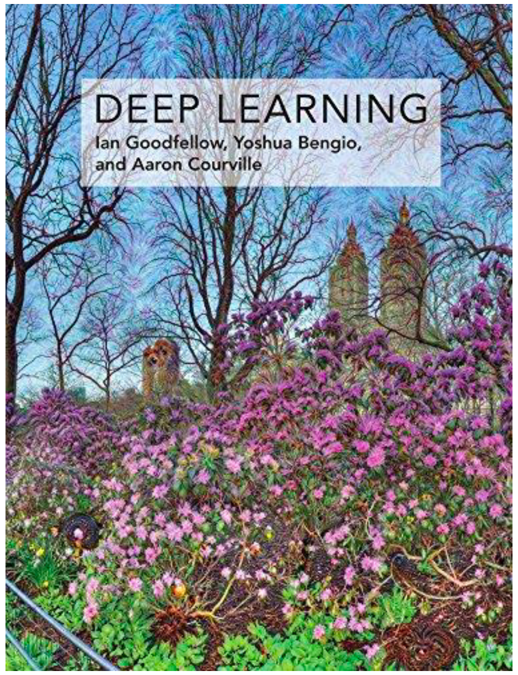
# Introduction

Lecture slides for Chapter 1 of *Deep Learning*
www.deeplearningbook.org
Ian Goodfellow

Adapted by: m.n. for CMPS 392



# Introduction

- Inventors have long dreamed of creating machines that think
- Today, artificial intelligence(AI) is a thriving field with many practical applications and active research topics.
- The field rapidly tackled and solved problems that are intellectually difficult for human beings but relatively straightforward for computers
  - ❑ Mathematical rules
- The true challenge to artificial intelligence proved to be solving the tasks that are easy for people to perform but hard for people to describe formally
  - ❑ Recognizing spoken words or faces in images.

(Goodfellow 2016)

# Deep learning

- Learn from experience
- understand the world in terms of a hierarchy of concepts
  - ❑ each concept defined in terms of its relation to simpler concepts.
- If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers.
  - ❑ For this reason, we call this approach to AI **deep learning.**

(Goodfellow 2016)

# Computer vs. Human

- IBM's Deep Blue chess-playing system defeated world champion Garry Kasparov in 1997
  - ❑ Chess is of course a very simple world!
- A person's everyday life requires an immense amount of knowledge about the world.
  - ❑ Much of this knowledge is subjective and intuitive, and therefore difficult to articulate in a formal way.
- Computers need to capture this same knowledge in order to behave in an intelligent way.



(Goodfellow 2016)

# Previous AI appraches

- **Knowledge base:** A computer can reason about statements in a formal language automatically using logical inference rules.
- Cyc failed to understand a story about a person named Fred shaving in the morning
  - people do not have electrical parts,
  - but Fred was holding an electric razor
  - “FredWhileShaving” contained electrical parts.
  - Cyc asked whether Fred was still a person while he was shaving!

(Goodfellow 2016)

# Machine learning

- Extracting patterns from raw data.
  - A simple machine learning algorithm called **logistic regression** can determine whether to recommend cesarean delivery
  - A simple machine learning algorithm called **naive Bayes** can separate legitimate e-mail from spam e-mail.

- The performance of these simple machine learning algorithms depends heavily on the representation of the data they are given.

(Goodfellow 2016)

# Features

- Each piece of information included in the representation of the patient is known as a feature.
- Logistic regression learns how each of these features of the patient correlates with various outcomes
- the choice of representation has an enormous effect on the performance of machine learning algorithms

(Goodfellow 2016)

# Representations Matter

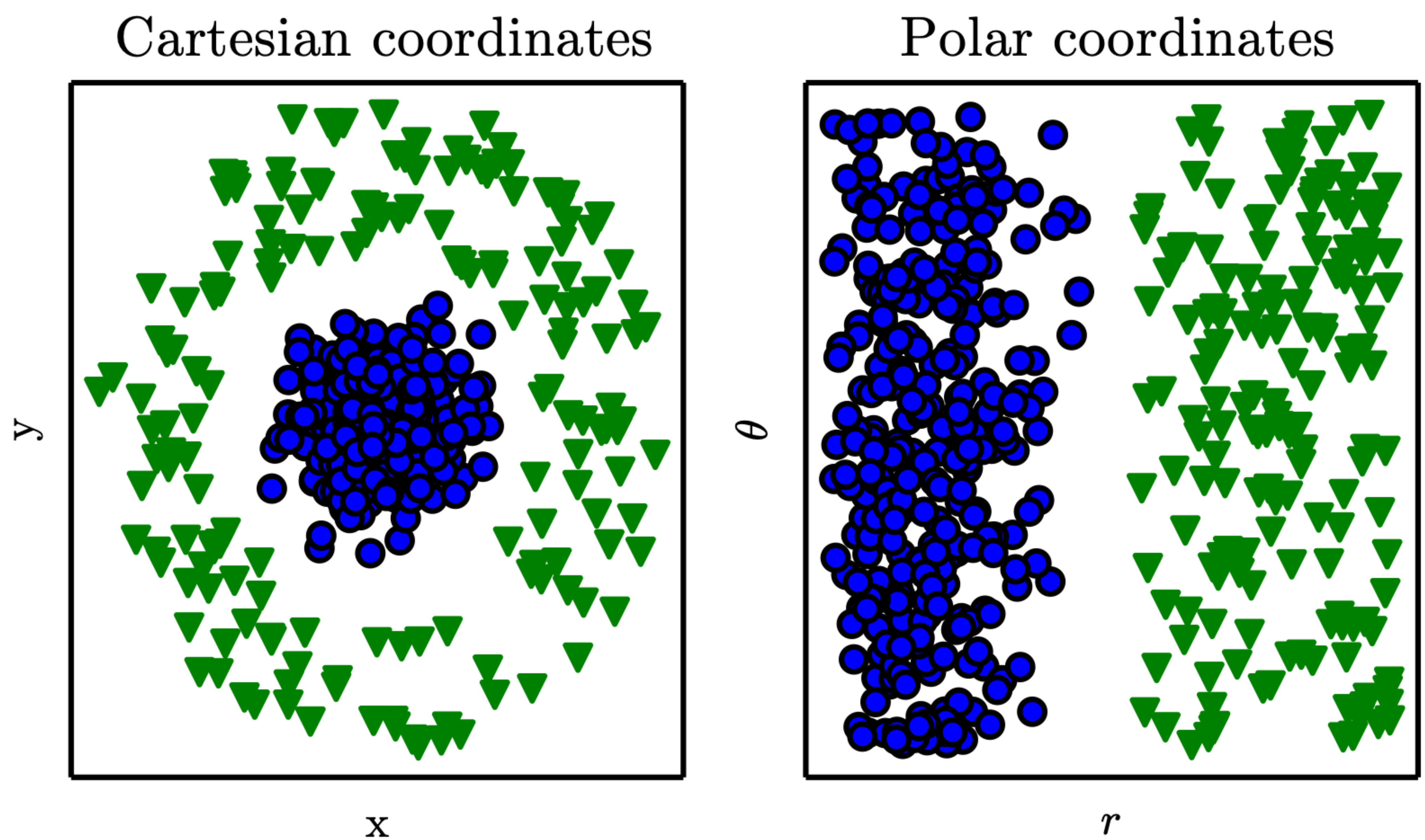



Figure 1.1

(Goodfellow 2016)

# Representation Learning

- However, for many tasks, it is difficult to know what features should be extracted.
  - For example, suppose that we would like to write a program to detect cars in photographs.
  - We know that cars have wheels,
  - But how to describe exactly what a wheel looks like in terms of pixel values?

