Large Scale Data Analysis Using Deep Learning

Introduction to Deep Learning

U Kang
Seoul National University

In This Lecture

- Overview of deep learning
- History of deep learning and its recent advances

Outline

- ☐ **Overview of Deep Learning**
- ☐ Historical Trends in Deep Learning

Deep Learning

- Branch of machine learning based on a set of algorithms that attempt to model high level abstractions in data
- Key technology in recent ‘AI revolution’

Artificial Intelligence (AI)

- Quickly growing field with many practical applications and active research topics

- Goal: intelligent software to automate routine labor, understand speech or images, make diagnoses in medicine, and support basic scientific research

Approaches to AI

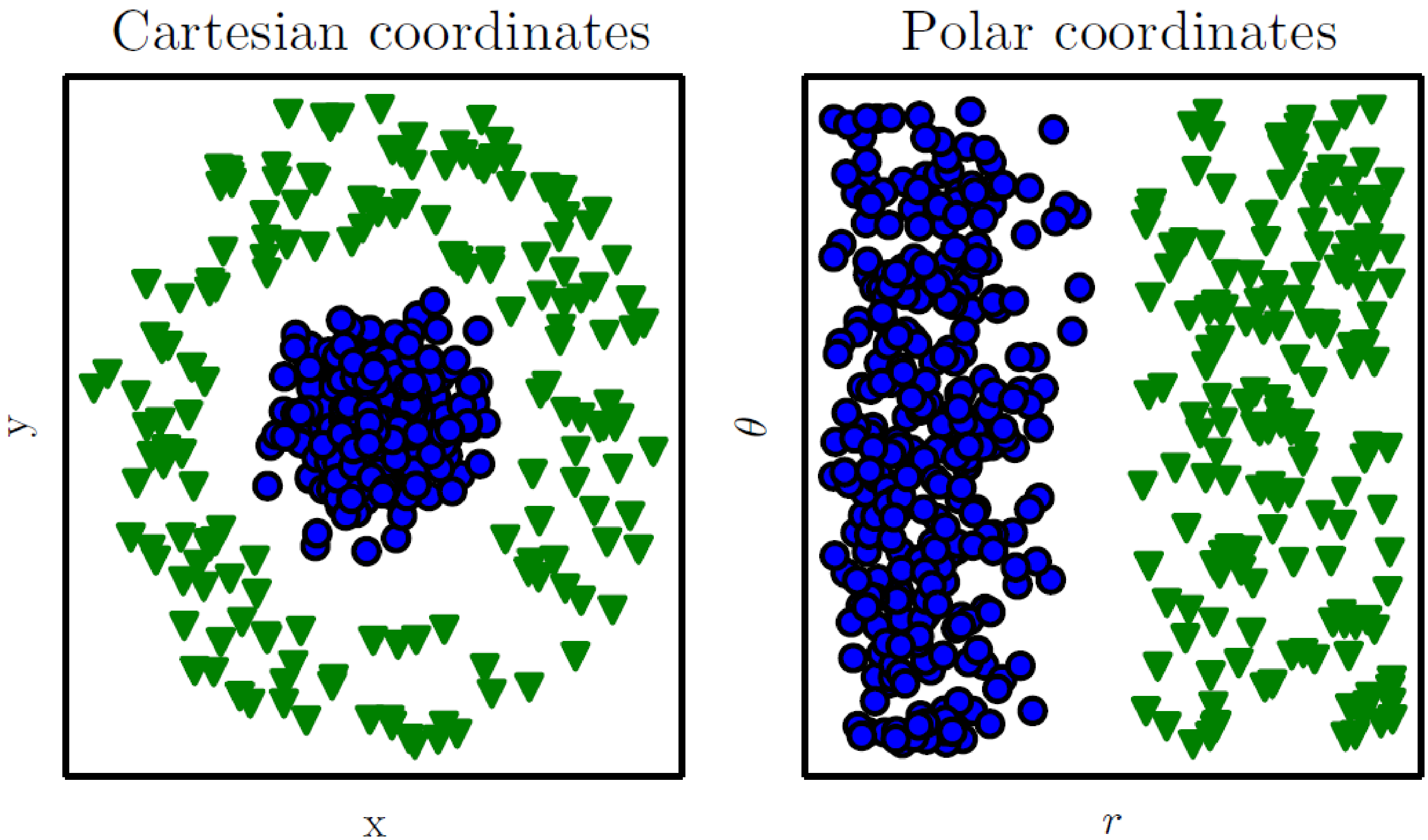
- Knowledge base approach
  - Hard-code knowledge about the world in formal language
  - A computer can reason about statements in these formal languages using logical inference rules

- Problem: not flexible, and hard to get exact knowledge

Machine Learning (ML)

- ML alg. acquires its own knowledge by extracting patterns from raw data
  - E.g., naïve Bayes can separate legitimate e-mail from spam e-mail, through training with e-mails and their labels
- ML depends heavily on the representation of the data
  - E.g., in the above e-mail example, each e-mail is represented by the set of words contained in it

Importance of Representations



Representation Learning

- It is difficult to know what feature should be extracted
  - E.g., features to detect cars in photographs?

- Representation learning: discover not only the mapping from representation to output, but also the representation itself

