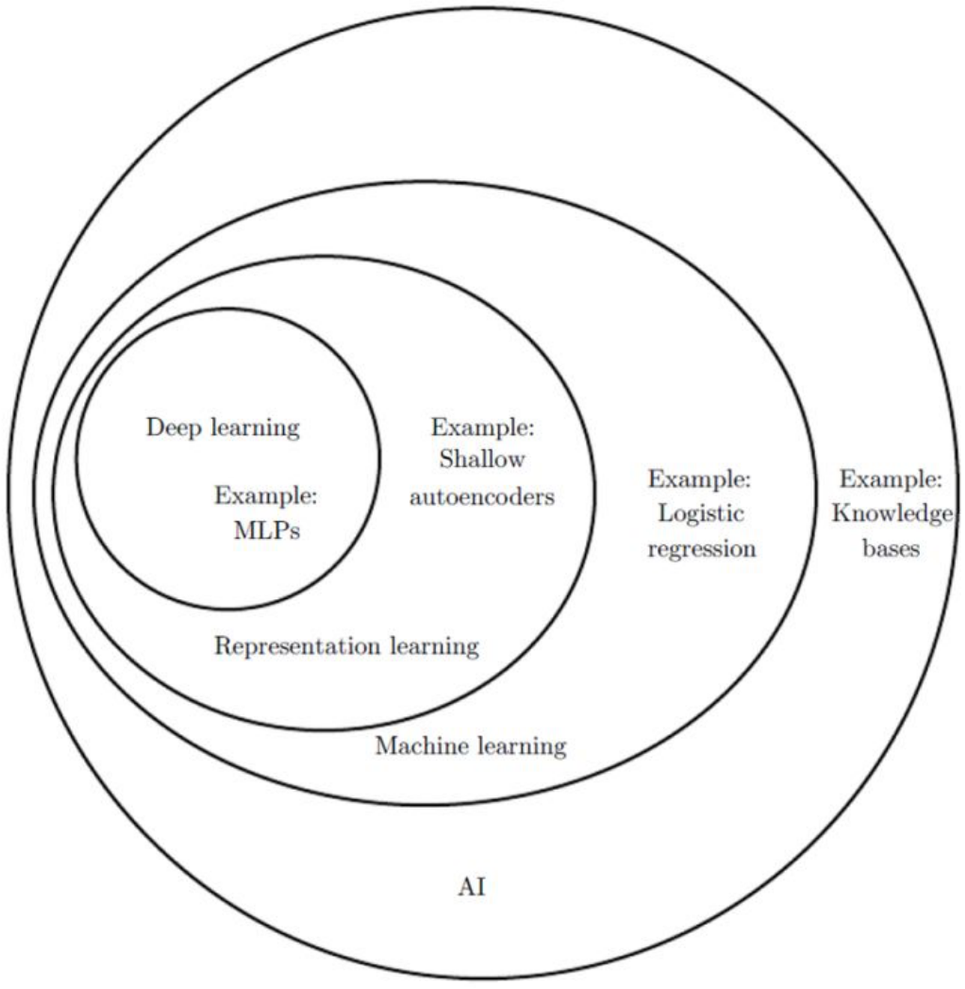
What's Deep Learning (DL)?