- We need to discover not only the mapping from representation to output
  - but also the representation itself.

(Goodfellow 2016)

# Autoencoders

- An autoencoder is the combination of an encoder function that converts the input data into a different representation,
- and a decoder function that converts the new representation back into the original format.
- Autoencoders are trained to preserve as much information
- but are also trained to make the new representation have various nice properties.

(Goodfellow 2016)

# Factors of variation

- When analyzing an image of a car, the factors of variation include the position of the car, its color, and the angle and brightness of the sun.
- When analyzing a speech recording, the factors of variation include the speaker’s age, their sex, their accent and the words that they are speaking
- How to *disentangle* the factors of variation and discard the ones that we do not care about?

(Goodfellow 2016)

# Depth: Repeated Composition

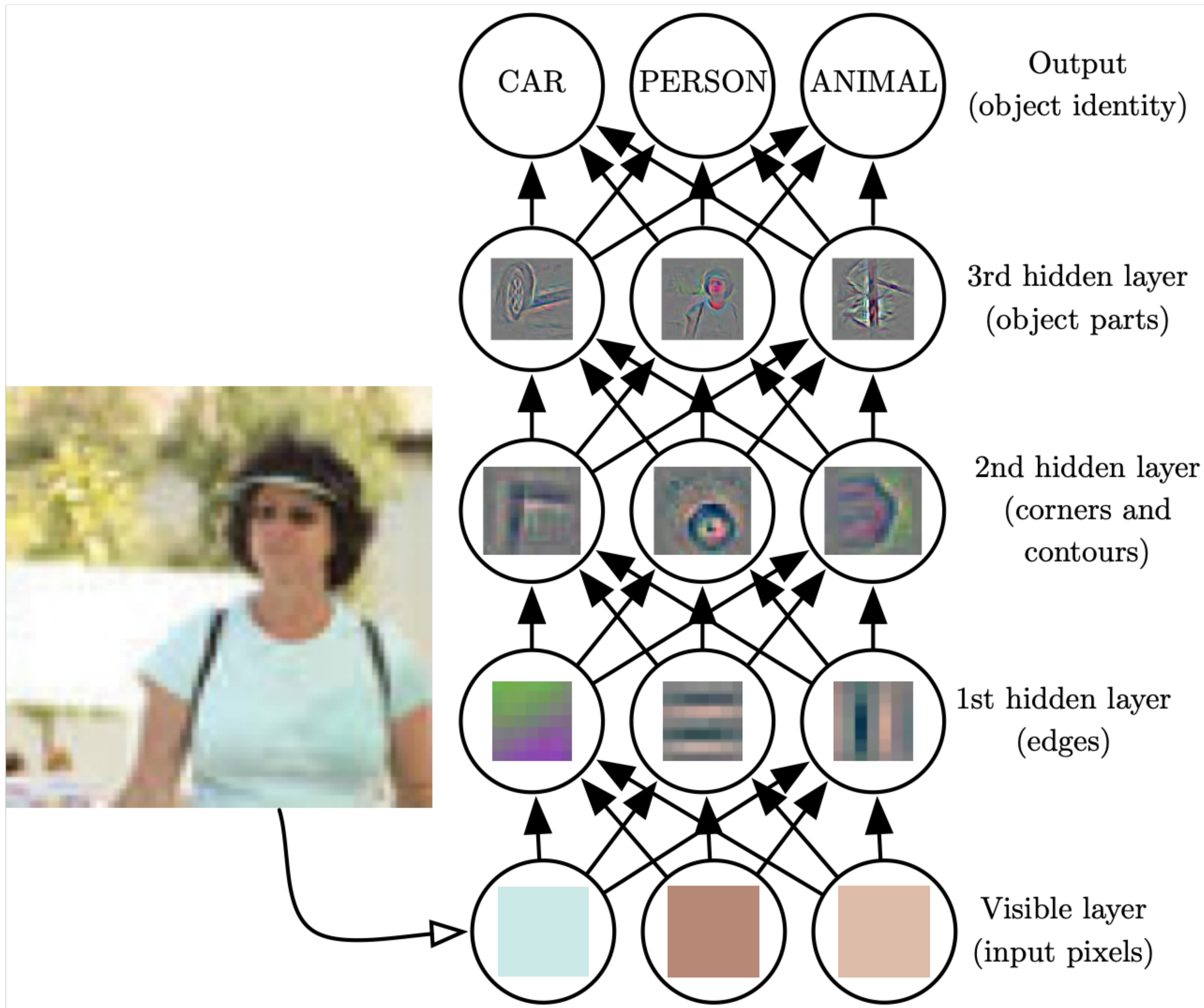



Figure 1.2

(Goodfellow 2016)

# Multilayer perceptron (MLP)

- A multilayer perceptron is just a mathematical function mapping some set of input values to output values.
- The function is formed by composing many simpler functions.
- We can think of each application of a different mathematical function as providing a new representation of the input.

(Goodfellow 2016)

# Multi-step computer program

- Another perspective is that depth allows the computer to learn a multi-step computer program.
- Each layer of the representation can be thought of as the state of the computer's memory after executing another set of instructions in parallel
- Networks with greater depth can execute more instructions in sequence.
- Sequential instructions offer great power because later instructions can refer back to the results of earlier instructions.

(Goodfellow 2016)

# Computational Graphs

Logistic regression: p(y = 1|**x**); $\boldsymbol{\theta}$) =σ($\boldsymbol{\theta}^{\mathsf{T}}$**x**).