Challenges in Representation Learning

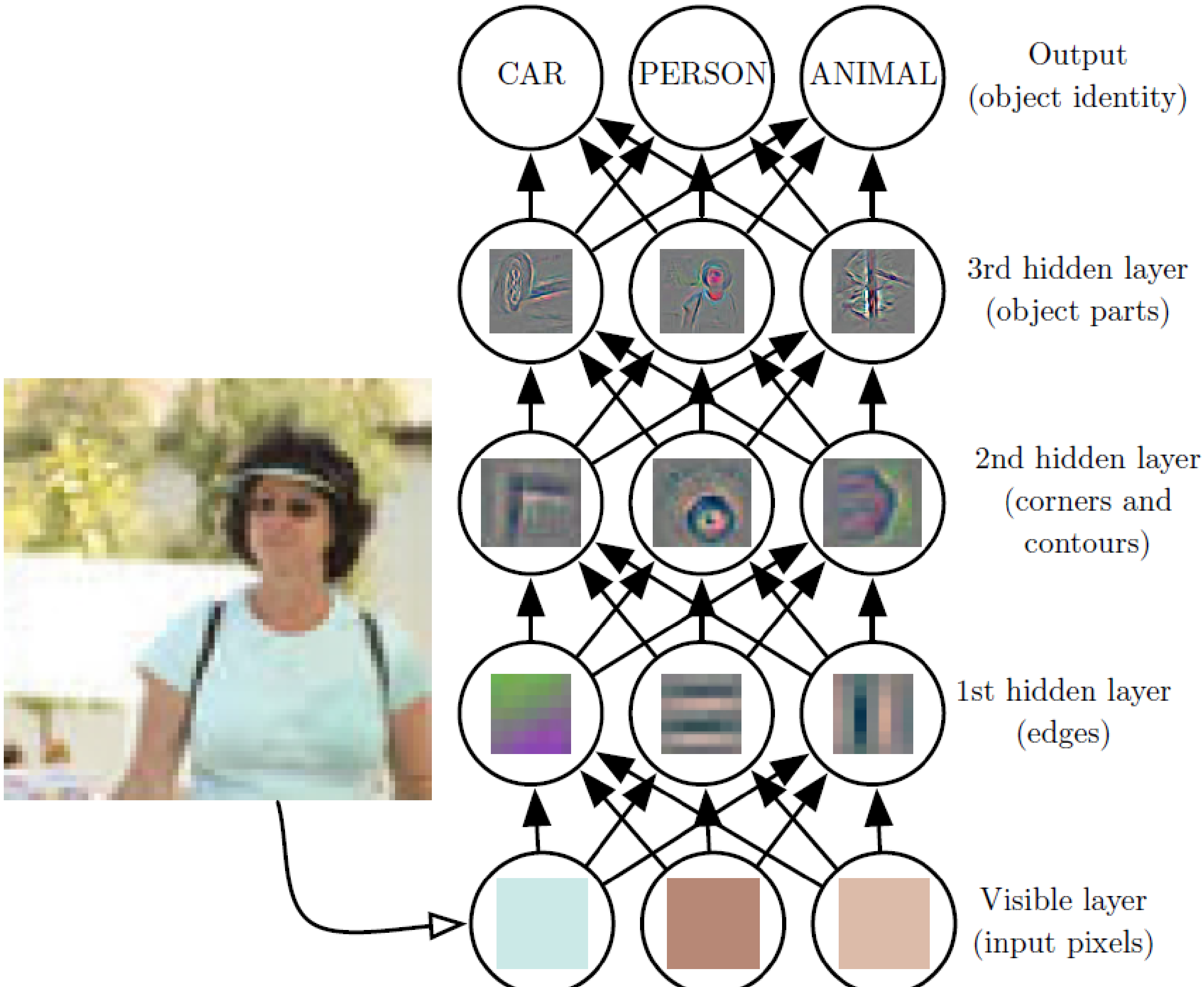
- How to separate factors of variation that explain the observed data?
  - A factor means a separate source of influence
  - E.g., image: a red car may look black at night
  - E.g., speech: a word may sound differently based on the speaker's age, sex, and accent

Deep Learning Representation

- Deep learning solves the problem in representation learning by introducing representations that are expressed in terms of other simple representations

