Lecture 1

Motivation

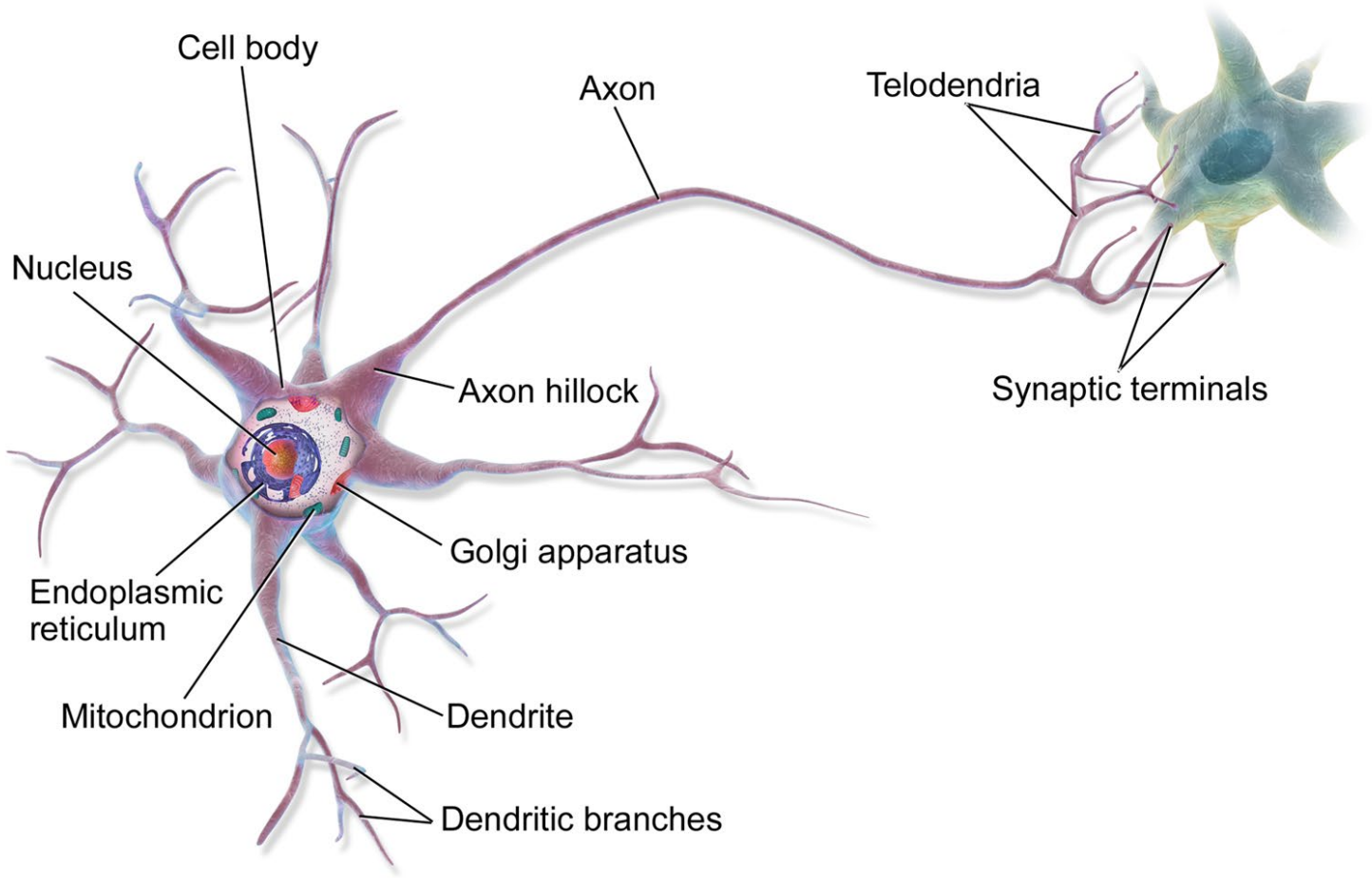
Deep Learning

- Deep learning attempts to learn representations of data with multiple levels of abstraction.

- Deep learning usually refers to a set of algorithms and computational models that are composed of multiple processing layers.

- These methods have significantly improved the state-of-the-art in many domains including, speech recognition, classification, pattern recognition, drug discovery, and genomics.