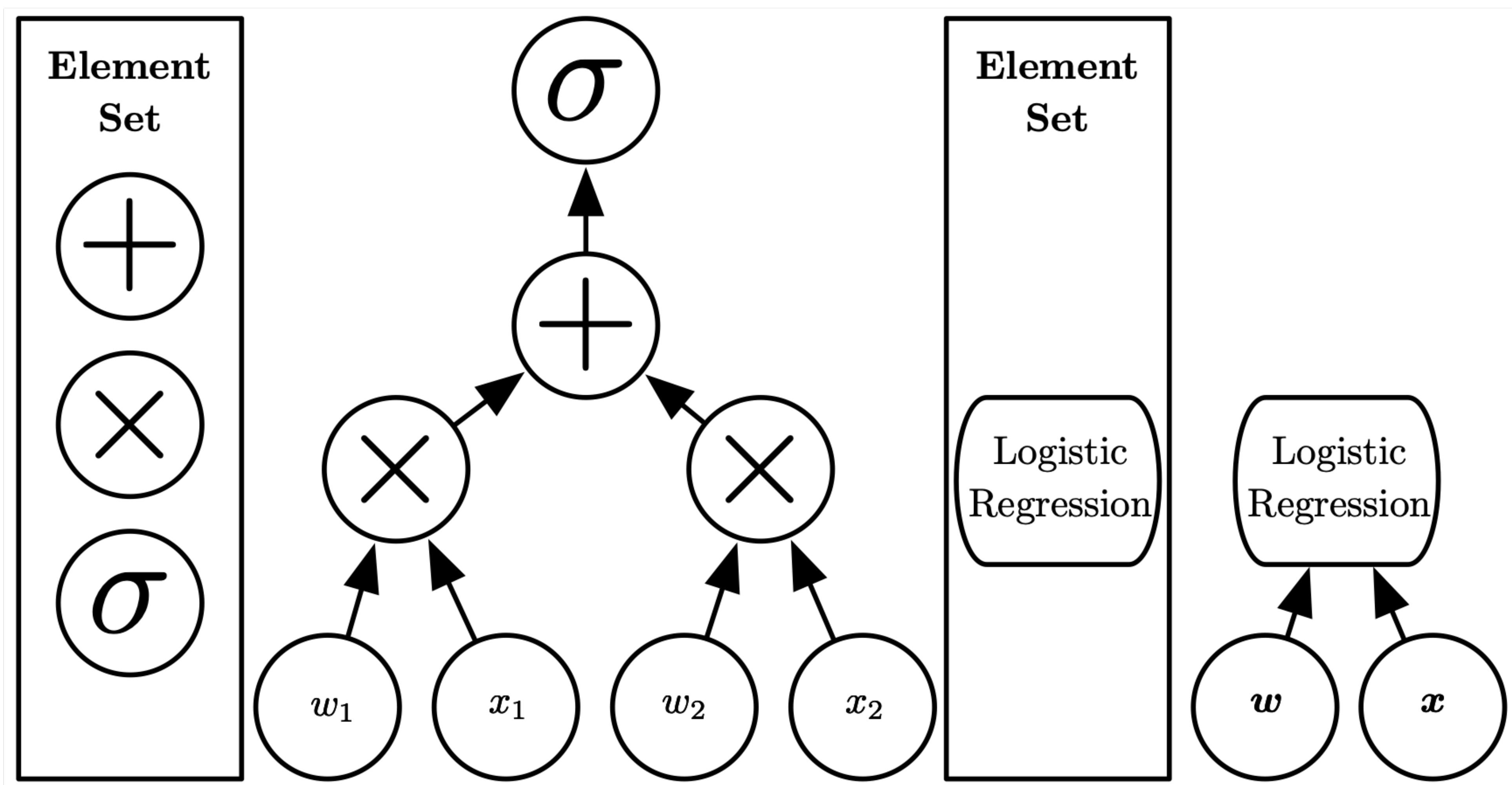



Figure 1.3

(Goodfellow 2016)

# Notion of depth

- Depth is the length of the longest path from input to output but depends on the definition of what constitutes a possible computational step.
  - ❑ If we use addition, multiplication and logistic sigmoids as the elements of our computer language, then this model has depth three.
  - ❑ If we view logistic regression as an element itself, then this model has depth one.

(Goodfellow 2016)

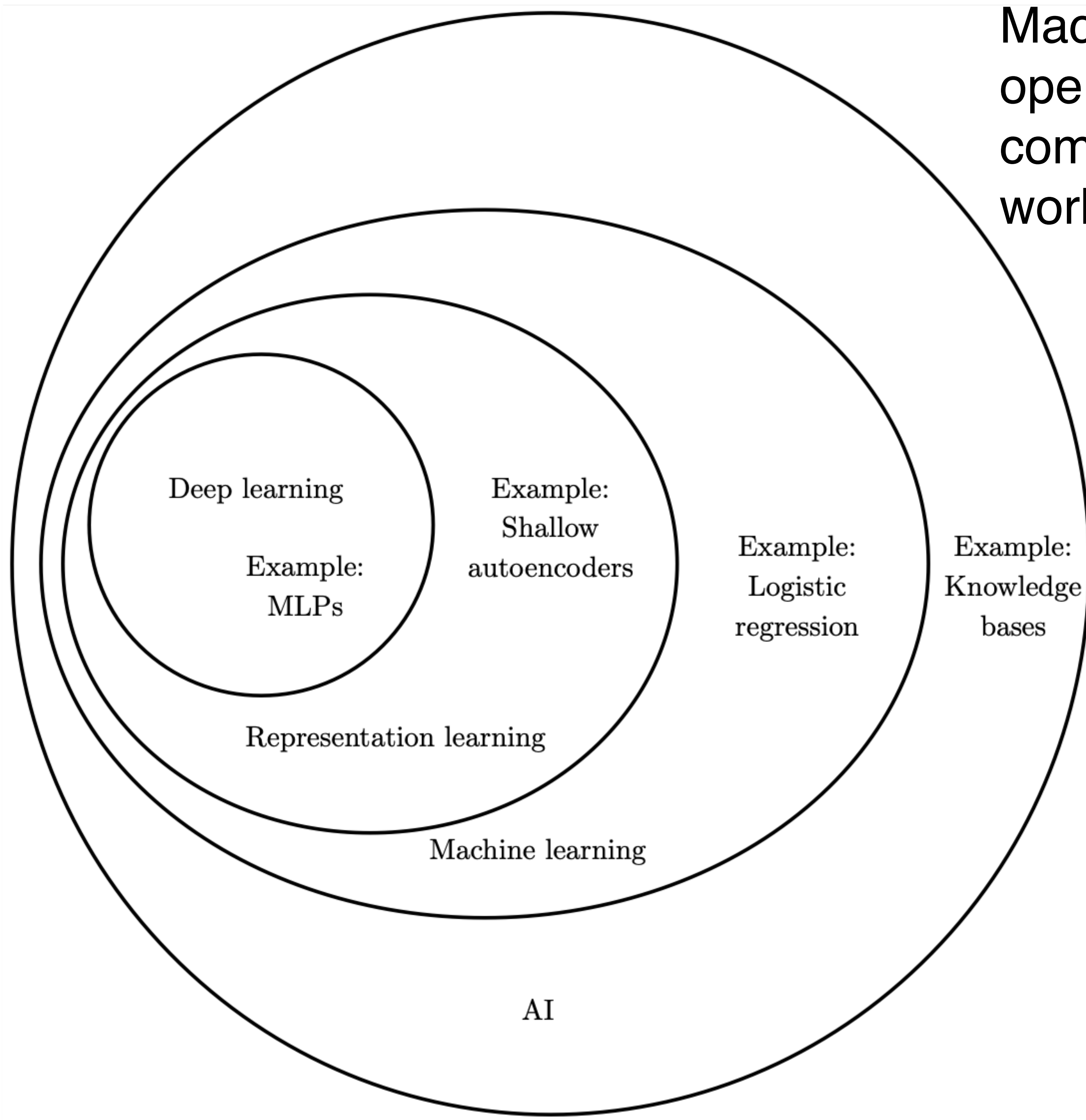
# Deep learning vs. machine learning

- Deep learning is:
  - ❑ An approach to AI
  - ❑ A type of machine learning
  - ❑ a technique that allows computer systems to improve with experience and data
  - ❑ can safely be regarded as the study of models that either involve a greater amount of composition of learned functions or learned concepts than traditional machine learning does.