- Deep learning builds complex concepts out of simpler concepts

Deep Learning Representation



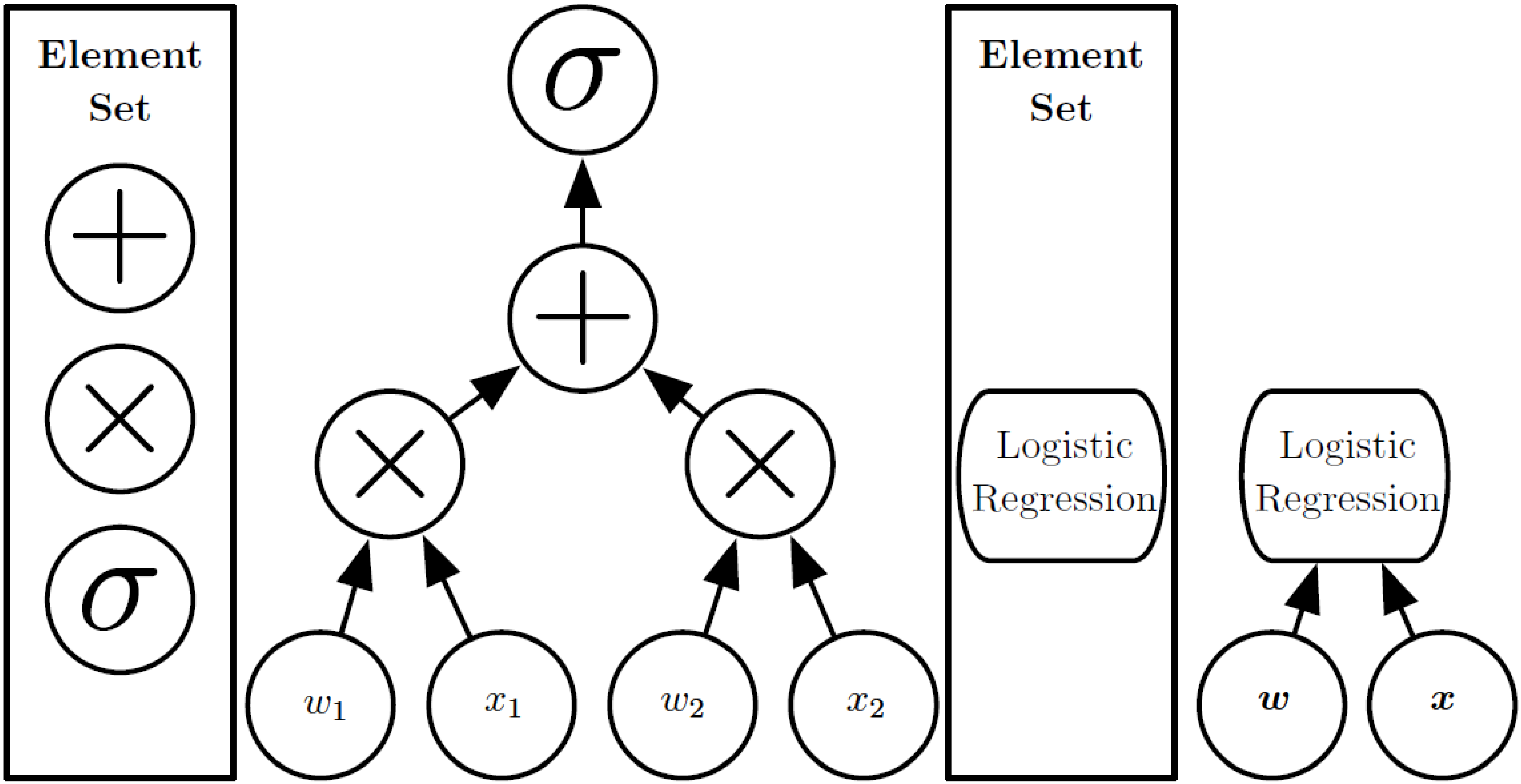
Multi-layer perceptron

Perspectives on Deep Learning

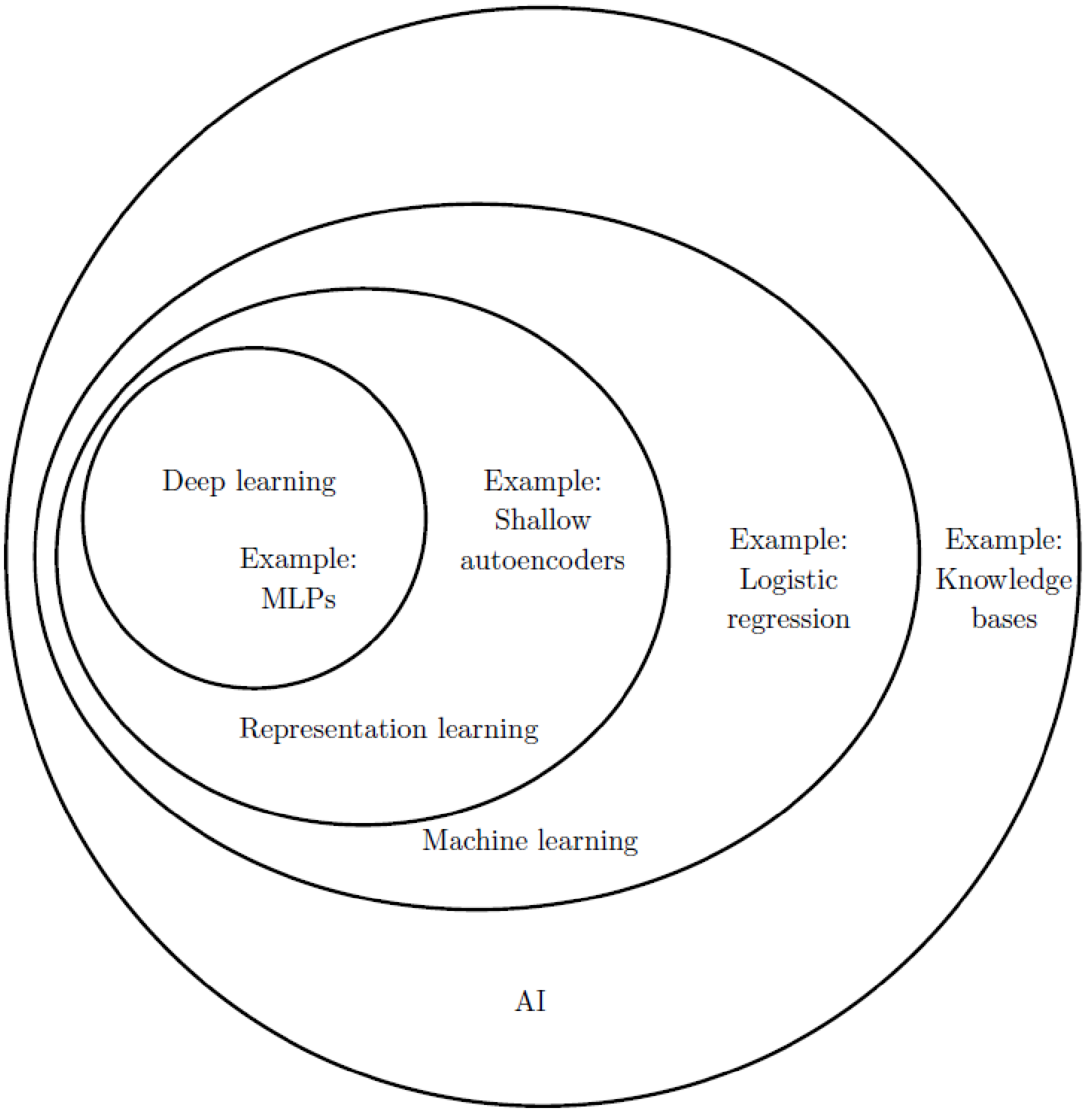
1. Learns the right representation
2. Depth allows the computer to learn a multi-step computer program
   - Each layer can be thought of as the state of computer's memory after executing another set of instructions
   - Networks with greater depth can execute more instructions in sequence
   - Sequential instructions offer great power since later instructions can refer back to the results of earlier instructions