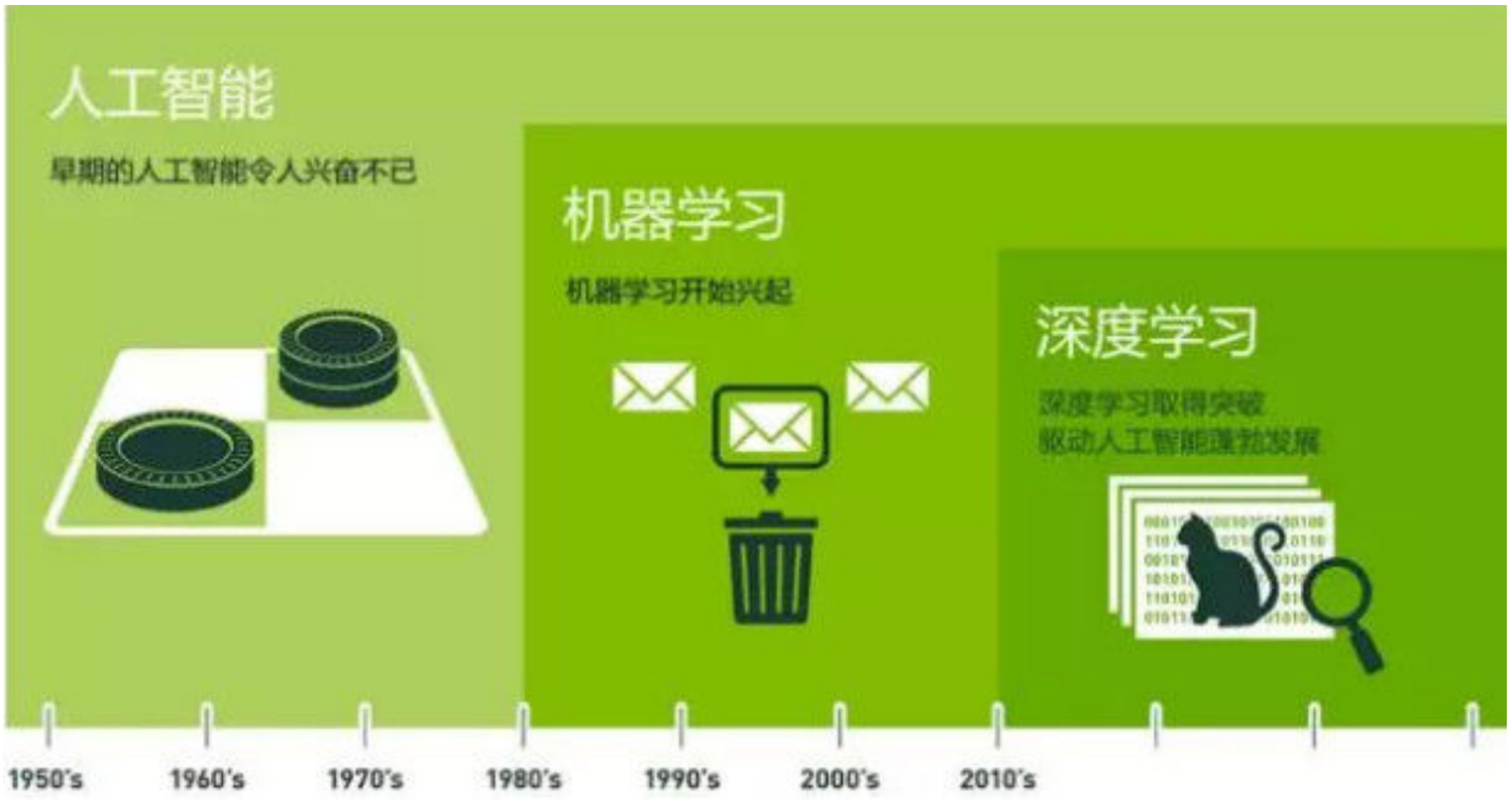






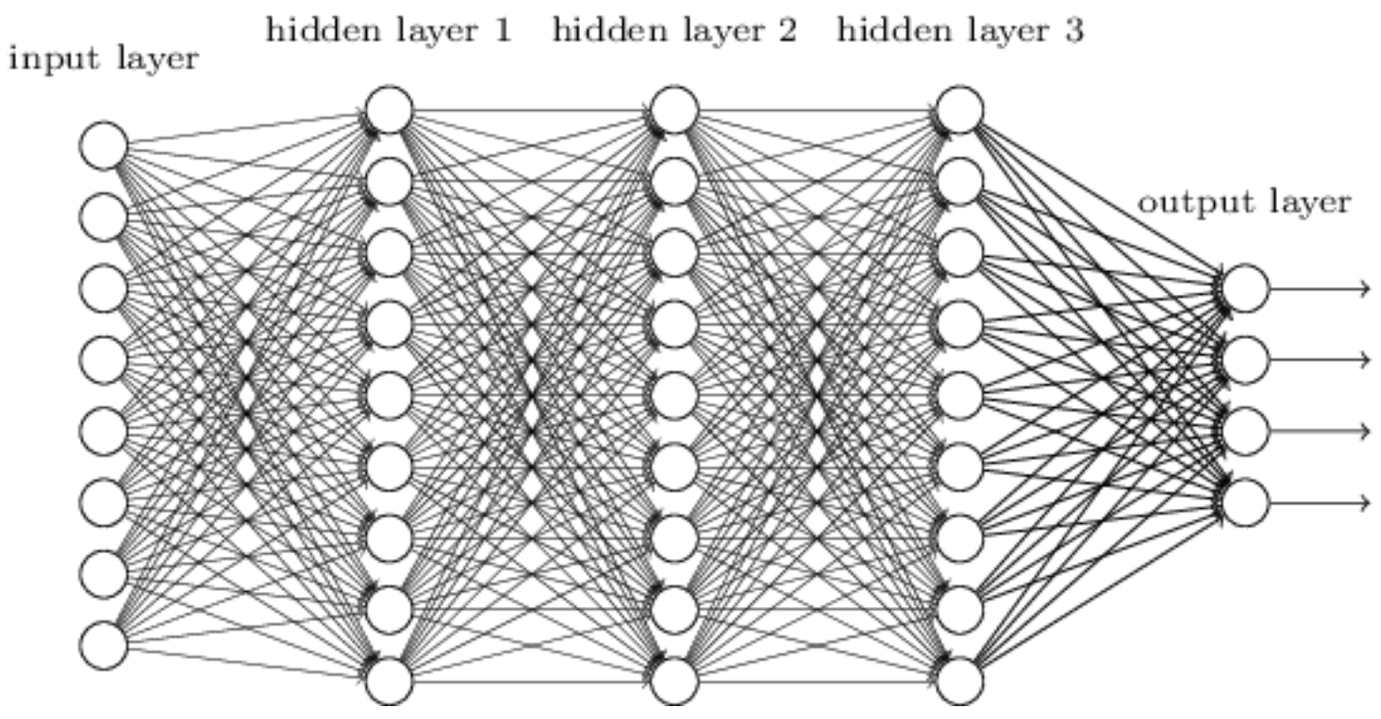
Machine Learning vs. Deep Learning





Deep Learning – WTF?

1. In contrast to **standard** machine learning,
2. Representation learning attempts to automatically learn good features or representations
3. Deep learning algorithms attempt to learn (multiple levels of) representations (here: h1,h2,h3) and an output (h4)
   - ❑ From "raw" inputs x （图像或文本）
   - ❑ (e.g. *sound, pixels/像素, words /文字*)



Huh?....

- Don’t worry, we’ll come back to that shortly….