(Goodfellow 2016)

# Machine Learning and AI

Machine learning can operate in complicated, real-world environments



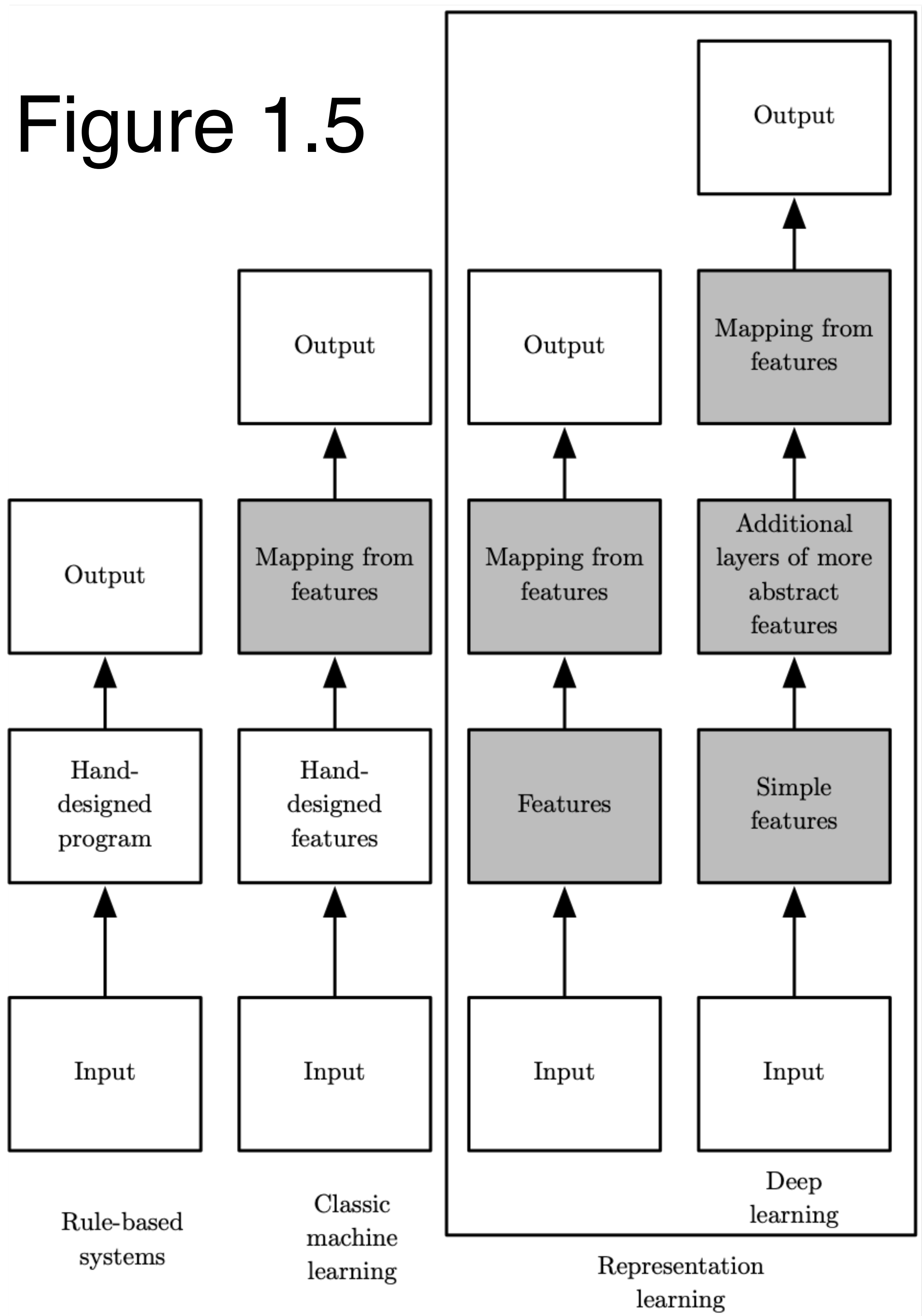

Deep learning is a particular kind of machine learning that achieves great power and flexibility

Figure 1.4

(Goodfellow 2016)

# Learning Multiple Components

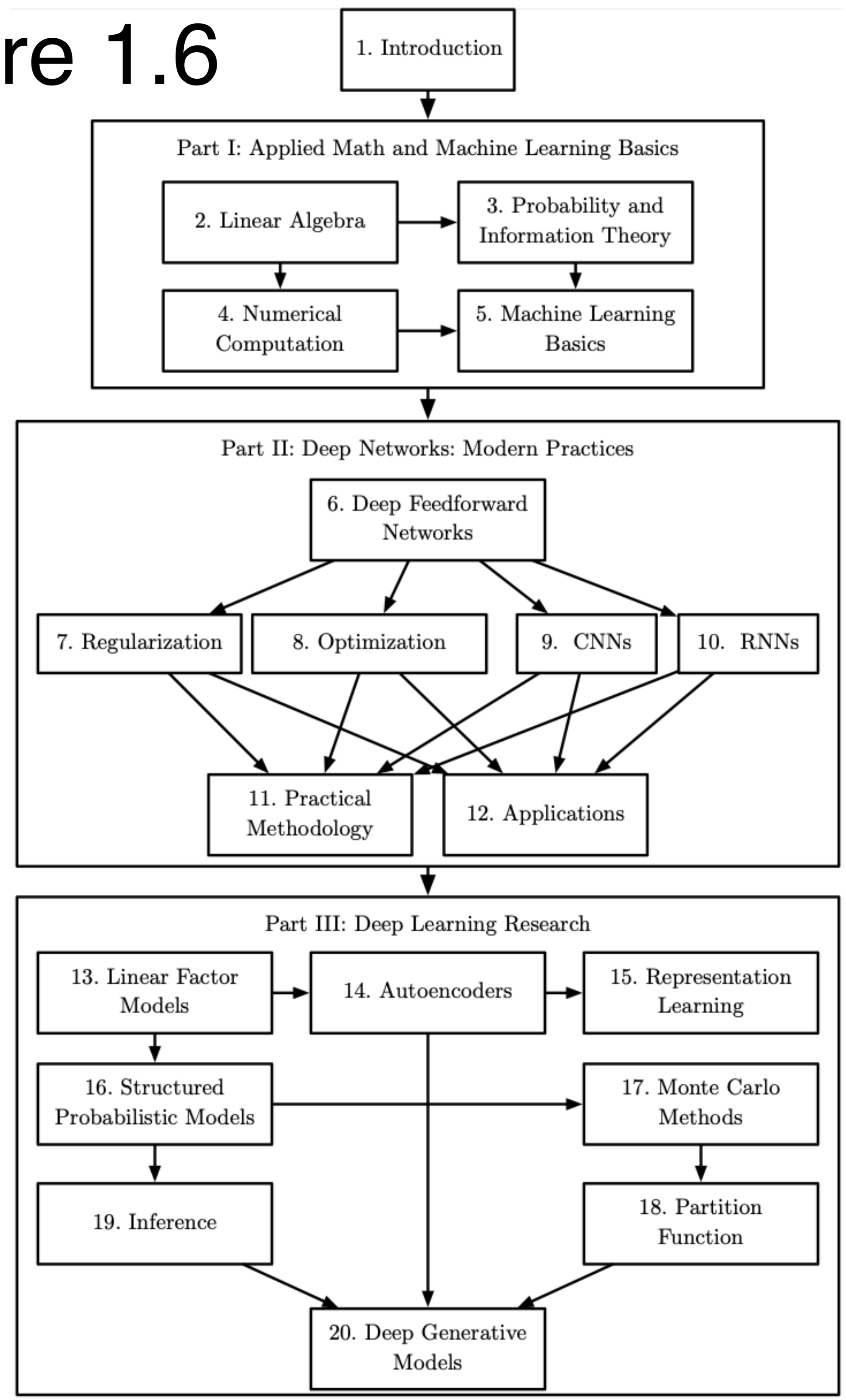
Figure 1.5




(Goodfellow 2016)

# Organization of the Book

Figure 1.6




(Goodfellow 2016)

# Who should take this course?

- University students (undergraduate or graduate)
  - ❑ If you want to begin a career in deep learning and artificial intelligence research
  - ❑ If you want to work as software engineer and want to rapidly acquire machine learning background and begin using deep learning in your product or platform.
- Applications:
  - ❑ computer vision, speech and audio processing, natural language processing, robotics, bioinformatics and chemistry, video games, search engines, online advertising and finance.

(Goodfellow 2016)

# Prerequisities

- We do assume that all readers come from a computer science background.
- We assume familiarity with
  - programming,
  - a basic understanding of computational performance issues, complexity theory,
  - introductory level calculus
  - and some of the terminology of graph theory.