Measuring the Depth of a Model

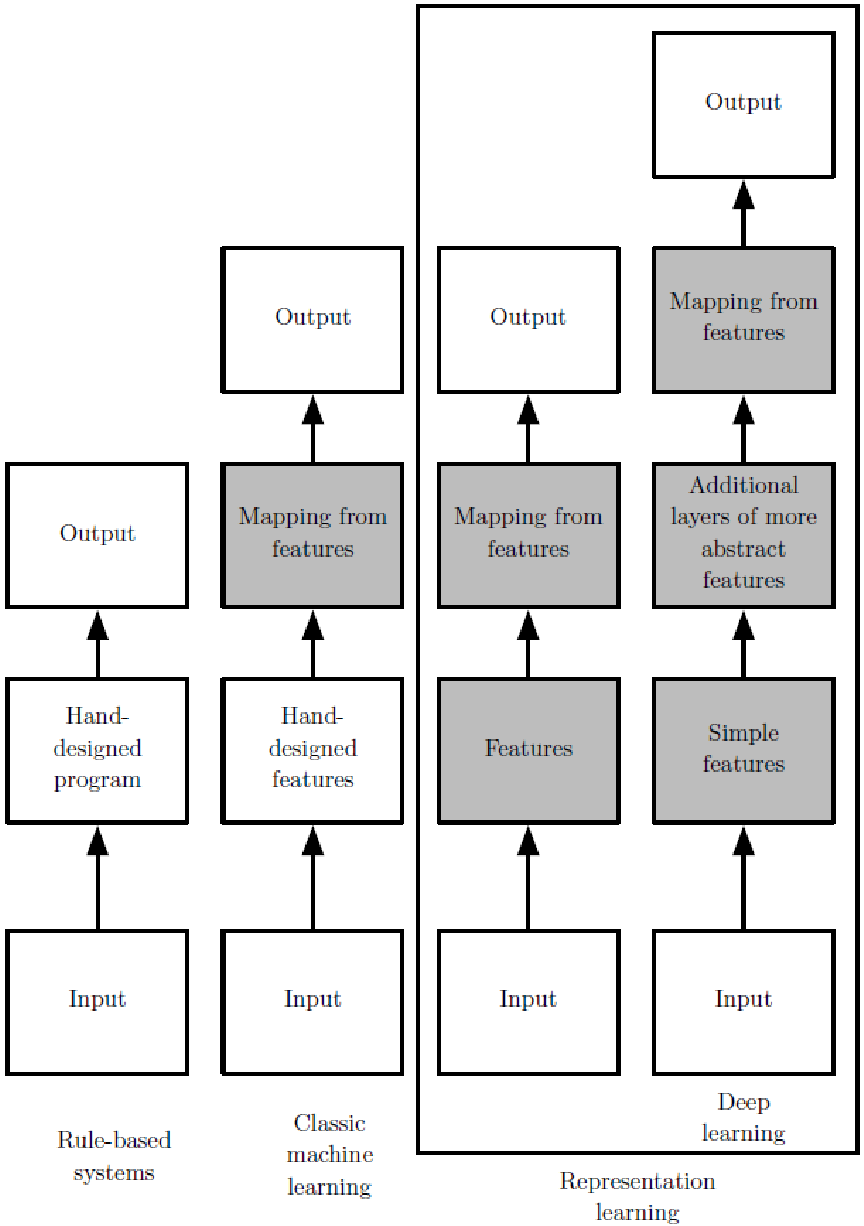
- Computational graph



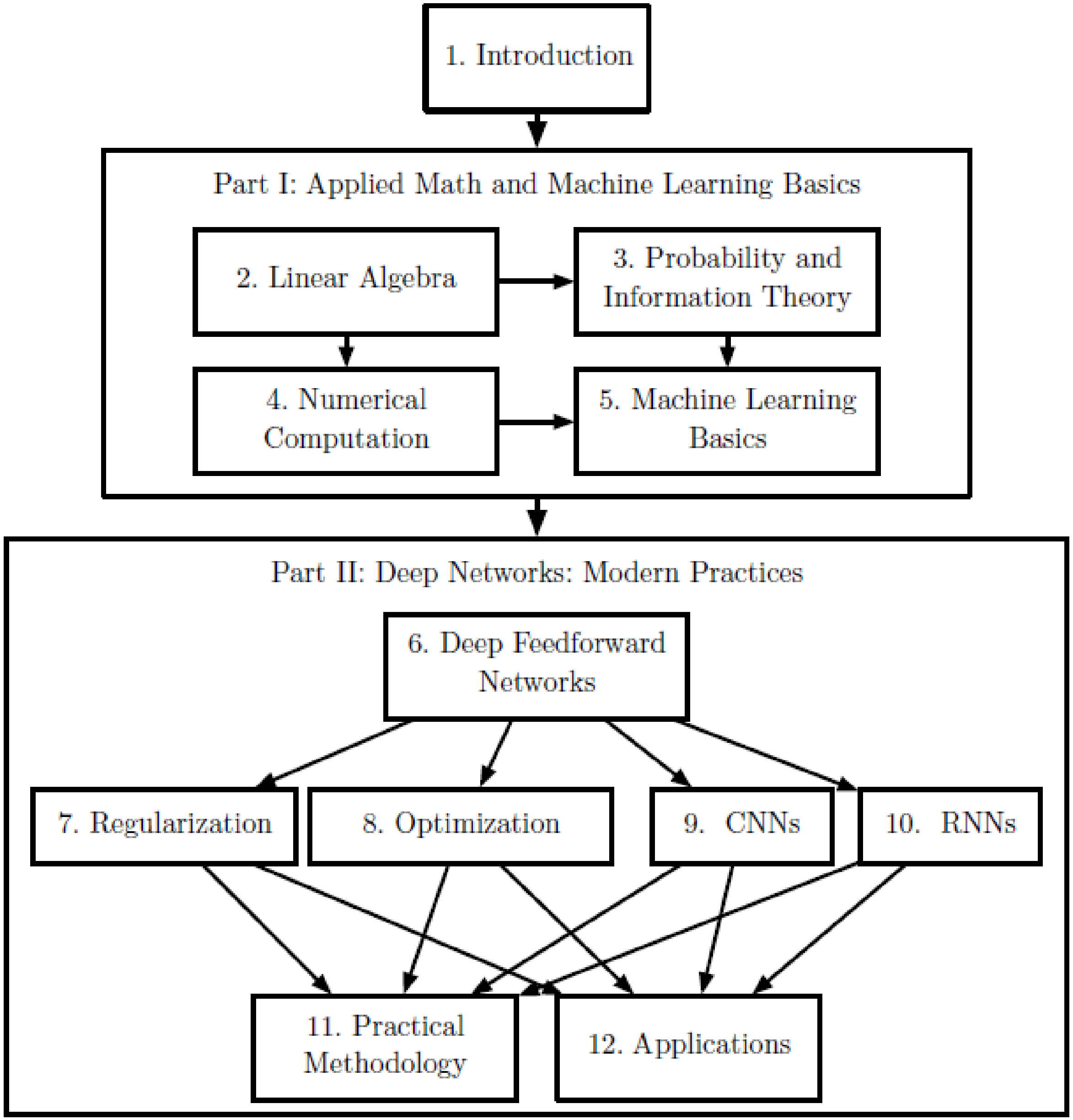
AI hierarchy



Learning Multiple Components



Plan of Study



Outline

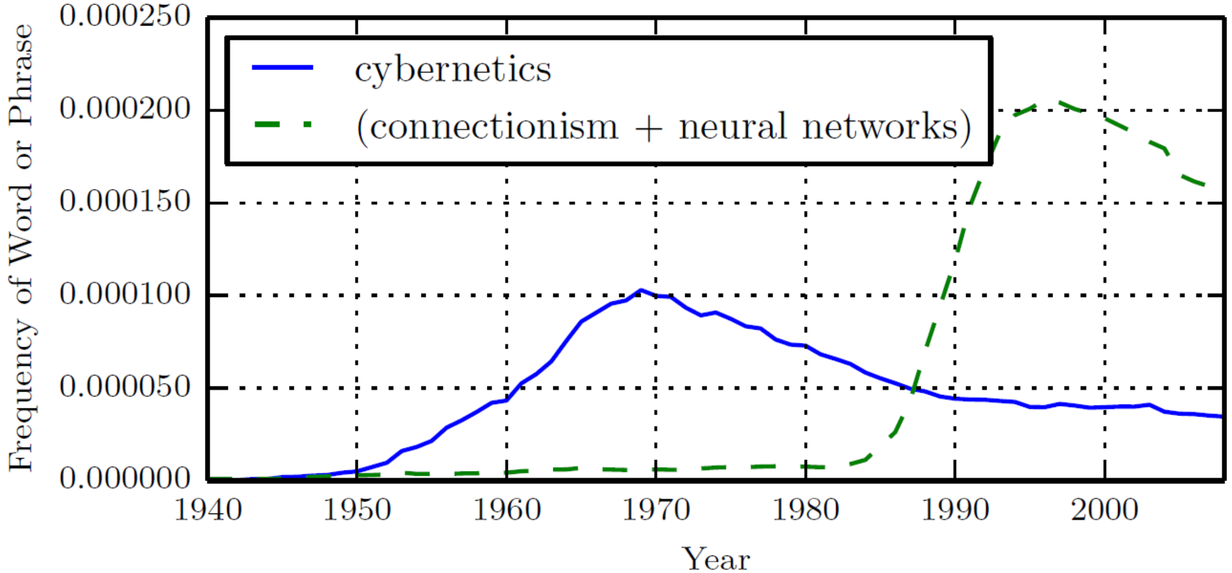
- ☑ Overview of Deep Learning
- ☐ **Historical Trends in Deep Learning**

Key Trends

1. Deep learning has a long and rich history with varying popularity over time
2. Deep learning has become more powerful as the amount of available training data has increased
3. Deep learning models have grown in size over time as computer hardware and software infrastructure for deep learning has improved
4. Deep learning has solved increasingly complicated applications with increasing accuracy over time

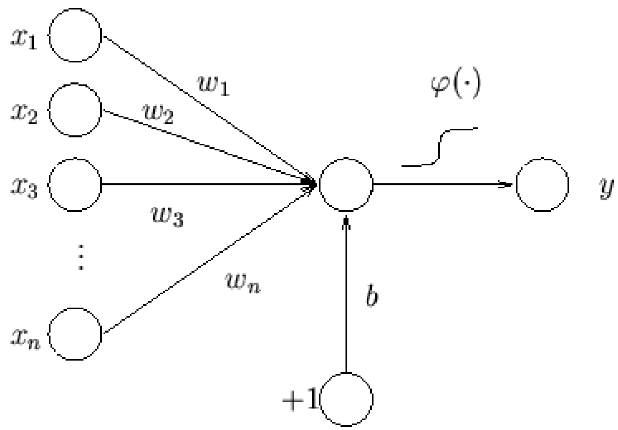
Waves in Deep Learning

- Cybernetics (1940s - 1960s)
  - Theories of biological learning: perceptron
- Connectionism (1980s - 1990s)
  - Back-propagation to train a neural network with one or two hidden layers
- Deep Learning (2006 -)



Cybernetics (1940s - 1960s)

- Theories of biological learning
- Implementations of the first models such as the perceptron allowing the training of a single neuron