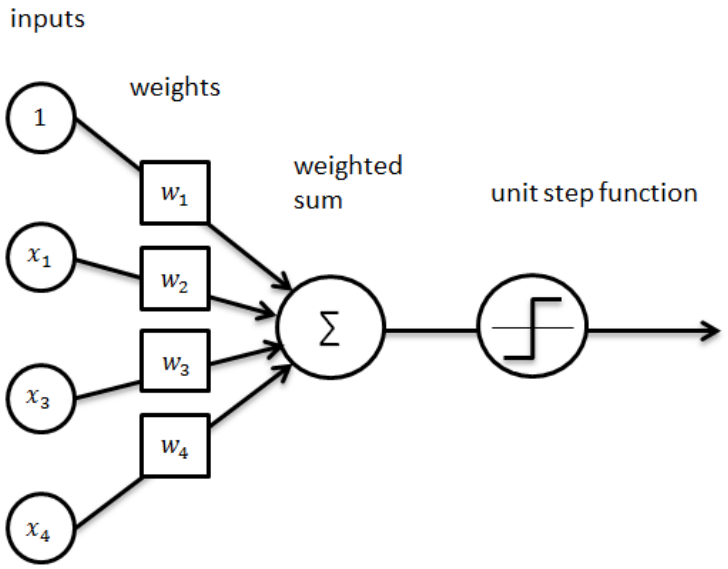
Biological Inspiration



History, Perceptron

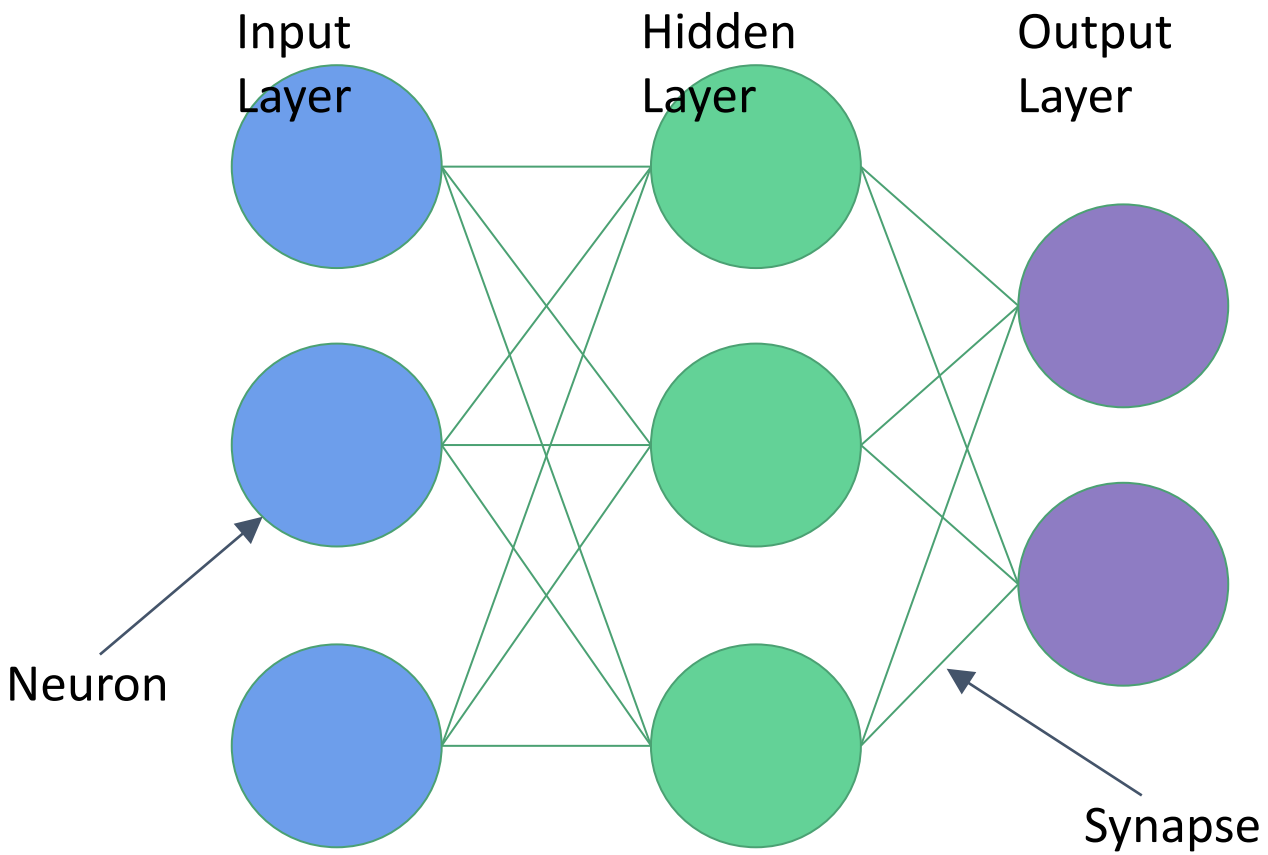
1958

The perceptron was developed by Rosenblatt (physiologist).



Credit:

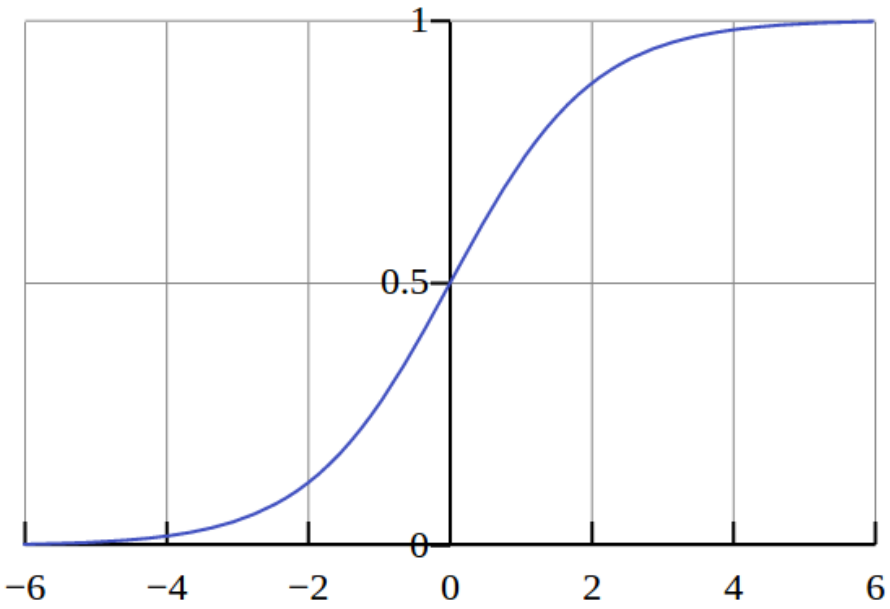
Neural Network Architecture



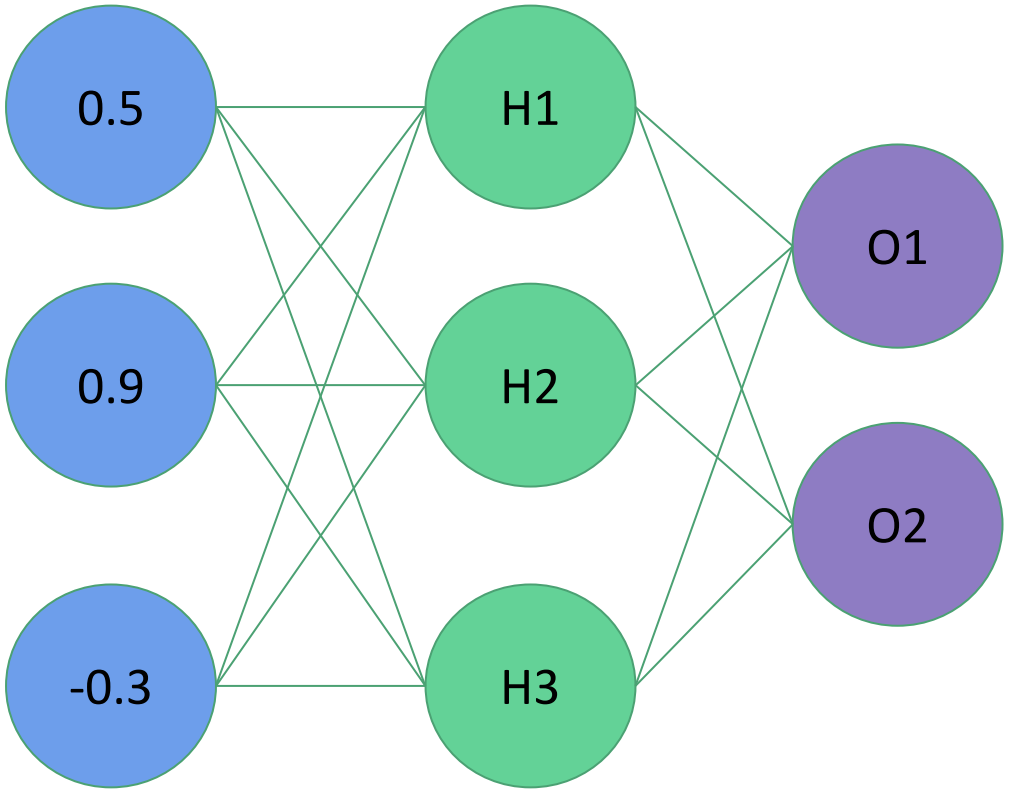
Activation Functions

- Activation Functions are applied to the inputs at each neuron
  - A common activation function is the Sigmoid

$$S(t) = \frac{1}{1 + e^{-t}}$$



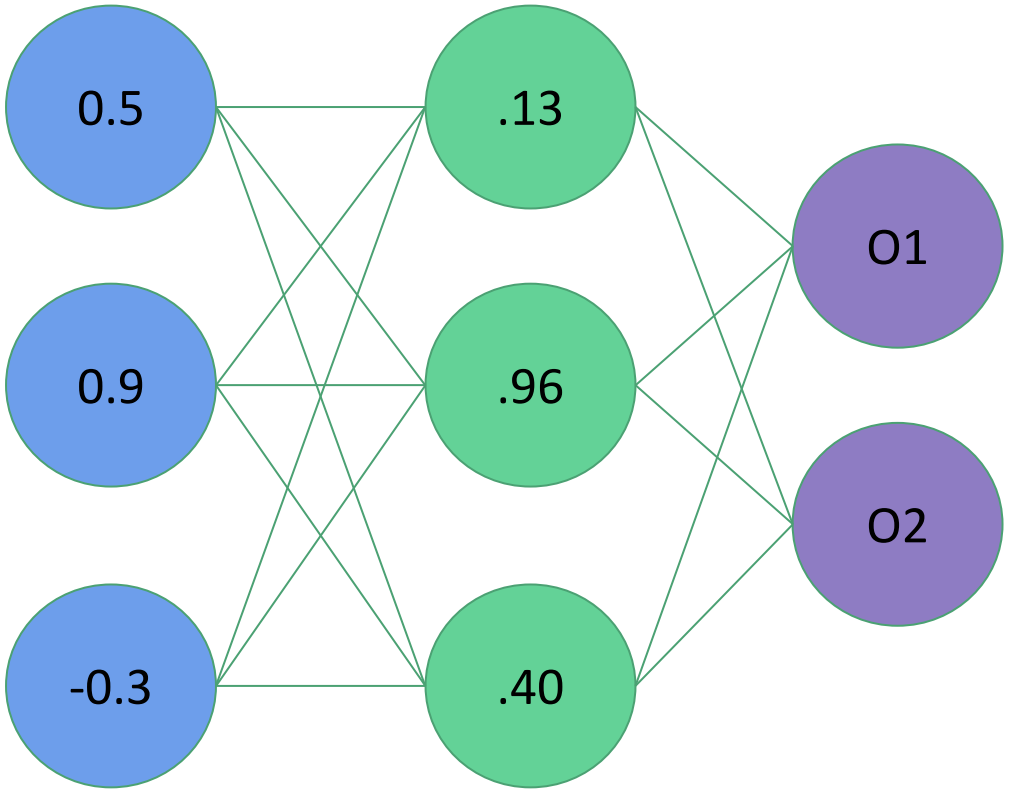
Inference



H1 Weights = (1.0, -2.0, 2.0)
H2 Weights = (2.0, 1.0, -4.0)
H3 Weights = (1.0, -1.0, 0.0)

O1 Weights = (-3.0, 1.0, -3.0)
O2 Weights = (0.0, 1.0, 2.0)