(Goodfellow 2016)

# Deep learning history

- DL has had a long and rich history, but has gone by many names reflecting different philosophical viewpoints, and has waxed and waned in popularity.
- DL has become more useful as the amount of available training data has increased.
- DL models have grown in size over time as computer infrastructure (both hardware and software).
- DL has solved increasingly complicated applications with increasing accuracy over time.

(Goodfellow 2016)

# History

- Three waves of development of deep learning:
  - Cybernetics in the 1940s–1960s
  - Connectionism in the 1980s–1990s
  - Deep learning starting 2006
- Artificial neural networks (ANNs): engineered systems inspired by the biological brain
  - the brain provides a proof by example that intelligent behavior is possible
  - ANNs can help understanding the brain and the principles that underlie human intelligence
- Current deep learning frameworks are not necessarily neurally inspired

(Goodfellow 2016)

# Historical Waves

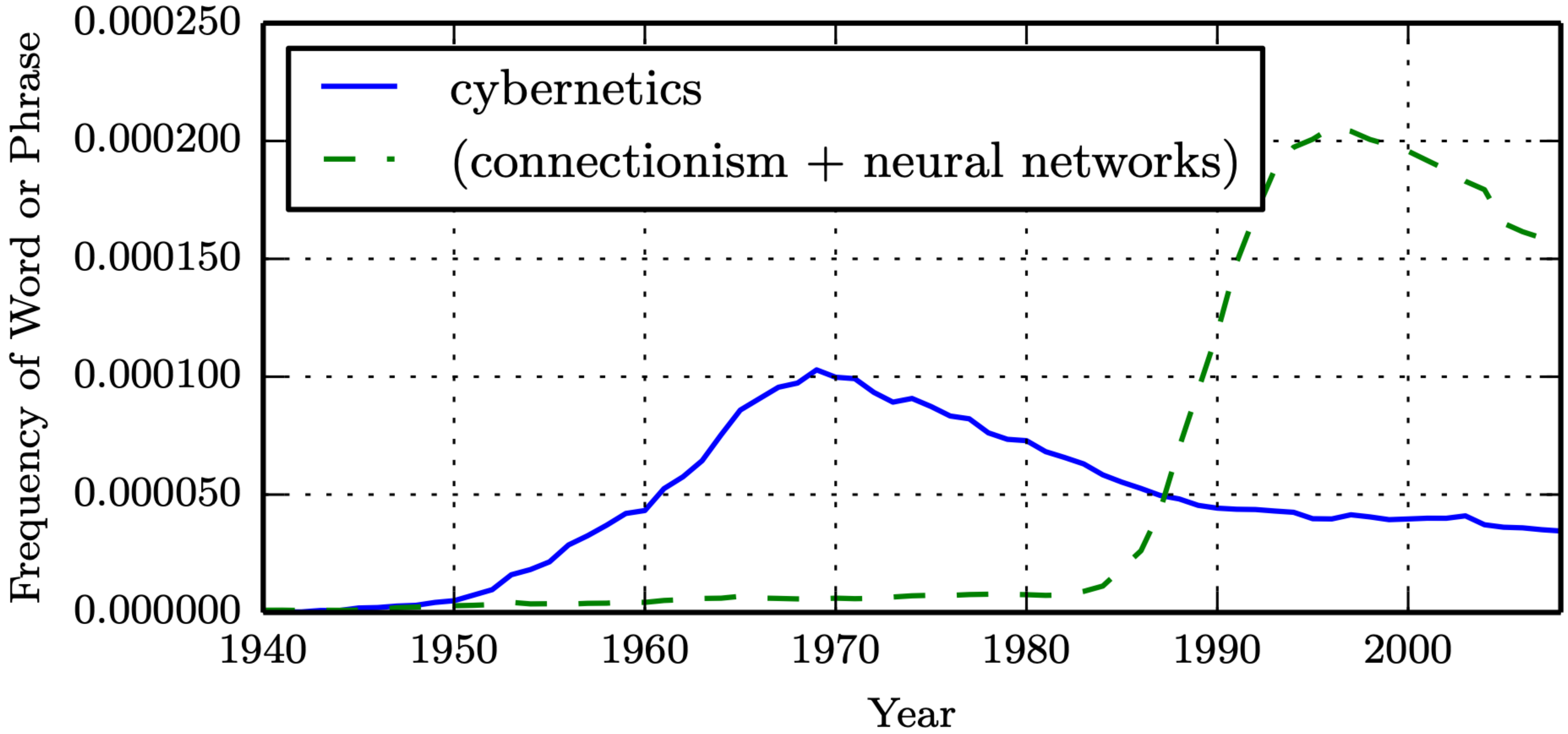



Figure 1.7

(Goodfellow 2016)

# Perceptron *(Rosenblatt, 1958, 1962)*

- These models were designed to take a set of n input values $x_1, \ldots, x_n$ and associate them with an output y.
- These models would learn a set of weights $w_1, \ldots, w_n$ and compute their output

$$f(x, w) = x_1 w_1 + \cdots + x_n w_n$$

Class is sign (f(x,w))

- The adaptive linear element (ADALINE) simply returned the value of f (x) itself to predict a real number *(Widrow and Hoff, 1960)*

(Goodfellow 2016)

# Linear models
## (e.g. Perceptron, Adaline)

- The training algorithm used to adapt the weights of the ADALINE was a special case of an algorithm called **stochastic gradient descent.**

- Linear models have many limitations. Most famously, they cannot learn the XOR function, where $f\ ([0,1],w)\ =\ 1$ and $f([1,0],w)\ =\ 1$ but $f\ ([1,1],w)\ =\ 0$ and $f\ ([0,0],w)\ =\ 0$.

- Critics who observed these flaws in linear models caused a backlash against biologically inspired learning in general (Minsky and Papert, 1969).

(Goodfellow 2016)

# Neuroscience

- Neuroscience has given us a reason to hope that a single deep learning algorithm can solve many different tasks.
- Neuroscientists have found that ferrets can learn to “see” with the auditory processing region of their brain if their brains are rewired to send visual signals to that area (Von Melchner et al., 2000).
- Today, we simply do not have enough information about the brain to use it as a guide.

# Connectionism

- Distributed representation (Hinton et al., 1986)
  - Each input to a system should be represented by many features,
  - and each feature should be involved in the representation of many possible inputs.
  - Example: shape vs. color
- Backpropagation: (Rumelhart et al., 1986; LeCun, 1987).
  - currently the dominant approach to training deep models.

(Goodfellow 2016)

# Second winter

- Ambitious claims while seeking investments.
- other fields of machine learning made advances. Kernel machines (Boser et al., 1992; Cortes and Vapnik, 1995; Schölkopf et al., 1999) and graphical models (Jordan, 1998)
  - These two factors led to a decline in the popularity of neural networks that lasted until 2006-2007.

(Goodfellow 2016)

# Third wave

- Researchers showed that they were able to train deeper neural networks than had been possible before, and focused attention on the theoretical importance of depth
- We have the computational resources to run much larger models today.
- As of 2016, a rough rule of thumb is that a supervised deep learning algorithm will generally achieve acceptable performance with around **5,000 labeled examples** per category, and will match or exceed human performance when trained with a dataset containing at least 10 million labeled examples.