- Linear model: $f(x,w) = x_1 w_1 + \ldots + x_n w_n + b$

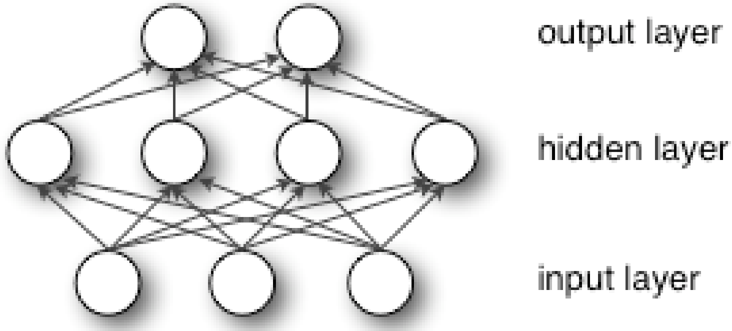


- Limitation: cannot learn the XOR function (Minsky 1969)
  - The first major dip in the popularity of neural network

Connectionism (1980s - 1990s)

- Main idea: a large number of simple computational units can achieve intelligent behavior when networked together

- Universal approximation theorem (Cybenko 1989, Hornik 1991)
  - A feed-forward network with a single hidden layer containing a finite number of neurons can approximate any continuous function
  - It means simple neural networks can represent a wide variety of interesting functions when given appropriate parameters; however, it does not guarantee the algorithmic learnability of those parameters



Connectionism (1980s - 1990s)

- Key concepts arose during connectionism movement of the 1980s
  - Distributed representation
  - Back-propagation
  - Modeling sequences with neural networks
    - RNN, LSTM

- Limitation: believed to be very difficult to train model
  - Especially for ‘deep’ model
  - The second major dip of neural network