On the history of "Deep Learning"

1. We will focus on different kinds of neural networks

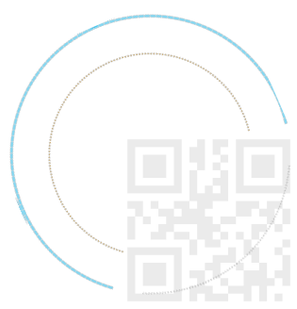
2. The dominant model family inside deep learning

3. Only clever terminology for stacked logistic regression units?

   - Maybe, but interesting modeling principles (end-to-end) and actual connections to neuroscience in some cases.
   - Recently: **Differentiable Programming** – becomes clear later

4. We will not take a historical approach but instead focus on methods which work well (e.g., RNN)on NLP problems now

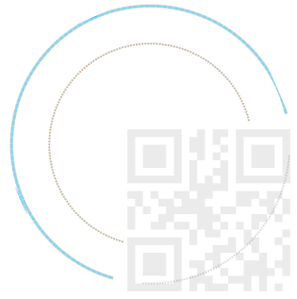
   - For a long history of deep learning models (starting ~1960s), see: Deep Learning in Neural Networks: An Overview by Jürgen Schmidhuber



Reasons for Exploring Deep Learning

Reasons for Exploring Deep Learning

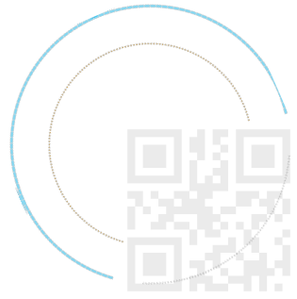
1. Manually designed features are often over-specified, incomplete and take a **long time to design and validate**
2. Learned Features are easy to **adapt**, fast to learn
3. Deep learning provides a very **flexible**, (almost?) **universal**, **learnable** framework for representing world, visual and linguistic information.
4. Deep learning can learn unsupervised (from raw text) and supervised (with specific labels like positive/negative)



Reasons for Exploring Deep Learning

1. In ~2010 deep learning techniques started outperforming other machine learning techniques. Why this decade?
2. Large amounts of training data favor deep learning (数据！！)
3. Faster machines and multicore CPU/GPUs favor Deep Learning
4. New models, algorithms, ideas

- ❑ Better, more flexible learning of intermediate representations
- ❑ Effective end-to-end joint system learning
- ❑ Effective learning methods for using contexts and transferring between tasks
- ❑ Better regularization and optimization methods
- ❑ Improved performance (first in speech and vision, then NLP)



What's Deep Learning (DL)?

1. Deep learning is a subfield of machine learning
2. Most machine learning methods work well because of human-designed representations and input features
3. Machine learning becomes just optimizing weights to best make a final prediction