Inference

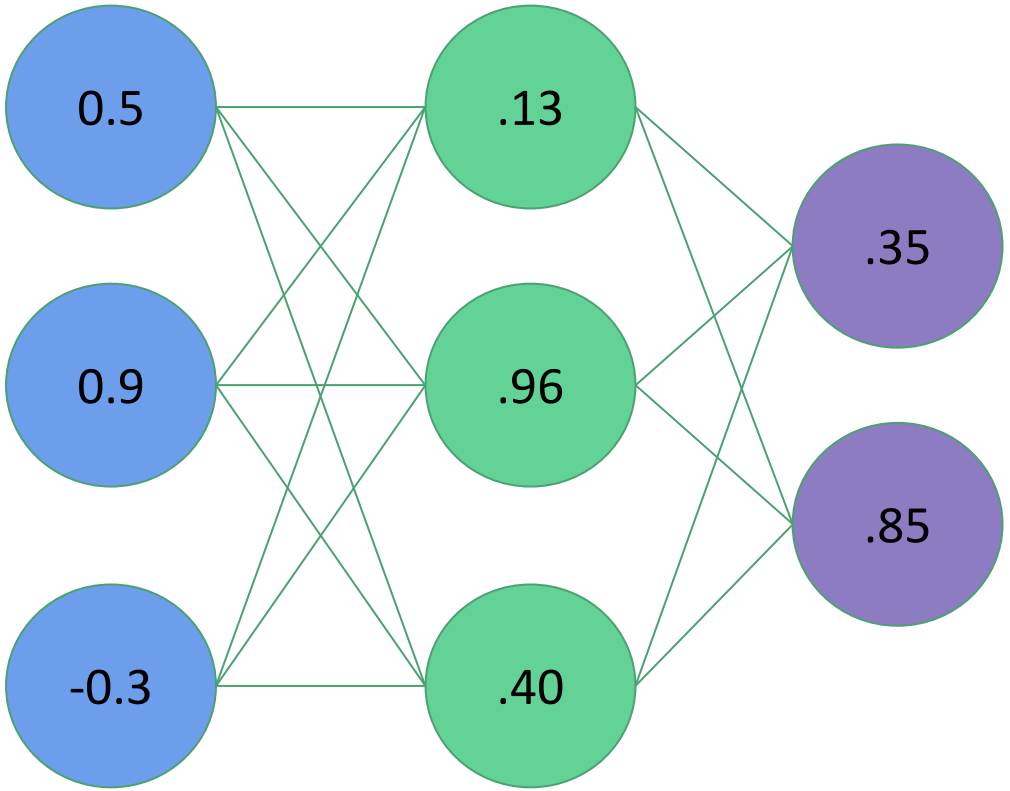


H1 Weights = (1.0, -2.0, 2.0)
H2 Weights = (2.0, 1.0, -4.0)
H3 Weights = (1.0, -1.0, 0.0)

O1 Weights = (-3.0, 1.0, -3.0)
O2 Weights = (0.0, 1.0, 2.0)

H1 = S(0.5 * 1.0 + 0.9 * -2.0 + -0.3 * 2.0) = S(-1.9) = .13
H2 = S(0.5 * 2.0 + 0.9 * 1.0 + -0.3 * -4.0) = S(3.1) = .96
H3 = S(0.5 * 1.0 + 0.9 * -1.0 + -0.3 * 0.0) = S(-0.4) = .40

Inference

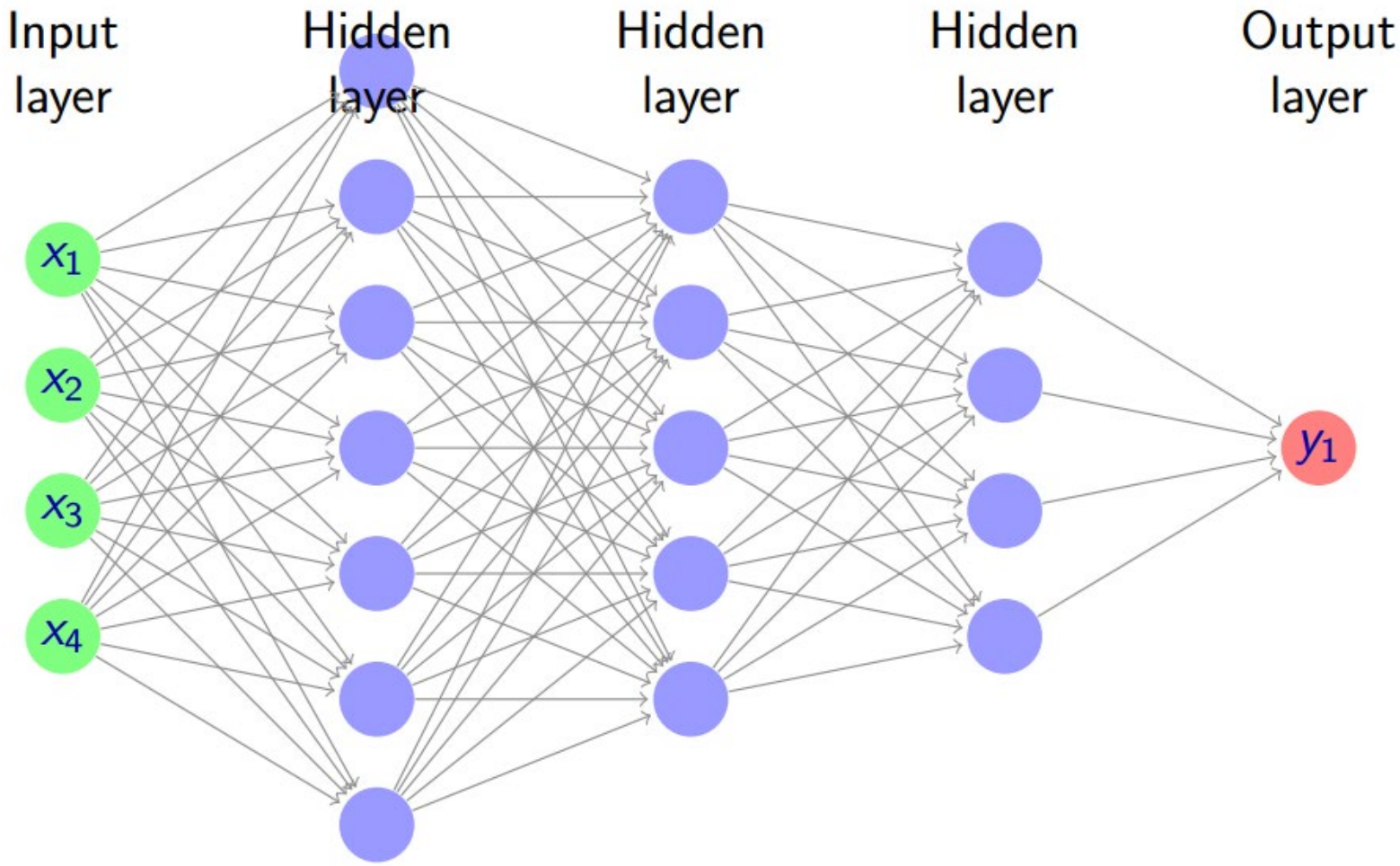


H1 Weights = (1.0, -2.0, 2.0)
H2 Weights = (2.0, 1.0, -4.0)
H3 Weights = (1.0, -1.0, 0.0)