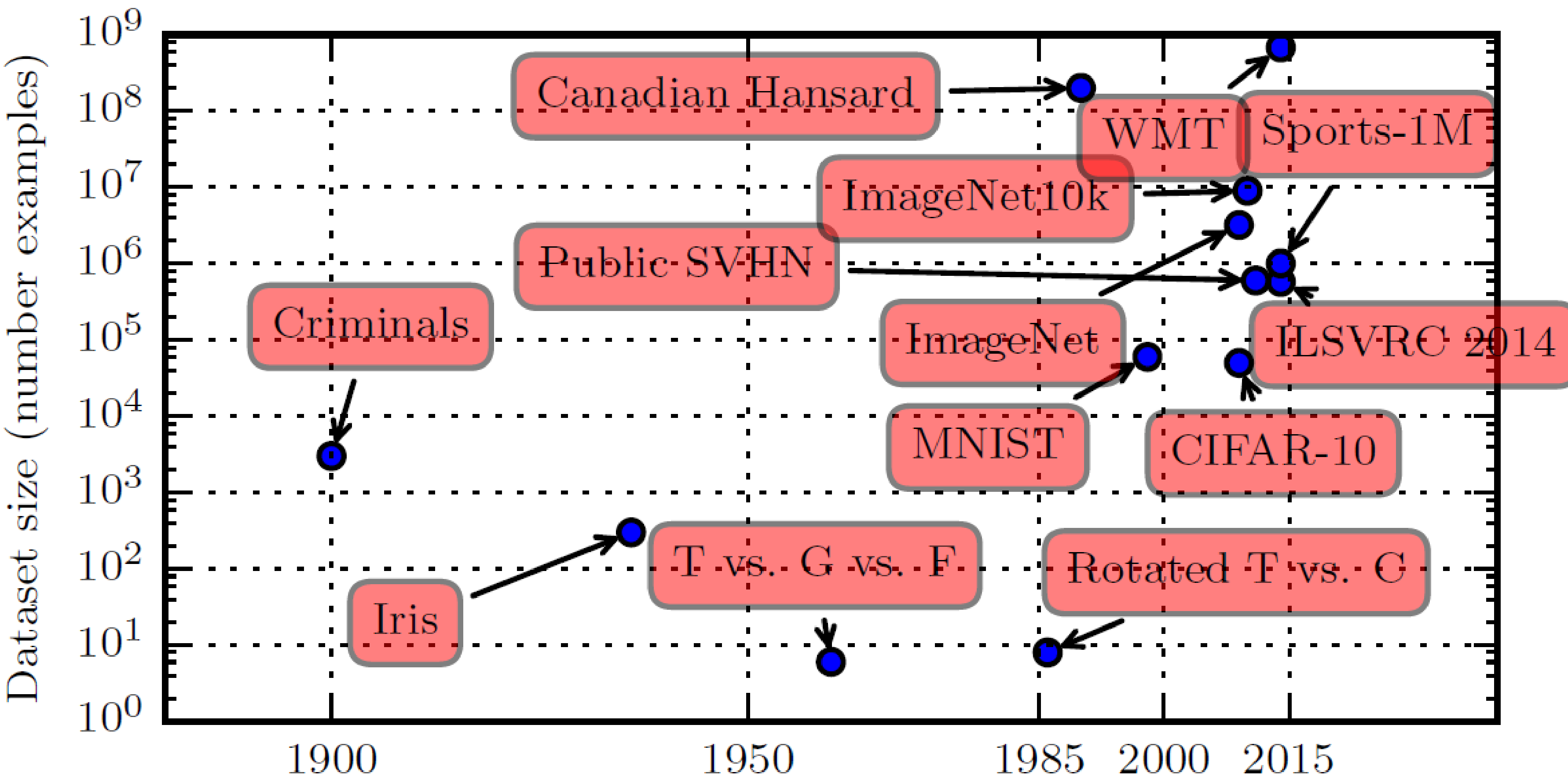
Connectionism (1980s - 1990s)

- Distributed representation
  - Each input to a system should be represented by many features, and each feature should be involved in the representation of many possible inputs
  - E.g.,
    - A vision system can recognize cars, trucks, and birds, and these objects can each be red, green, or blue
    - One way of representing these inputs is to have a separate neuron that activates for each of the nine possible combinations
    - Distributed representation: three neurons for objects, three neurons for colors => total six neurons

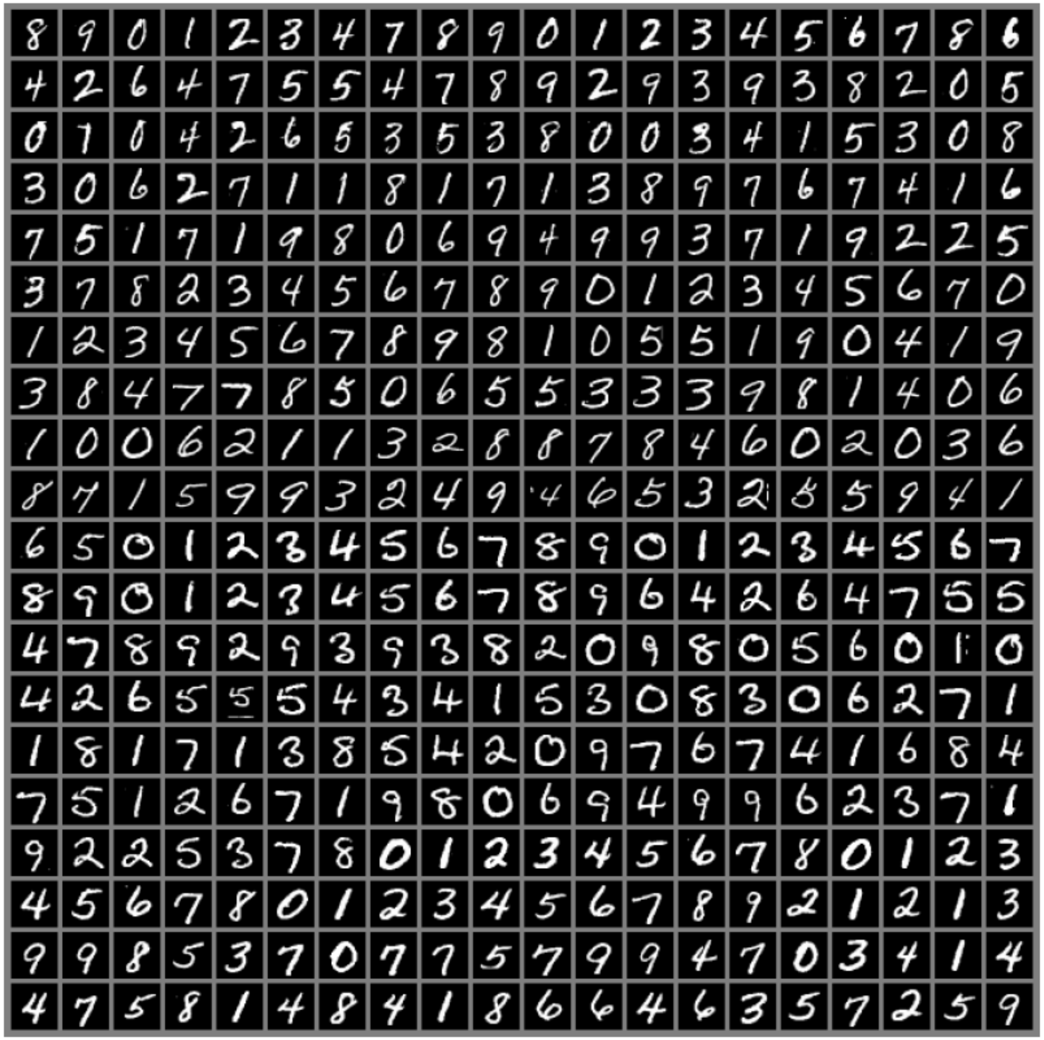
Deep Learning (2006-)

- New technologies that enabled training deep neural networks
  - New unsupervised learning techniques
    - Deep belief network (Hinton, 2006): greedy layer-wise pretraining
  - New activation functions (e.g., rectified linear unit)
  - Powerful computing architecture
    - Clusters and GPU