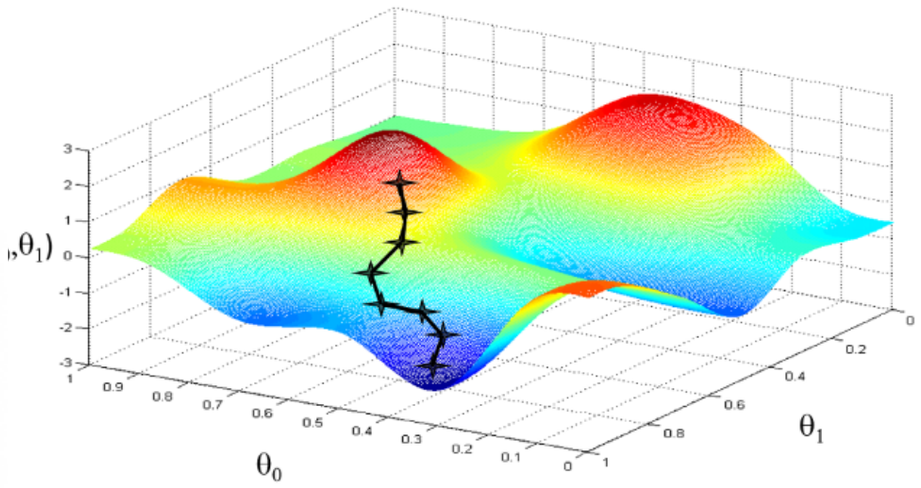
PYTORCH

theano

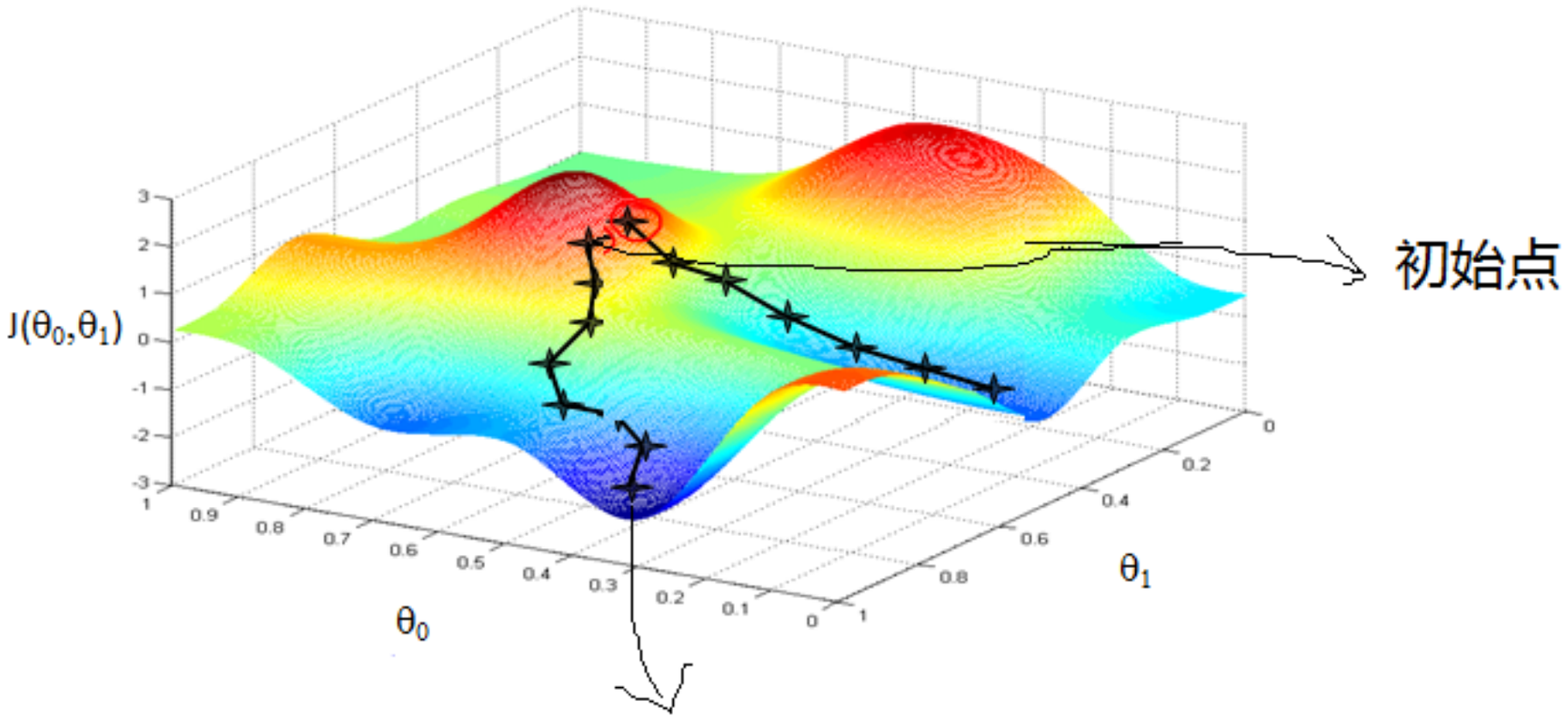
Spark

TensorFlow

neon



What's Deep Learning (DL)?



Natural Language Processing with Deep Learning

玖强

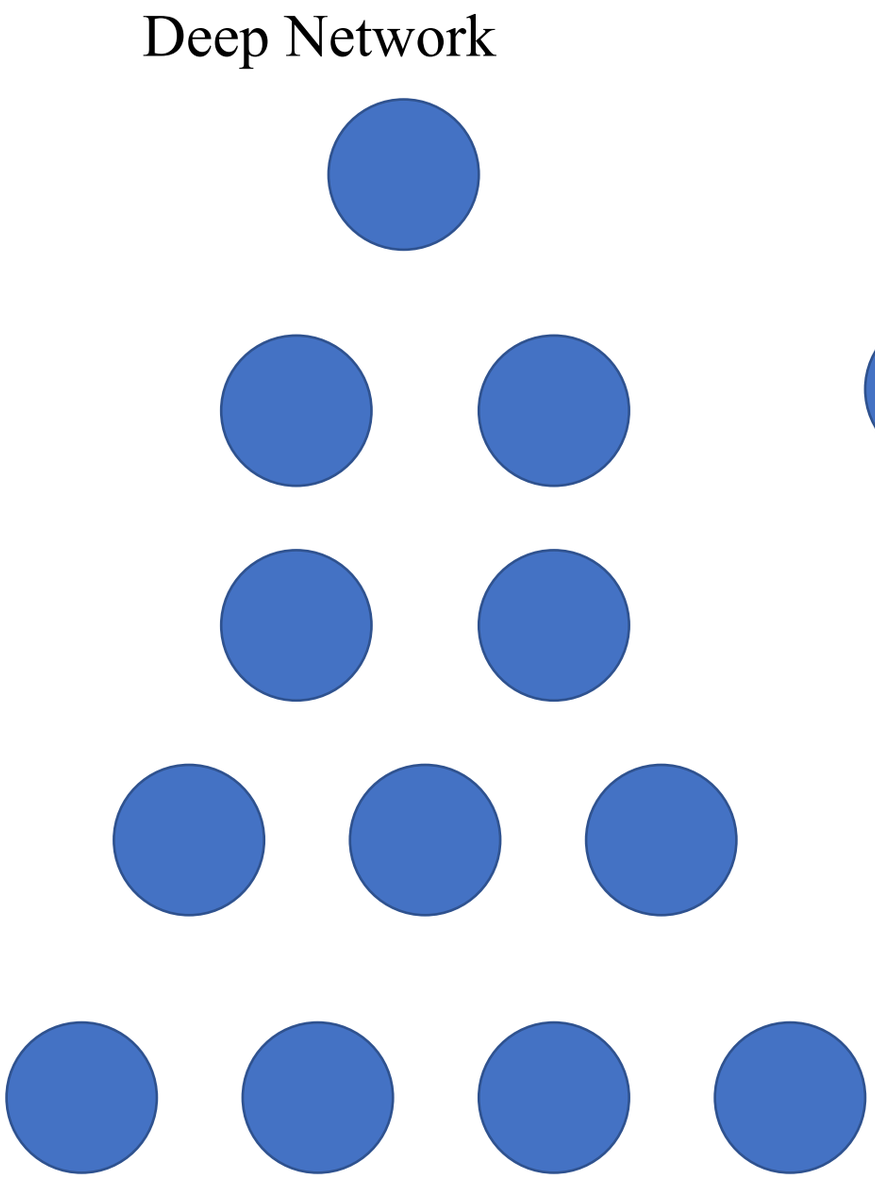
Huh?....