(Goodfellow 2016)

# Historical Trends: Growing Datasets

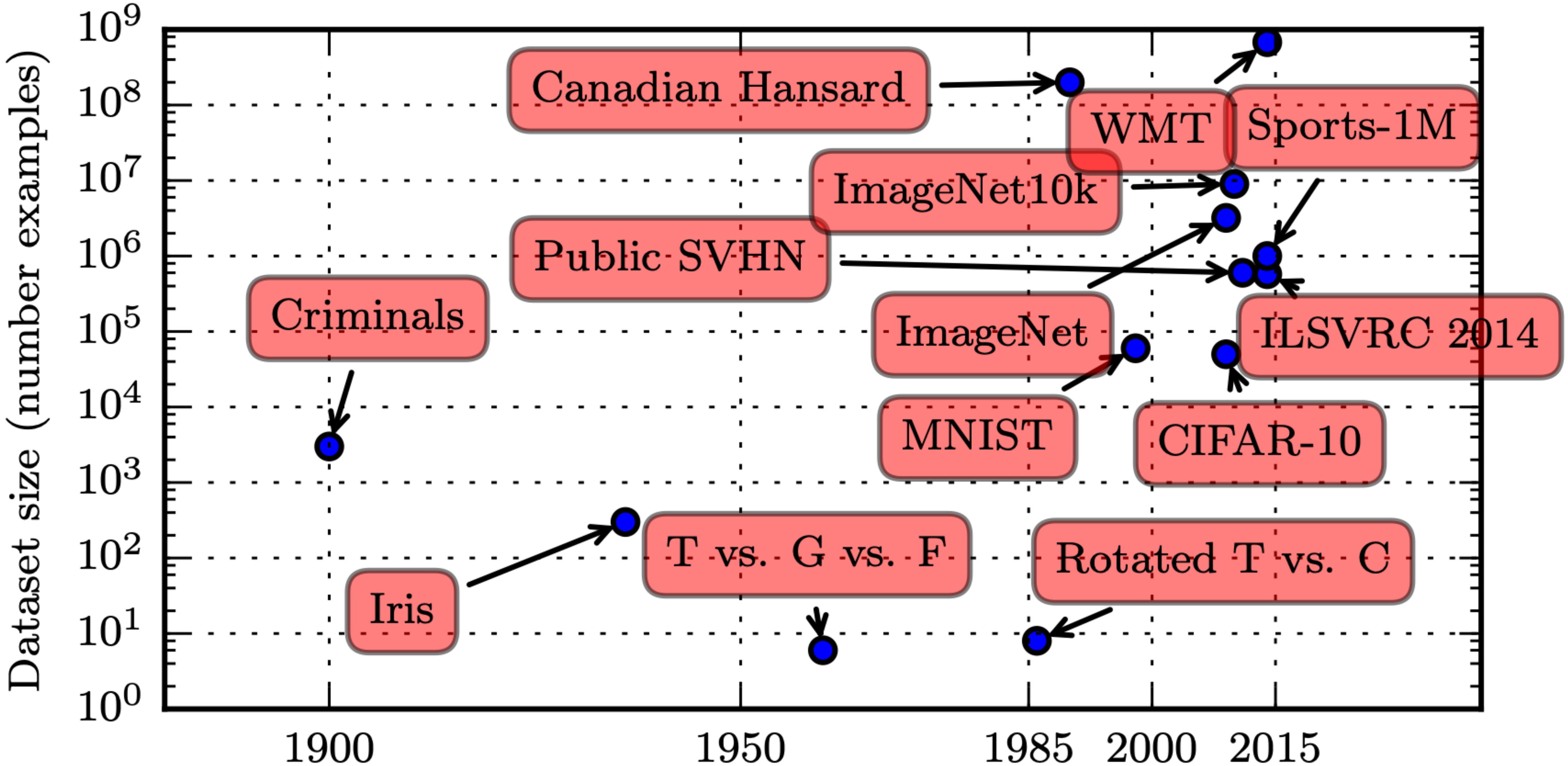



Figure 1.8

(Goodfellow 2016)

# Historical Trends: Increasing model sizes

- faster CPUs,
- the advent of general purpose GPUs,
- faster network connectivity,
- better software infrastructure for distributed computing.

(Goodfellow 2016)

# Connections per Neuron

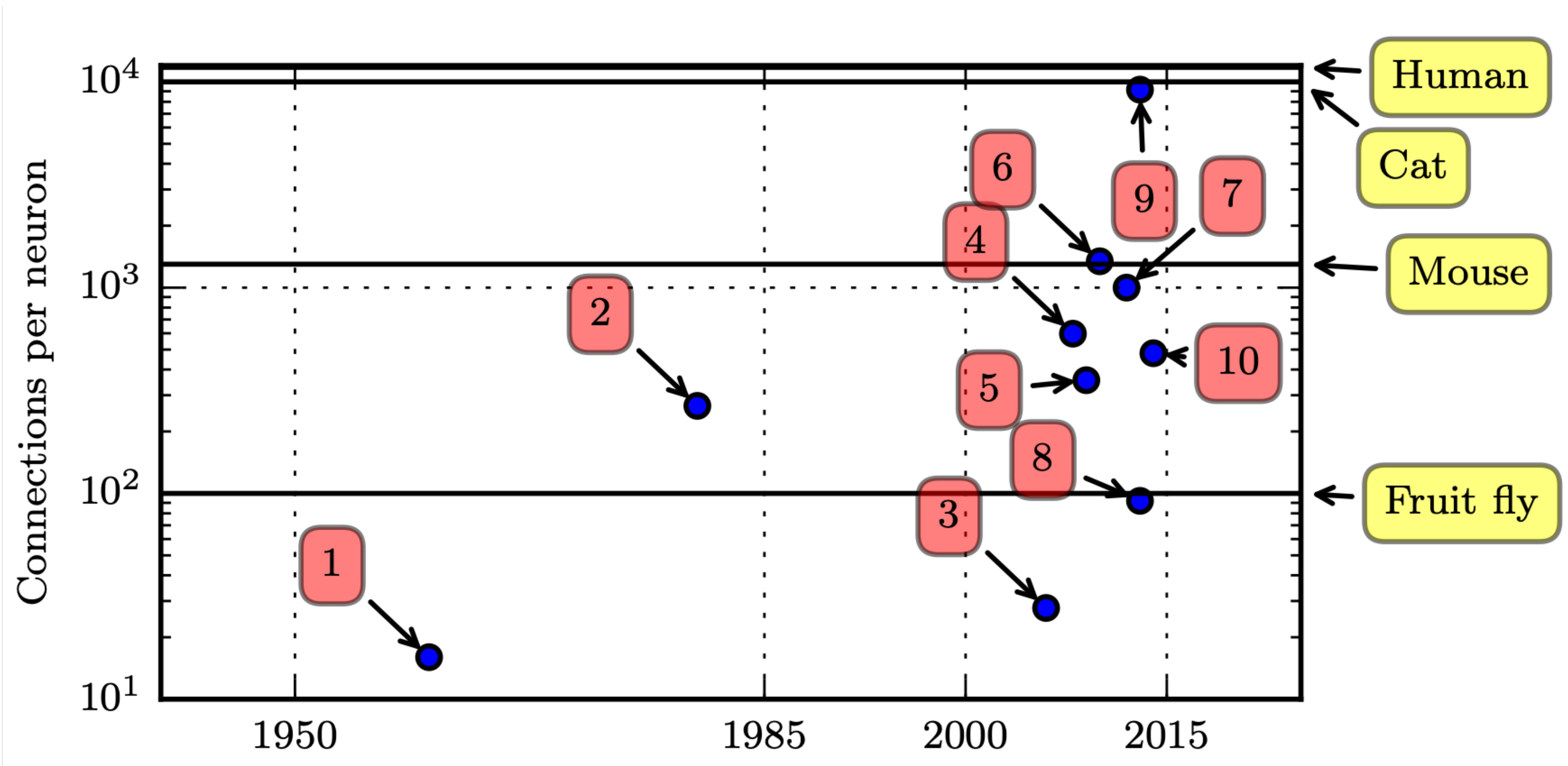



Figure 1.10

(Goodfellow 2016)