Growing Datasets



MNIST Dataset

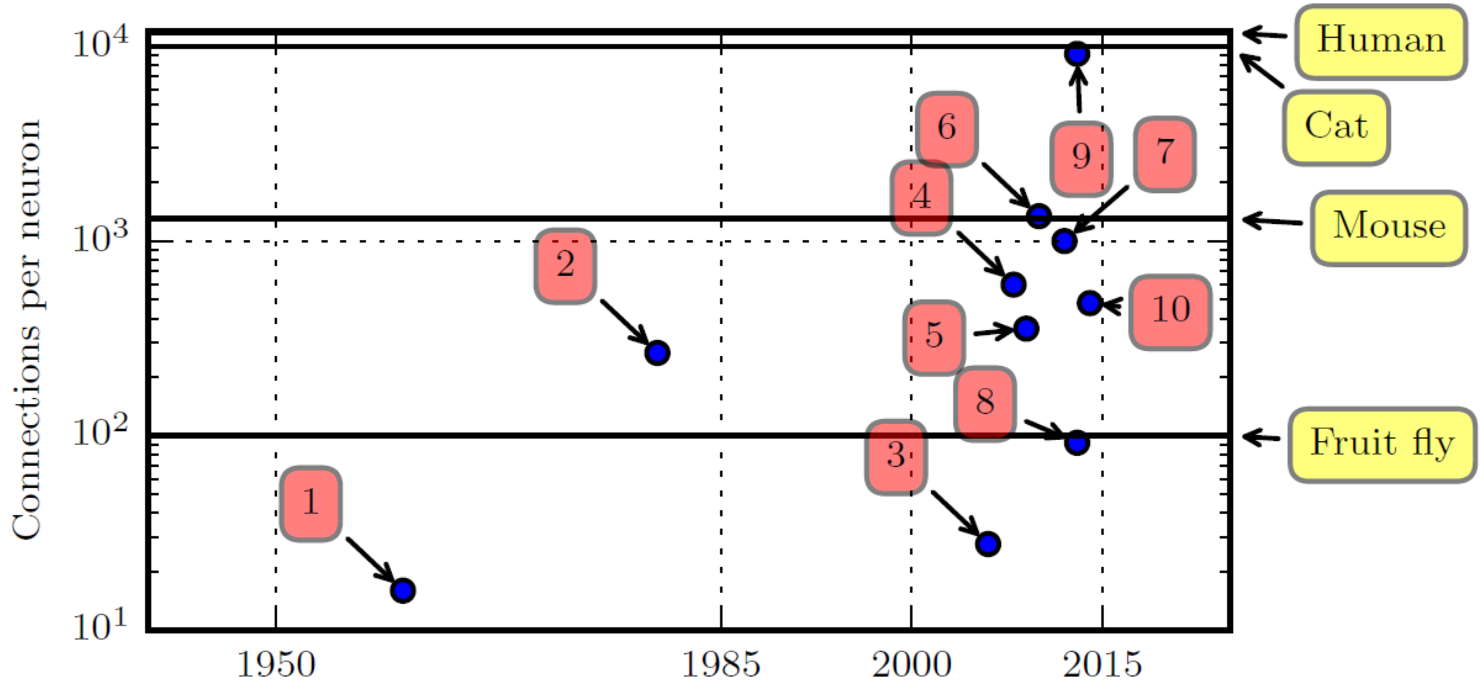


Why Growing Datasets Matters?

- The age of "Big Data" has made machine learning much easier because the key burden of statistical estimation (generalize well to new data after observing only a small amount of data) has been considerably lightened
- Rule of thumb
  - A supervised deep learning algorithm would achieve acceptable performance with ~5000 labeled examples per category
  - Deep learning algorithm would exceed human performance when trained with a dataset with ≥ 10 million labeled examples

Increasing Model Sizes

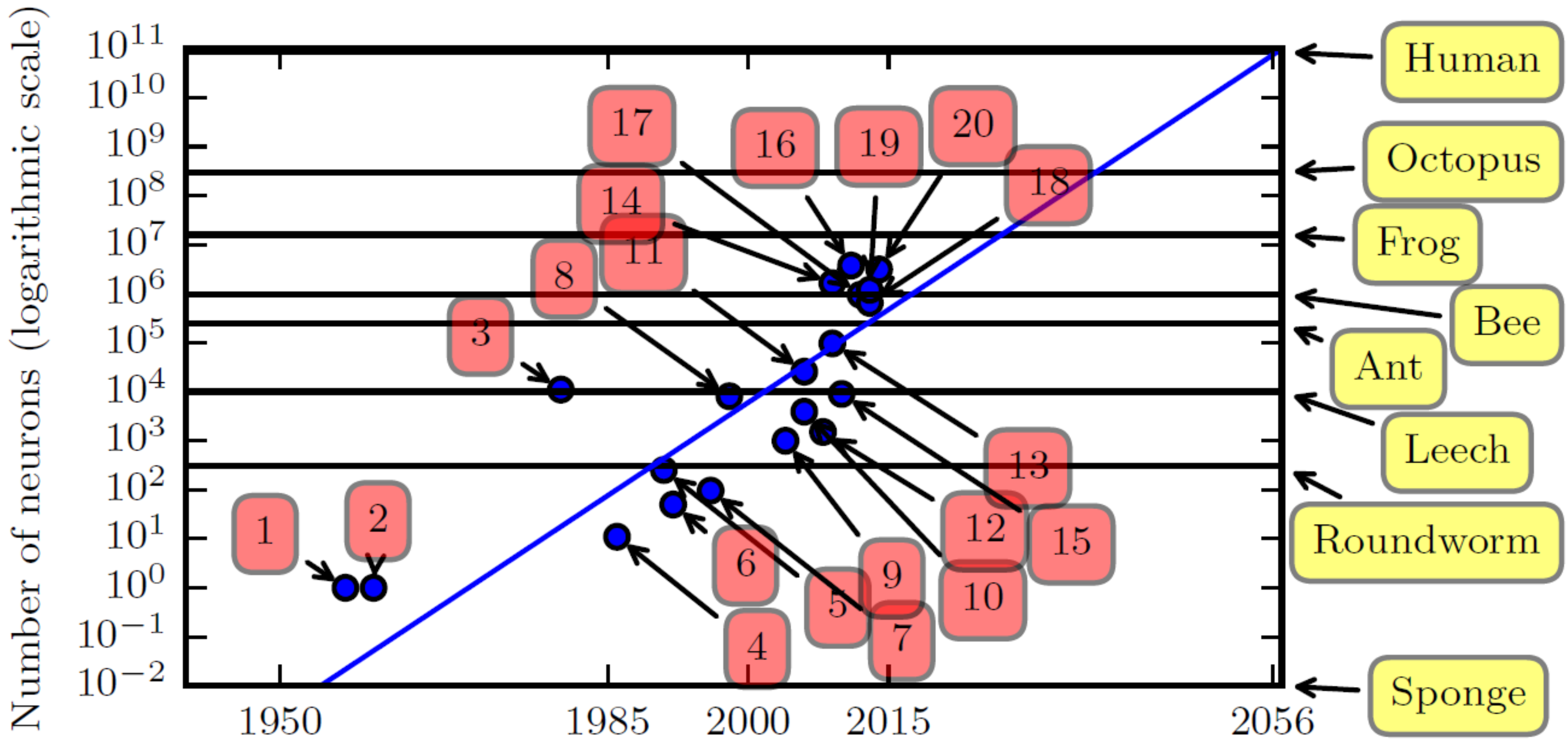
- A main insight of connectionism: animals become intelligent when many of their neurons work together
- The # of connections per neuron is continuously increasing
  - But, still smaller than that of human