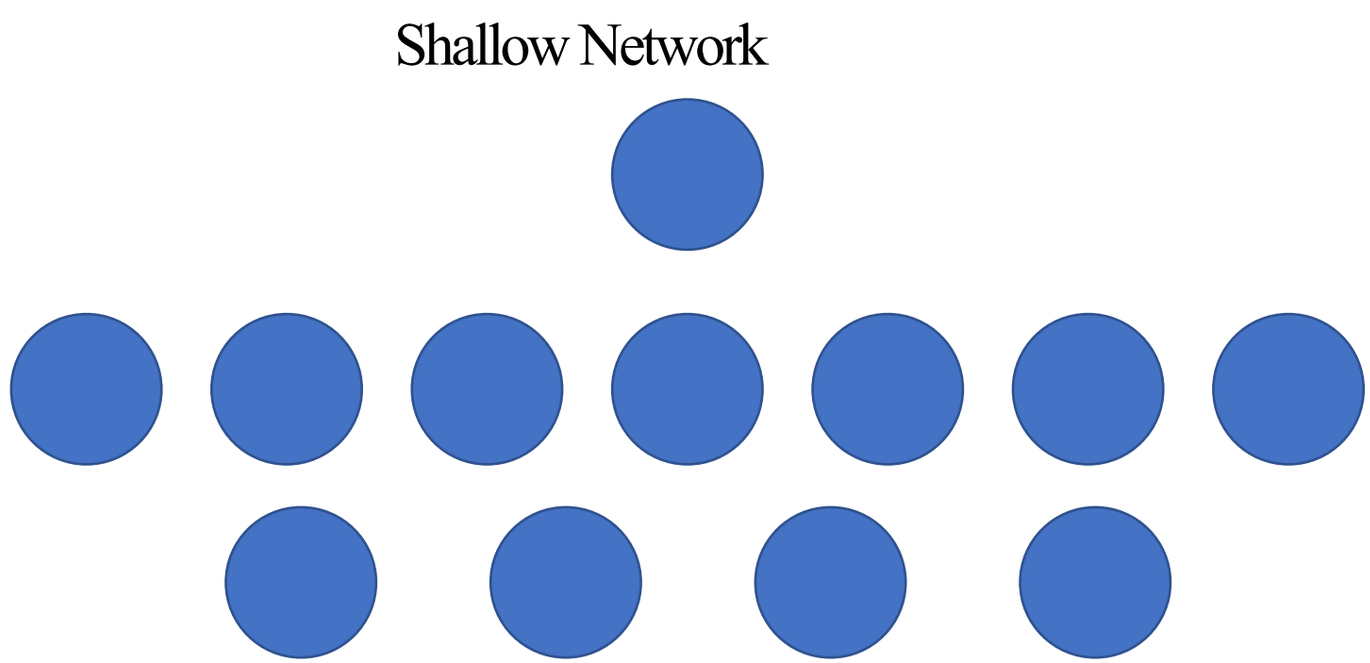
- Don’t worry, we’ll come back to that shortly….

Deep vs Shallow Networks

- Given the same number of non-linear (neural network) units, a deep architecture is more expressive than a shallow one (Bishop 1995)
- Two layer (plus input layer) neural networks have been shown to be able to approximate **any** function
- However, functions compactly represented in k layers may require exponential size when expressed in 2 layers

Deep vs Shallow Networks





Shallow (2 layer) networks need a lot more hidden layer nodes to compensate for lack of expressivity

In a deep network, high levels can express combinations between features learned at lower levels