# Number of Neurons

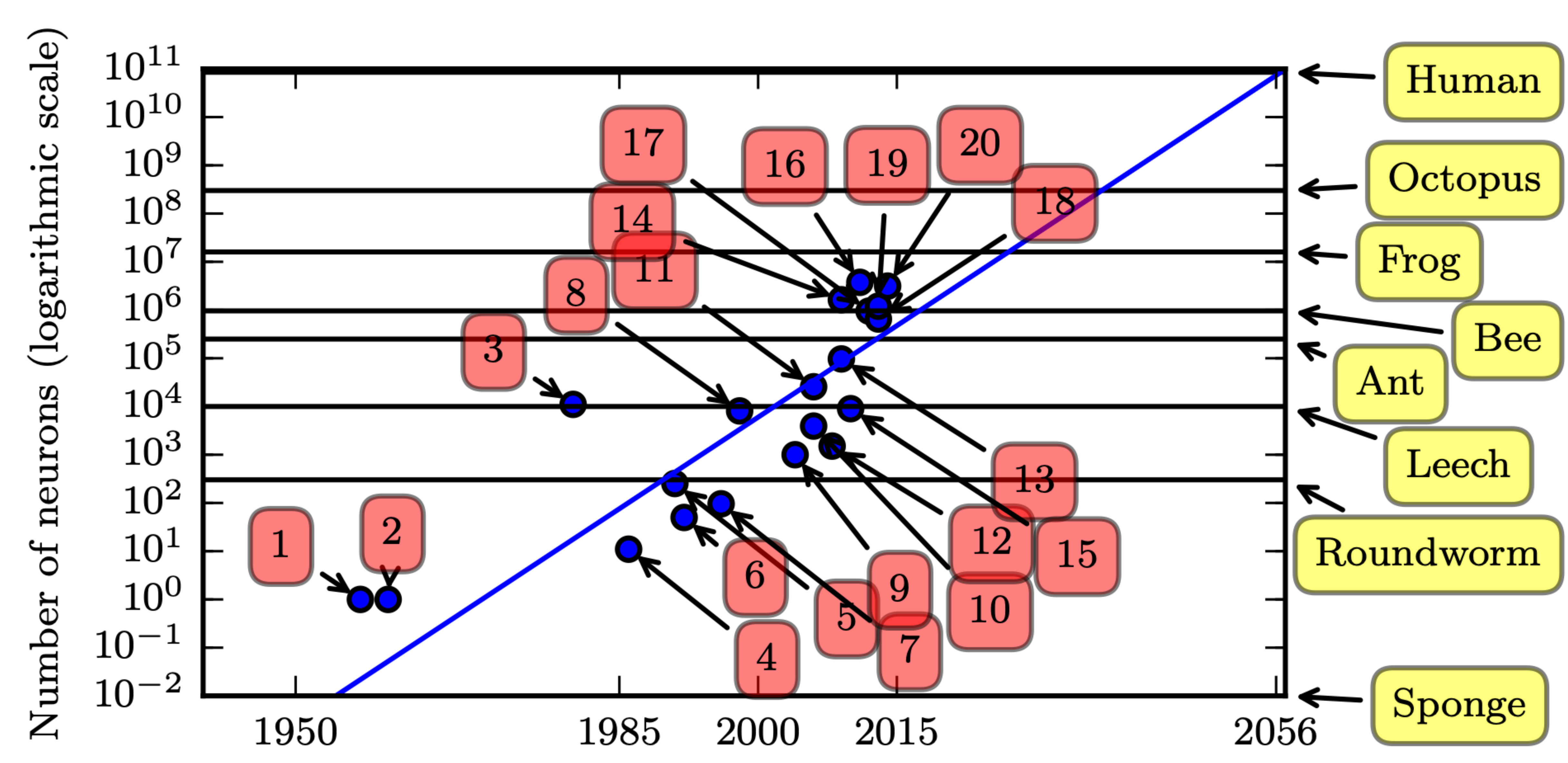



Figure 1.11

(Goodfellow 2016)

# The MNIST Dataset

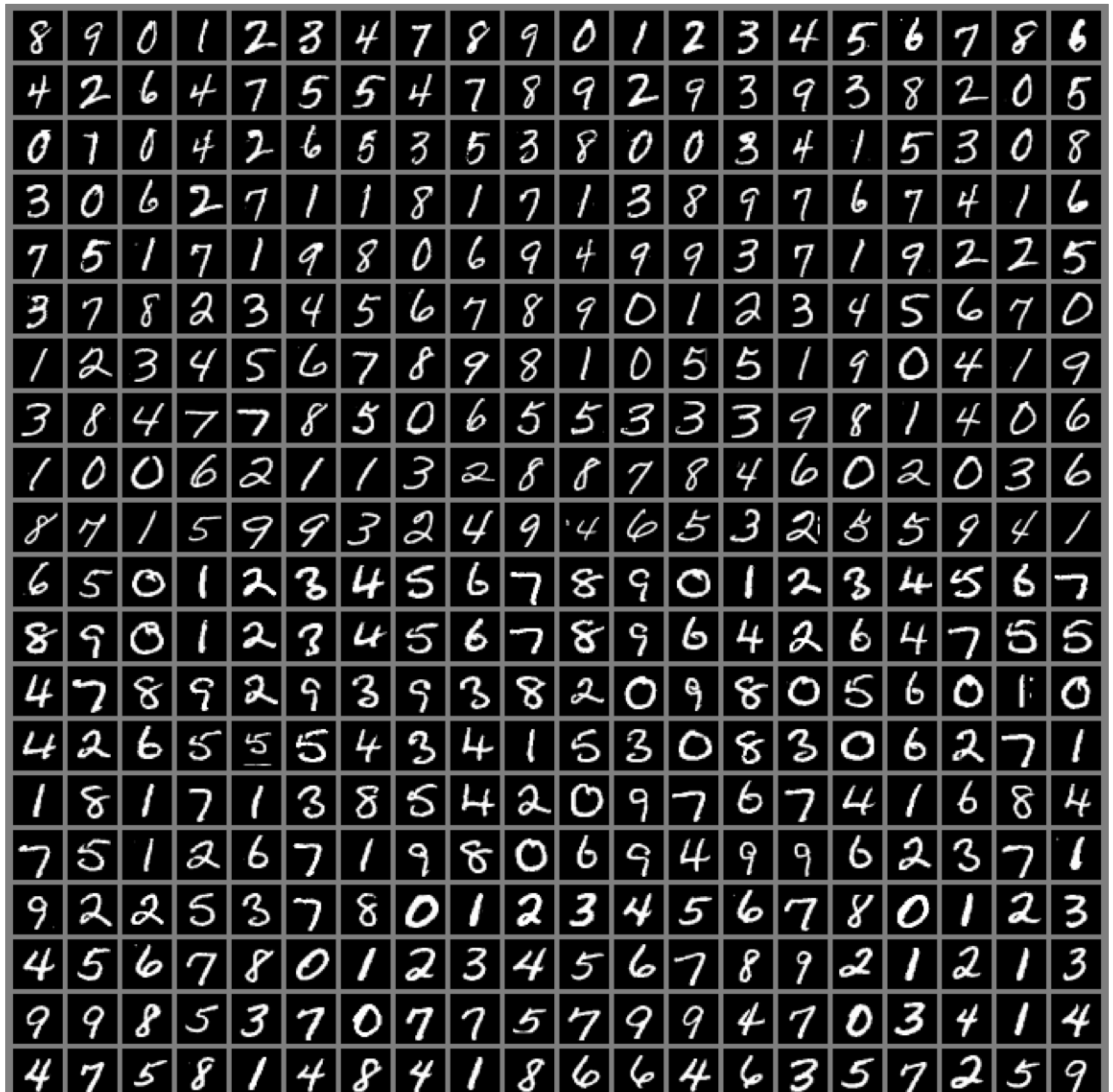


Figure 1.9

the
drosophila
of machine
learning

(Goodfellow 2016)