Number of Neurons

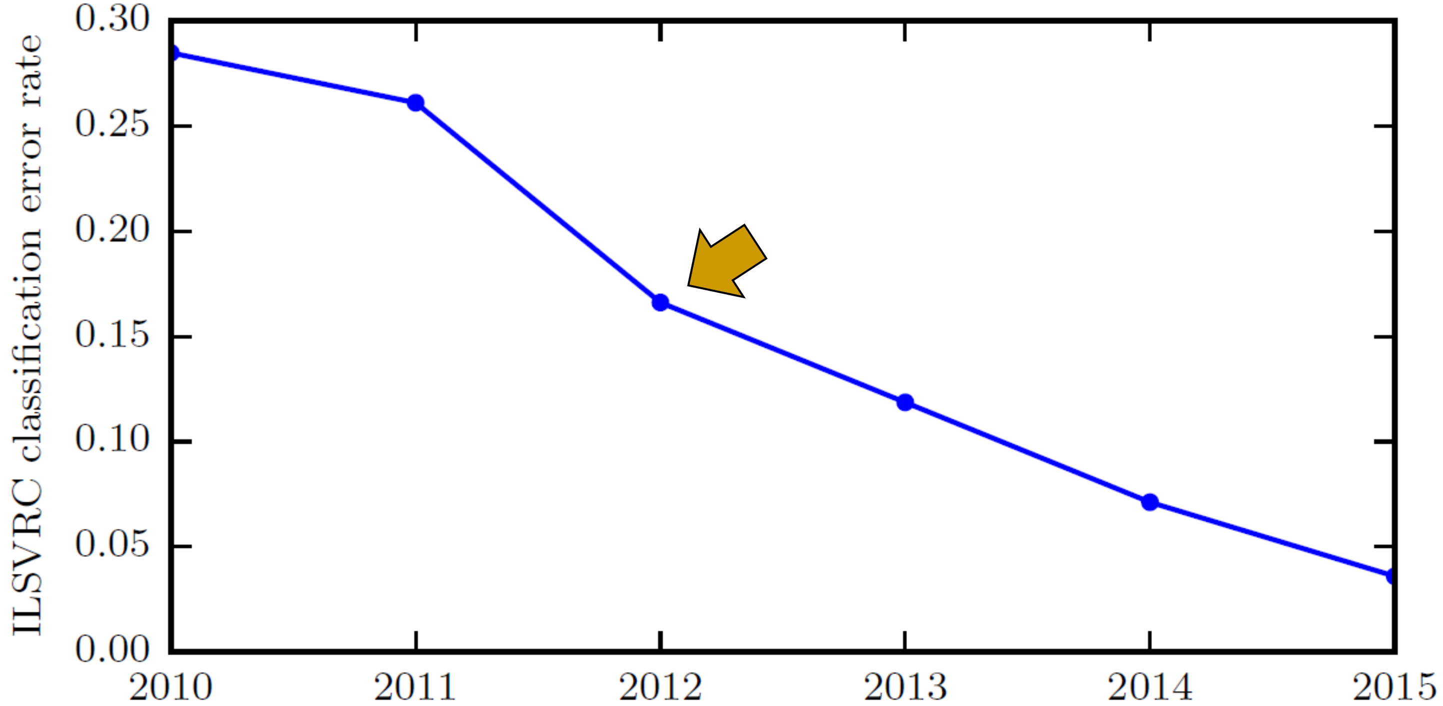
- The total # of neurons of neural networks has been very small until recently
- Since the introduction of hidden units, artificial neural networks (ANN) have doubled in size roughly every 2.4 years
- Unless new technologies allow faster scaling, ANN will reach the same number of neurons as the human brain in 2050
- The increase in model size is one of the most important trends in deep learning
  - Due to faster CPU, GPU, faster network connectivity, and better software infrastructure for distributed computing

Number of Neurons



Increasing Accuracy, Complexity, and Real-World Impact

- Increasing accuracy: object recognition
  - The deep learning revolution is recognized by many people when a CNN won the ILSVRC challenge by a wide-margin



More on Increasing Accuracy

- Increasing accuracy in other areas
  - Speech recognition
    - Deep learning decreased the error by 50%
  - Image segmentation
  - Machine translation
  - …

Increasing Complexity

- Neural networks become able to solve more complex problems
  - Automatic image transscription
  - Machine translation
  - Neural Turing machine
    - A neural network that learns to read from memory cells and write arbitrary content to memory cells
    - Enables self-programming: learn simple programs from examples of desired behavior
    - E.g., learn to sort list of numbers
  - Playing video games

Real World Impact

- DL used in many top technology companies
  - Google, Microsoft, Facebook, IBM, …
- Many software infrastructure developed
  - Tensorflow, Theano, Caffe, …
- DL has made contributions to other sciences
  - Neuroscience: CNN for object recognition provides a model of visual processing that neuroscientists can study
  - Help develop new medication
  - Automatically parse microscope images used to construct -3D map of the human brain

What you need to know

- Deep learning: an approach to machine learning
  - learning to represent the world as a nested hierarchy of concepts, with each concept defined in relation to simple concepts, and more abstract representations computed in terms of less abstract ones
- Deep learning benefits heavily from advances in human brain research, statistics, math, and computer science
- Recent tremendous growth of deep learning is based on powerful computers, larger datasets, and techniques for training deep networks
- Many opportunities and challenges for applications, theories, and methods

Questions?