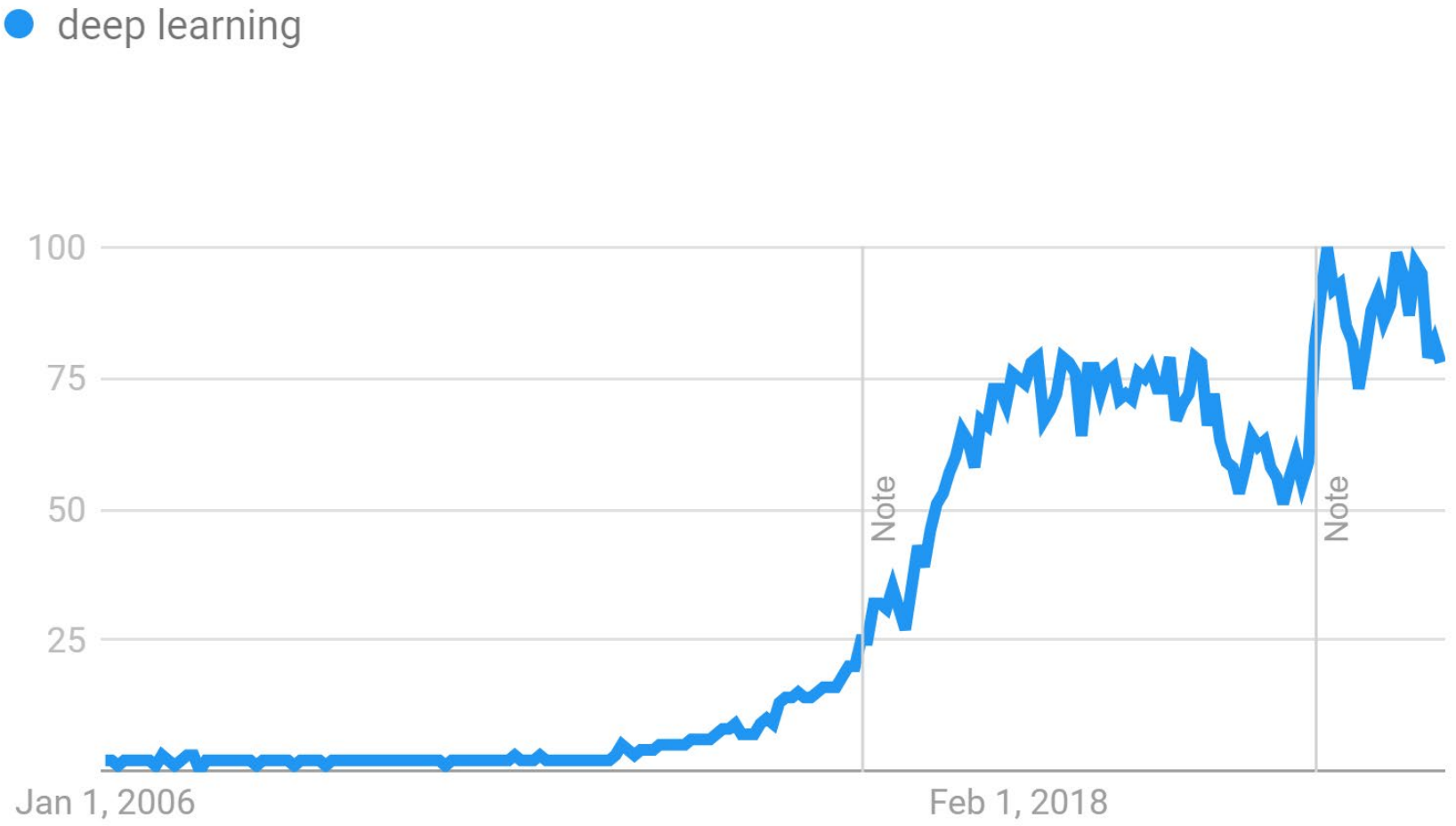
O1 Weights = (-3.0, 1.0, -3.0)
O2 Weights = (0.0, 1.0, 2.0)