Deep vs Shallow Networks

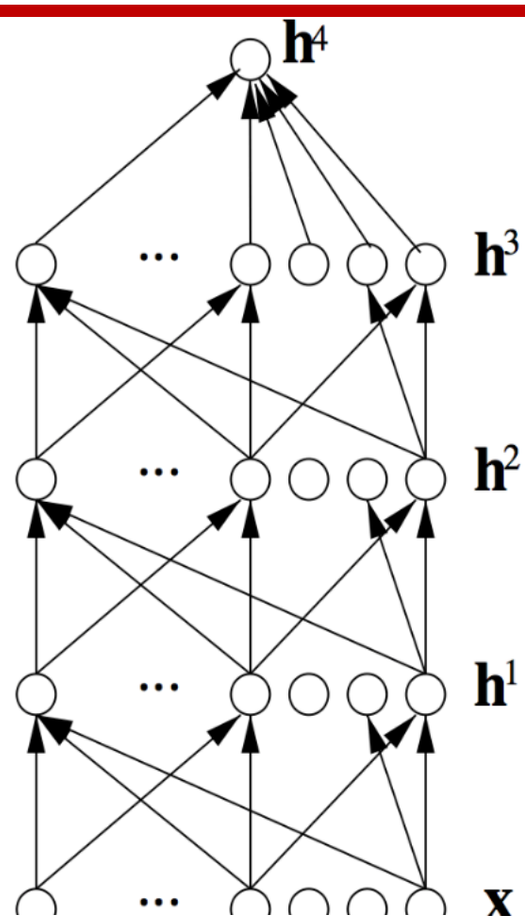
Output layer

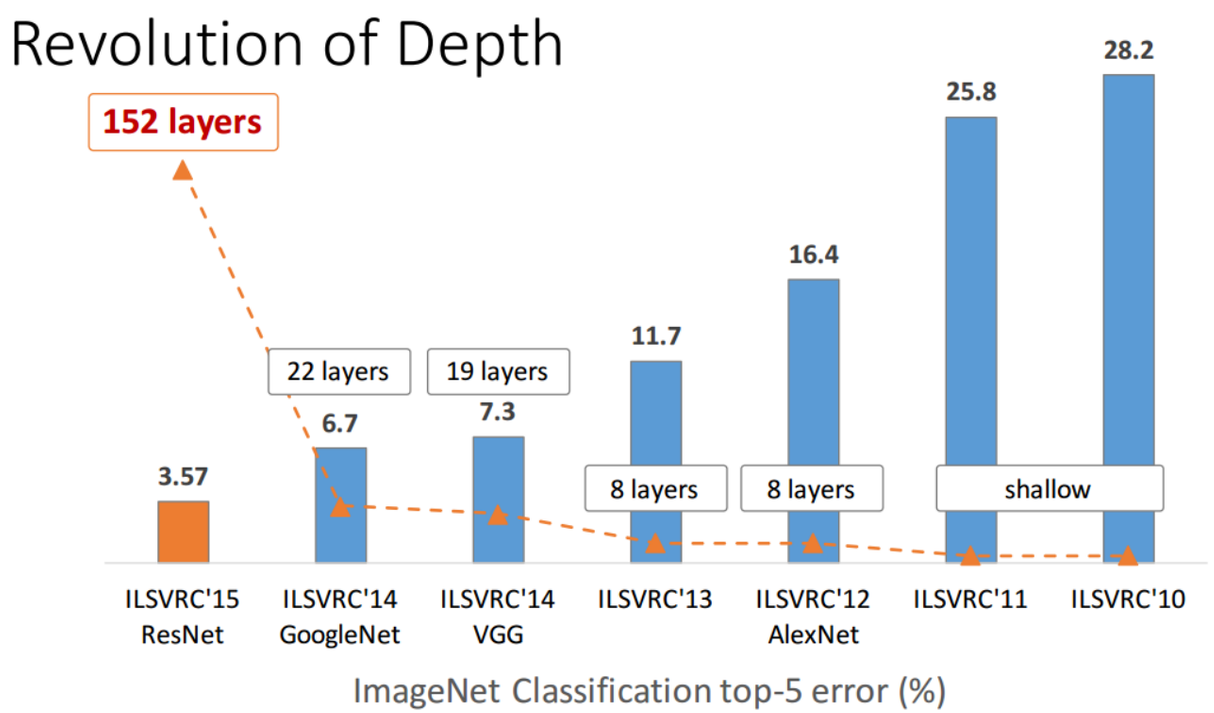
Here predicting a supervised target

Hidden layers

These learn more abstract representations as you head up

Input layer





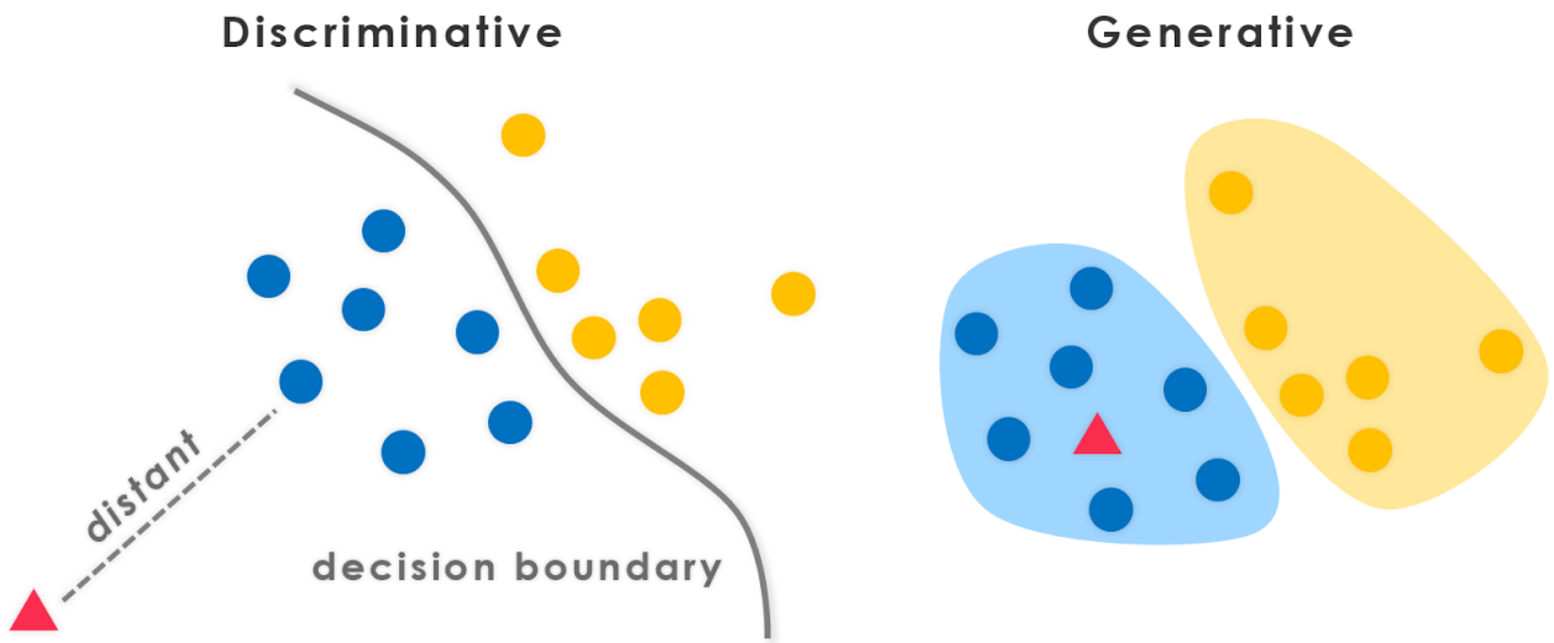
Unsupervised Training

- Today, most practical, good NLP& ML methods require labeled training data (i.e., supervised learning)
  - But almost all data is unlabeled
  - Far more un-labeled data in the world (i.e. online) than labeled data:
  - Websites; Books; Videos; Pictures
  - Fortunately, a good model of observed data can really help you learn classification decisions

- Deep networks take advantage of unlabelled data by learning **good representations** of the data **through unsupervised learning**
- Humans learn initially from unlabelled examples
  - Babies learn to talk without labeled data

Discriminative Vs Generative Models

- 2 types of classification algorithms



- 1. Generative – Model Joint Distribution
  - p(Class /\ Data)
- 2. Discriminative – Conditional Distribution
  - p(Class\Data)