# Increasing Accuracy, and Real-World Impact

- A dramatic moment in the meteoric rise of deep learning came when a convolutional network won ILSVRC challenge for the first time and by a wide margin, bringing down the state-of-the-art top-5 error rate from 26.1% to 15.3% (Krizhevsky et al., 2012),
  - ❑ Since then, these competitions are consistently won by deep convolutional nets
- The introduction of deep learning to speech recognition resulted in a sudden drop of error rates, with some error rates cut in half.

(Goodfellow 2016)

# Solving Object Recognition

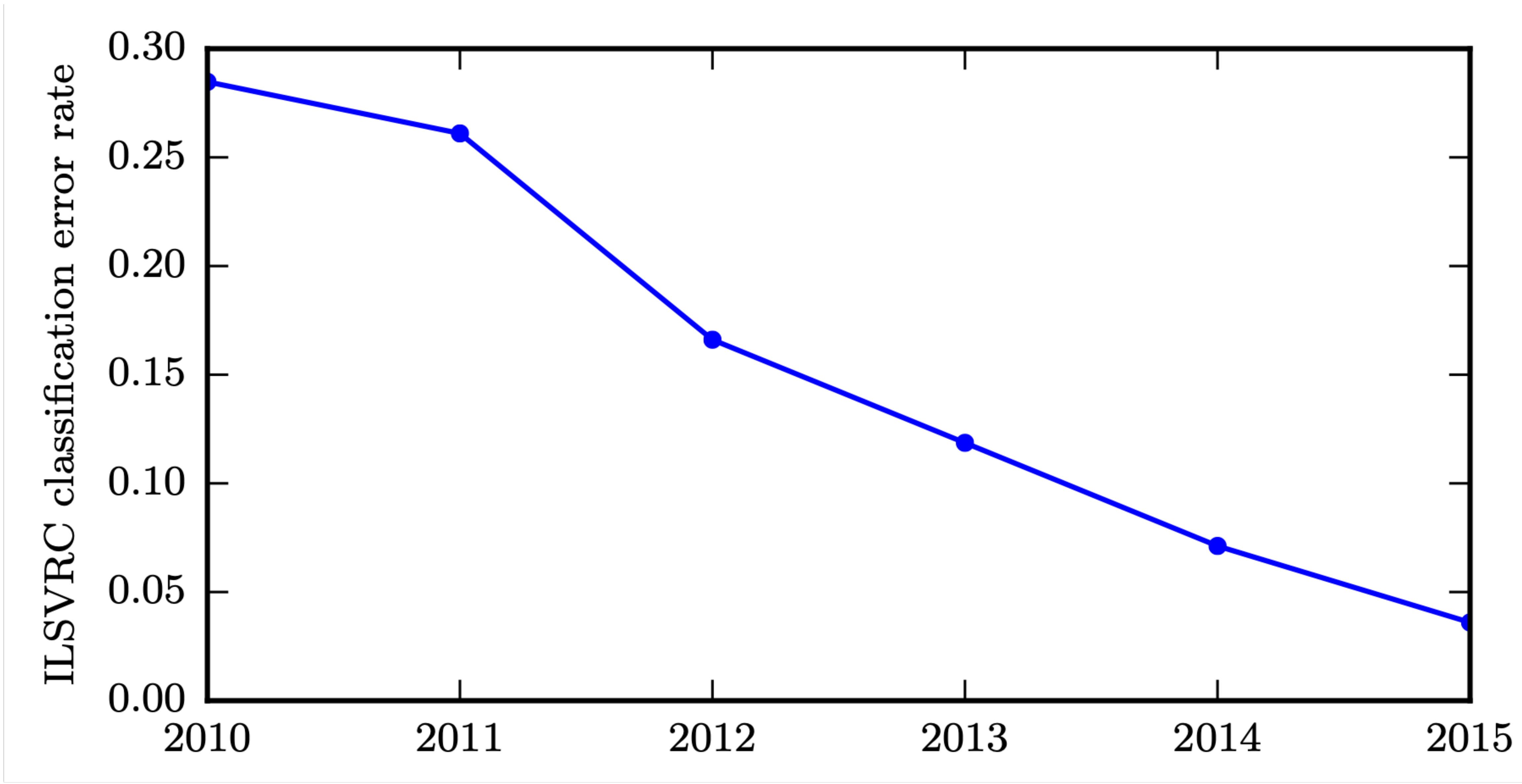



Figure 1.12

(Goodfellow 2016)

# Increasing complexity

- Deep networks have also had spectacular successes for pedestrian detection and image segmentation
  - and yielded superhuman performance in traffic sign classification
- neural networks could learn to output an entire sequence of characters transcribed from an image, rather than just identifying a single object.

(Goodfellow 2016)

# Other applications

- Recurrent neural networks, such as the LSTM sequence model are now used to model relationships between sequences and other sequences rather than just fixed inputs.
- In the context of reinforcement learning, an autonomous agent must learn to perform a task by trial and error, without any guidance from the human operator.
  - ❑ DeepMind demonstrated that a deep reinforcement learning system is capable of learning to play Atari video games, reaching human-level performance
  - ❑ Deep learning has also significantly improved the performance of reinforcement learning for robotics

(Goodfellow 2016)

# Companies and tools

- Google, Microsoft, Facebook, IBM, Baidu, Apple, Adobe, Netflix, NVIDIA and NEC.
- Competition and Convergence of Deep Learning Libraries:
  - ❑ TensorFlow 2.0
  - ❑ PyTorch 1.3

Python 2 support ended on Jan 1, 2020.

**>>> print “Goodbye World”**

(Goodfellow 2016)

# Turing award

- Yann LeCun
- Geoffrey Hinton
- Yoshua Bengio

Turing Award given for:

> "The conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing."

# Online courses

- Fast.ai: Practical Deep Learning for Coders
  - Jeremy Howard et al.
- Stanford CS231n: Convolutional Neural Networks for Visual Recognition
- Stanford CS224n: Natural Language Processing with Deep Learning
- Deeplearning.ai (Coursera): Deep Learning
  - Andrew Ng
- Reinforcement Learning
  - David Silver: Introduction to Reinforcement Learning
  - OpenAI: Spinning Up in Deep RL

# Summary

- Deep learning is an approach to machine learning that has drawn heavily on our knowledge of the human brain, statistics and applied math as it developed over the past several decades.

- In recent years, it has seen tremendous growth in its popularity and usefulness, due in large part to more
  - powerful computers,
  - larger datasets and
  - techniques to train deeper networks.

- The years ahead are full of challenges and opportunities to improve deep learning even further and bring it to new frontiers.

(Goodfellow 2016)

# Watch

- https://www.youtube.com/watch?v=vi7lACKOUao
- https://www.youtube.com/watch?v=0VH1Lim8gL8

(Goodfellow 2016)