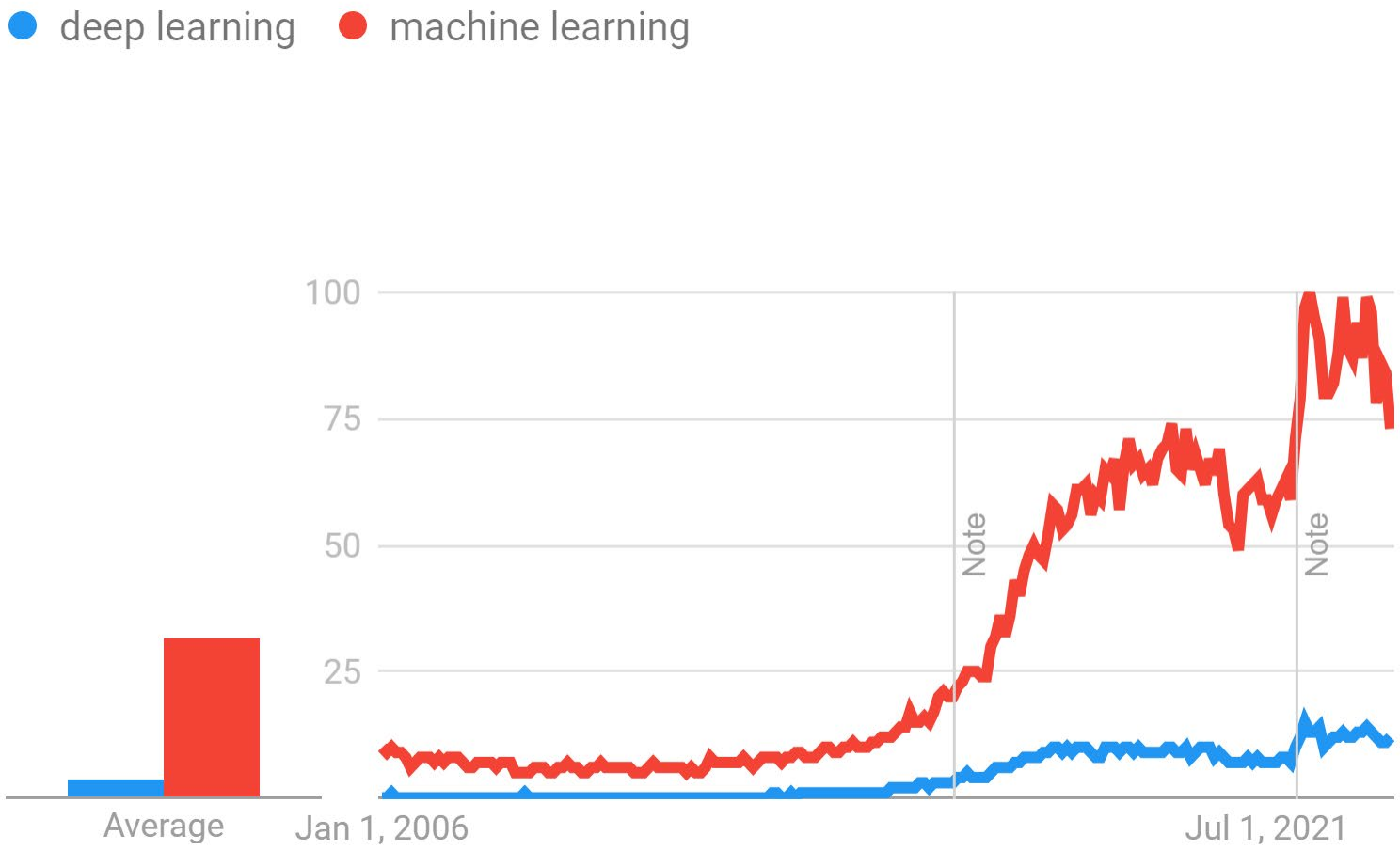
O1 = S(.13 * -3.0 + .96 * 1.0 + .40 * -3.0) = S(-.63) = .35
O1 = S(.13 * 0.0 + .96 * 1.0 + .40 * 2.0) = S(1.76) = .85

Feedforward Deep Networks



Interest over time



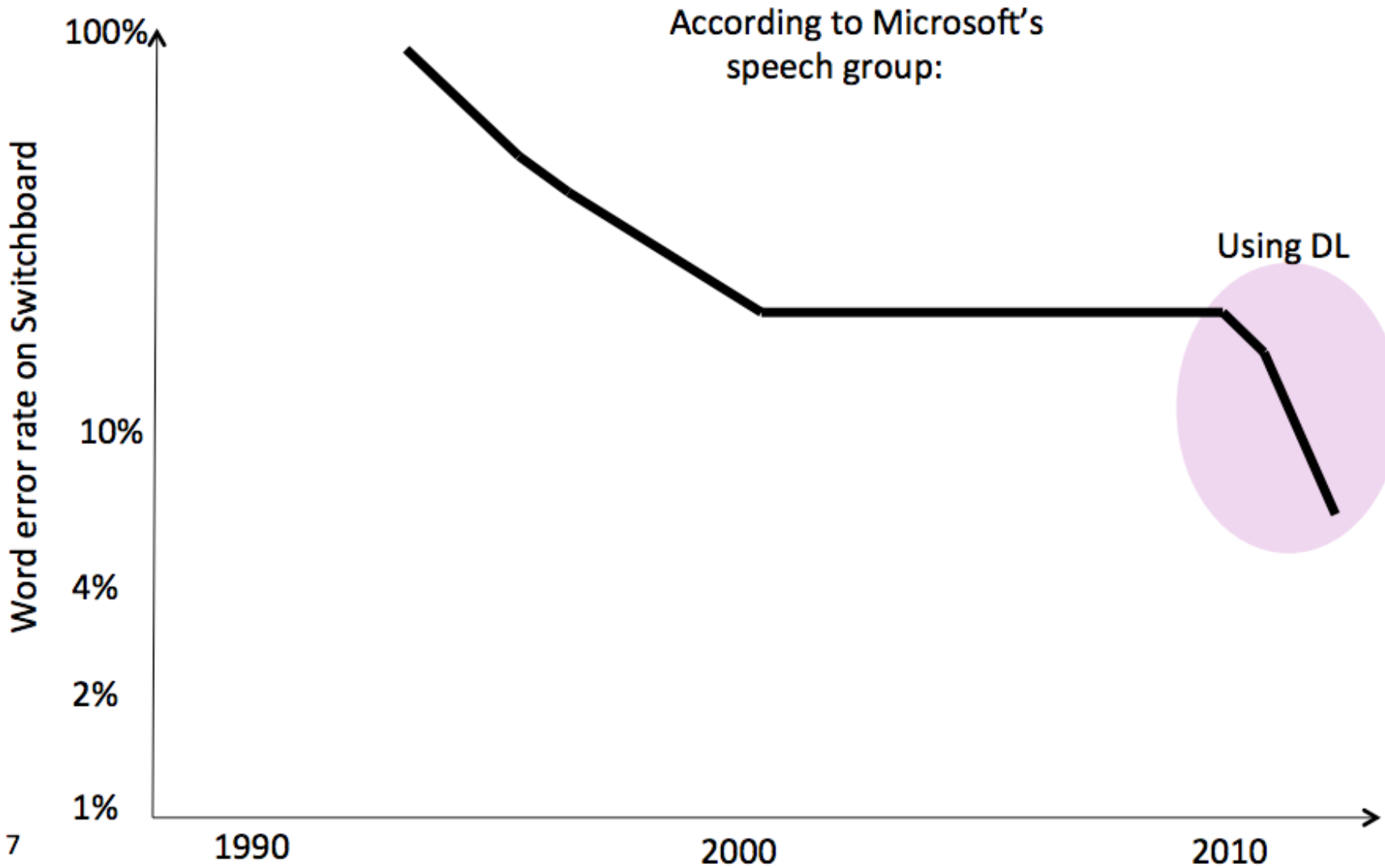

deep learning
machine learning
100
75
50
25
Note
Note
Average
Jan 1, 2006
Jul 1, 2021