Applications

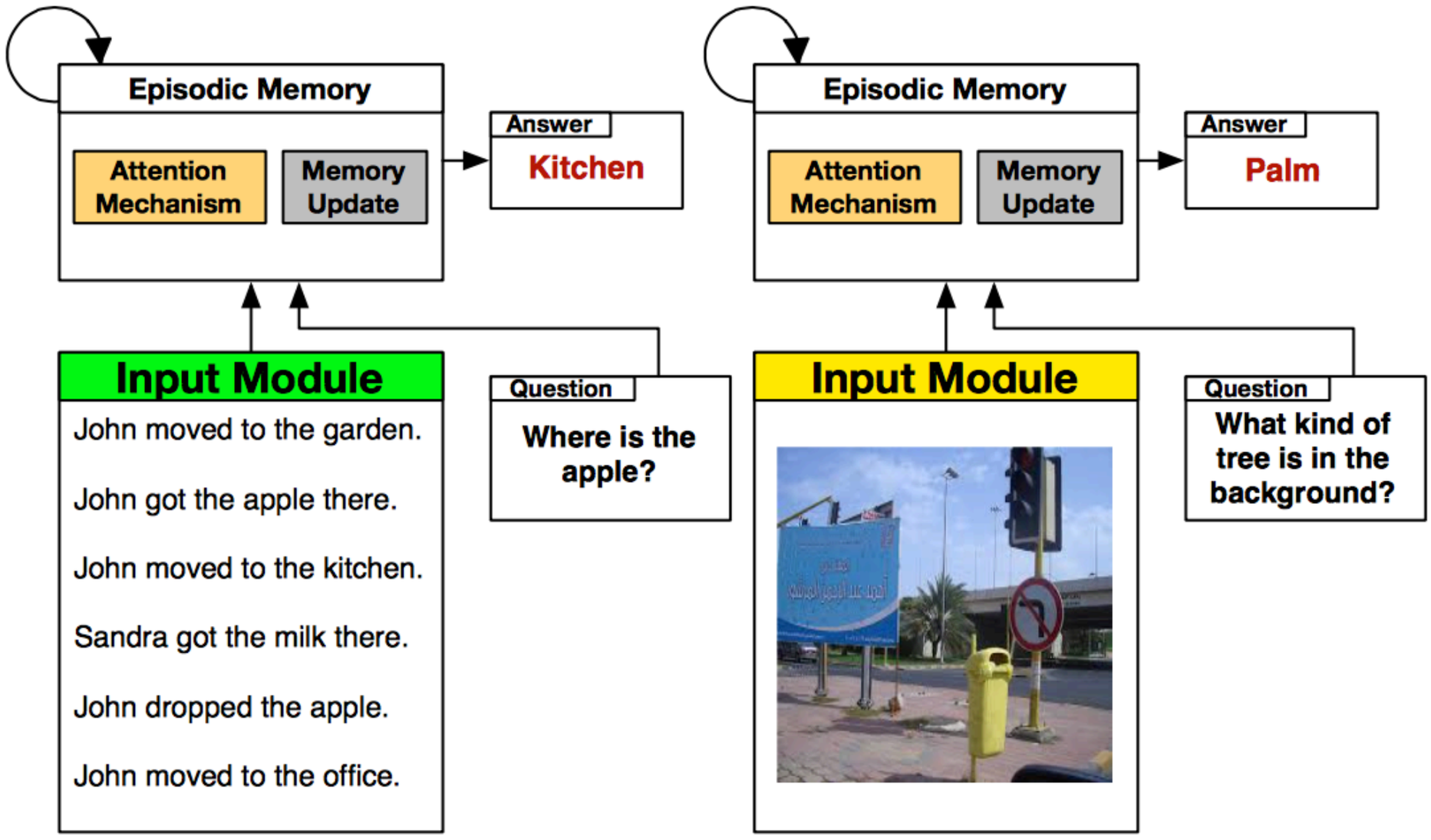
COMMON TASKS OF DEEP LEARNING IN NLP

- 自然语言处理/Natural Language Processing (NLP)
- 词性标注/Labeling or tagging: drink(v.) milk(n.) at (p.) night (n.)
- 分词/ Word Segmentation：小明/很/开心. 分词工具:中文/Jieba
- 文本分类/Text classification：体育/科技/娱乐, 高兴/悲伤/平静
- 自动文摘/Text summarization：长文归纳成短文
- 机器翻译/Machine translation：很重要，很火，很成熟
- 问答系统/Question answering： Siri和Watson
- 图像自动描述生成/Image captioning：和图像结合是未来的趋势
- 基于文本的检索/Text-based image/video retrieval：很有实际价值

COMMON TASKS OF DEEP LEARNING IN NLP

Deep Learning Algorithms	NLP Usage
Neural Network – NN (feed)	•Part-of-speech Tagging •Tokenization •Named Entity Recognition •Intent Extraction
Recurrent Neural Networks -(RNN)	•Machine Translation •Question Answering System •Image Captioning
Recursive Neural Networks	•Parsing sentences •Sentiment Analysis •Paraphrase detection •Relation Classification •Object detection
Convolutional Neural Network -(CNN)	•Sentence/ Text classification •Relation extraction and classification •Spam detection •Categorization of search queries •Semantic relation extraction