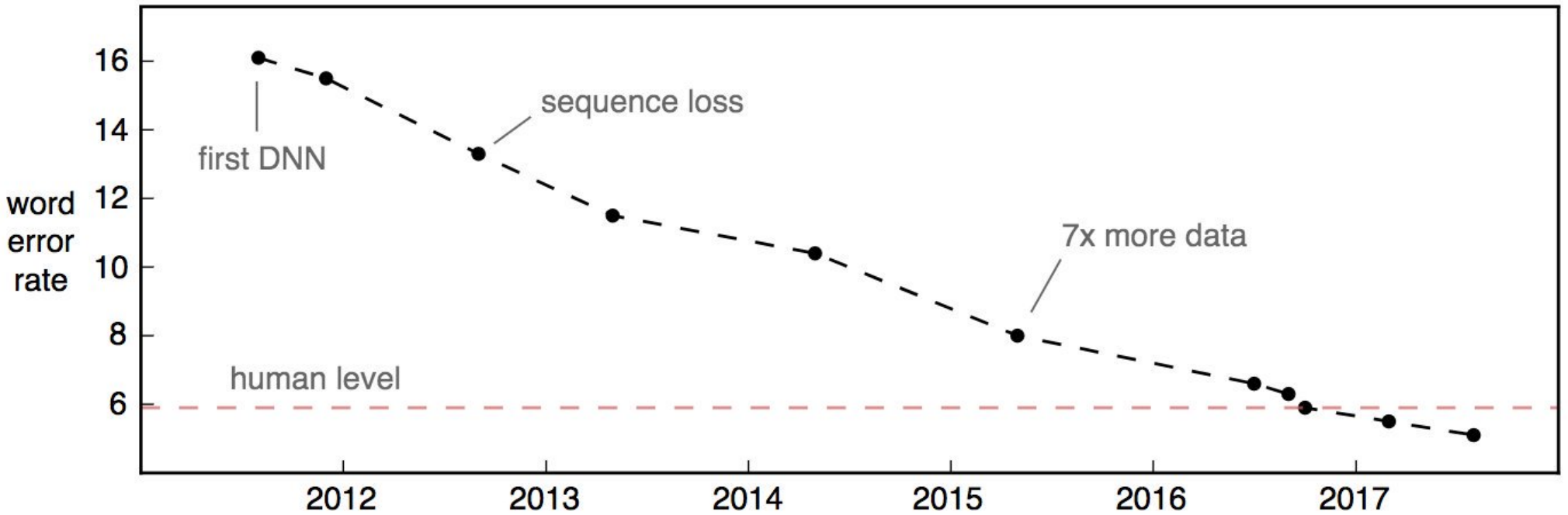
Milestones in Deep Learning History

- 2006: Geoffrey Hinton's "Deep Belief Networks" - The modern era of deep learning begins.
- 2012: AlexNet's ImageNet victory brings mainstream attention.
- 2014-2015: Introduction of Generative Adversarial Networks (GANs).
- 2016: Google's AlphaGo uses deep learning to defeat a Go world champion.

Milestones in Deep Learning History

- 2017: Transformer architecture revolutionizes Natural Language Processing.
- 2018-2019: GPT-2 and BERT gain popularity.
- 2020-2021: COVID-19 drives deep learning in healthcare.
- 2022: LLM (Large Language Model) and ChatGPT
- 2022-2023: Anticipated breakthroughs in the maturing field.