Question Answering

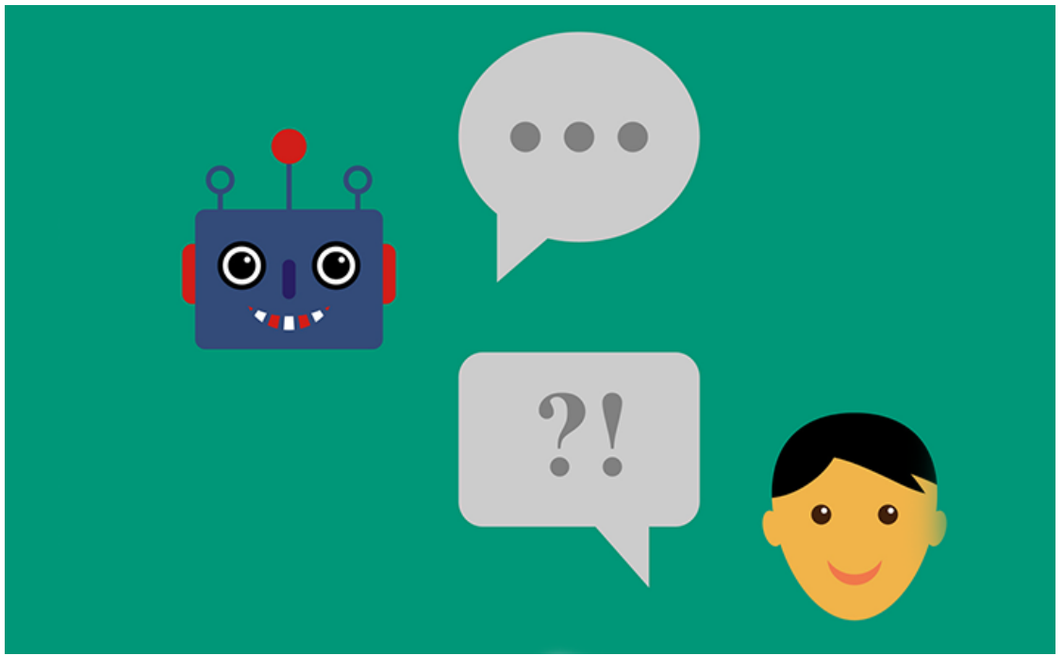


(a) Text Question-Answering

(b) Visual Question-Answering

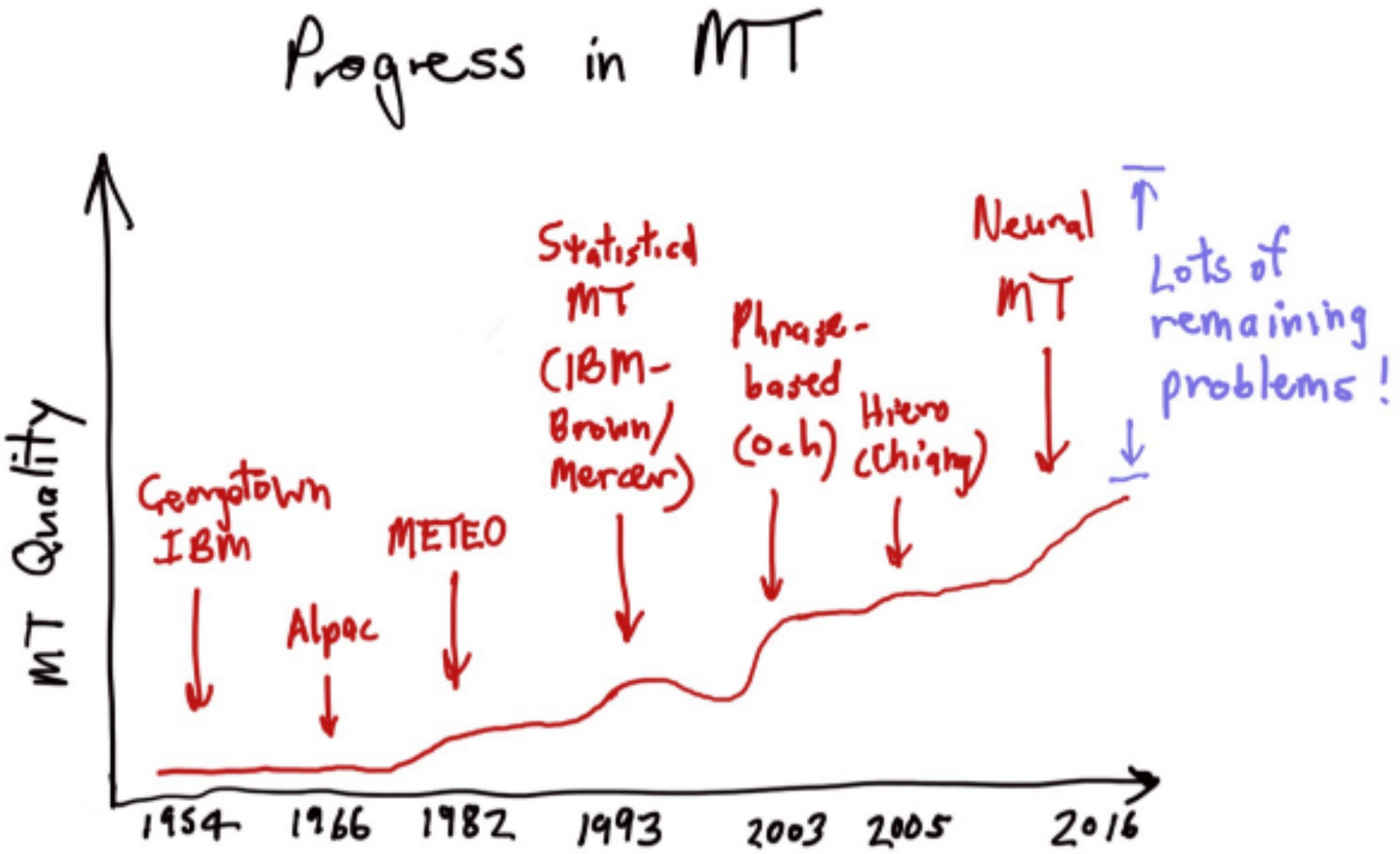
Dialogue agents / Response Generation

- A simple, successful example is the auto-replies available in the Google Inbox app
- An application of the powerful, general technique of Neural Language Models, which are an instance of Recurrent Neural Networks



Machine Translation

- Source sentence is mapped to vector, then output sentence generated [Sutskever et al. 2014, Bahdanau et al. 2014, Luong and Manning 2016]



Conclusion

- 下节课我们将学习数学理论基础

时间	课时安排
2018/2/6	第一课 NLP发展历史介绍和展望 1.NLP发展现状 2.传统NLP方法面临的挑战 3.Big Data和Deep Learning给NLP带来的变革和机遇 4.NLP的发展趋势，以及和各行各业的结合应用
2018/2/13	第二课 数学理论基础 1. 概率和信息论 2. 监督学习、半监督学习和非监督学习 3. 分类与回归模型
2018/2/20	第三课 自然语言基础 1. Word vector与Word embedding 2. 什么是分词、词性标注、依存句法分析等？如何利用开源工具包完成 3. 什么是统计自然语言处理？

END