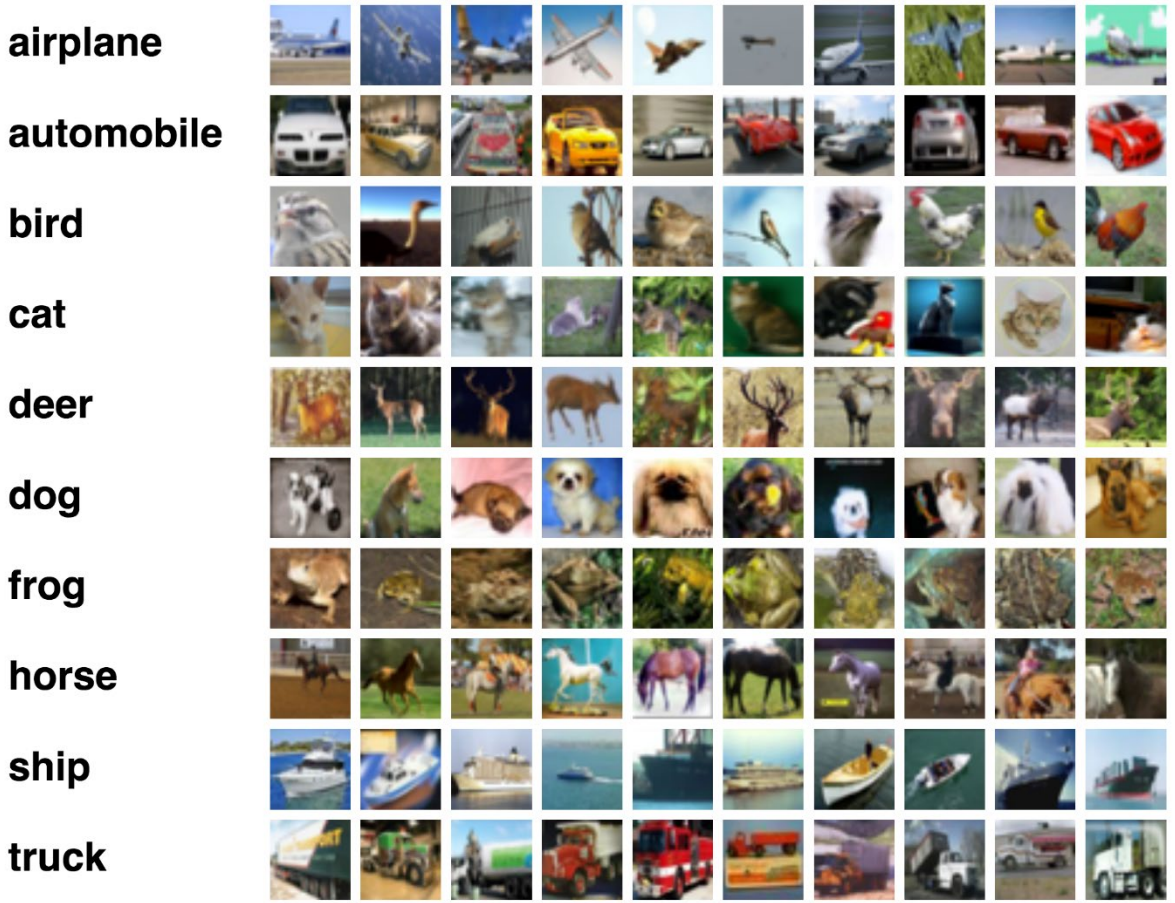

According to Microsoft's
speech group:
Using DL
Word error rate on Switchboard
100%
10%
4%
2%
1%
1990
2000
2010

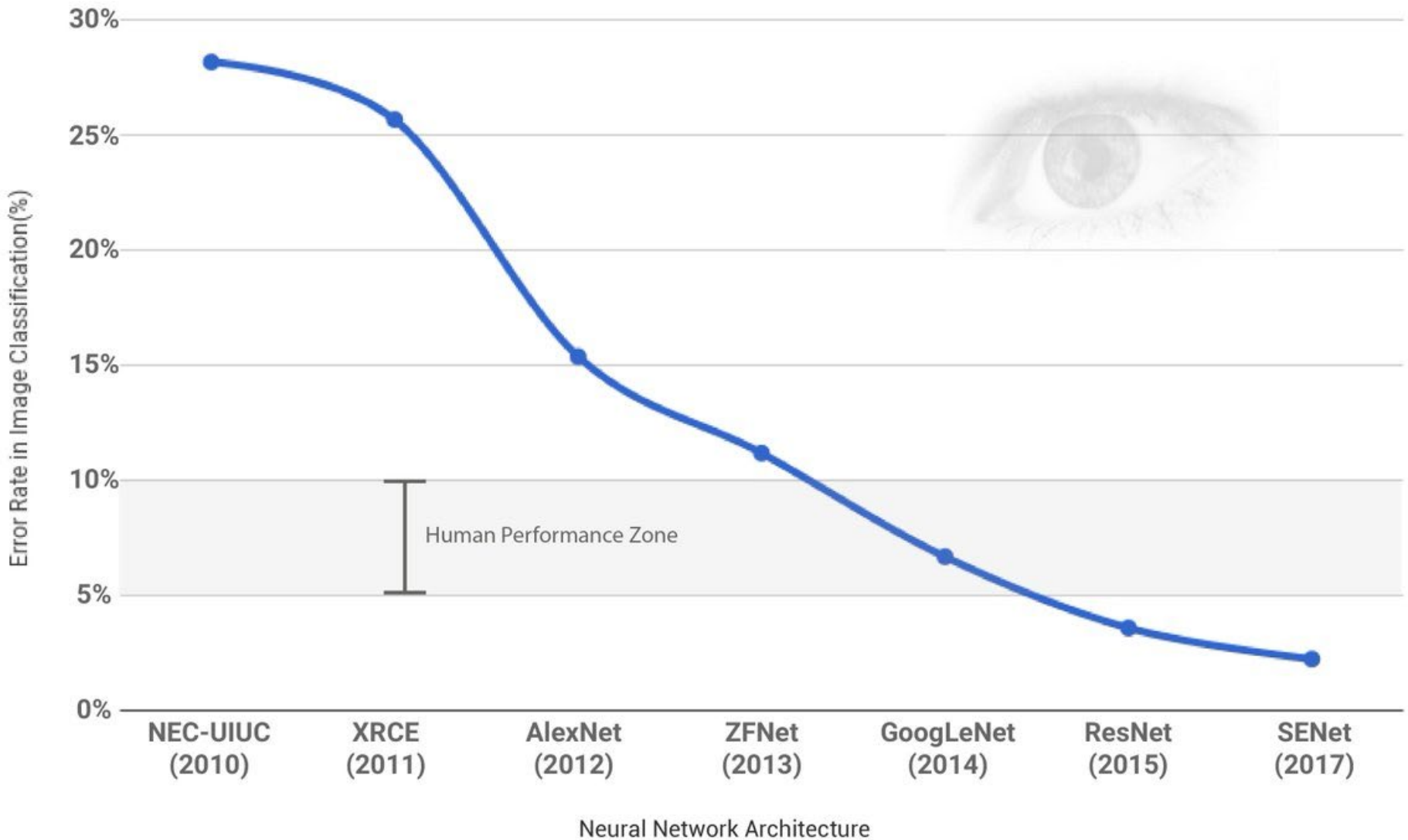


Improvements in word error rate over time on the Switchboard conversational speech recognition benchmark. The test set was collected in 2000. It consists of 40 phone conversations between two random native English speakers.

Credit: ML review

ImageNet

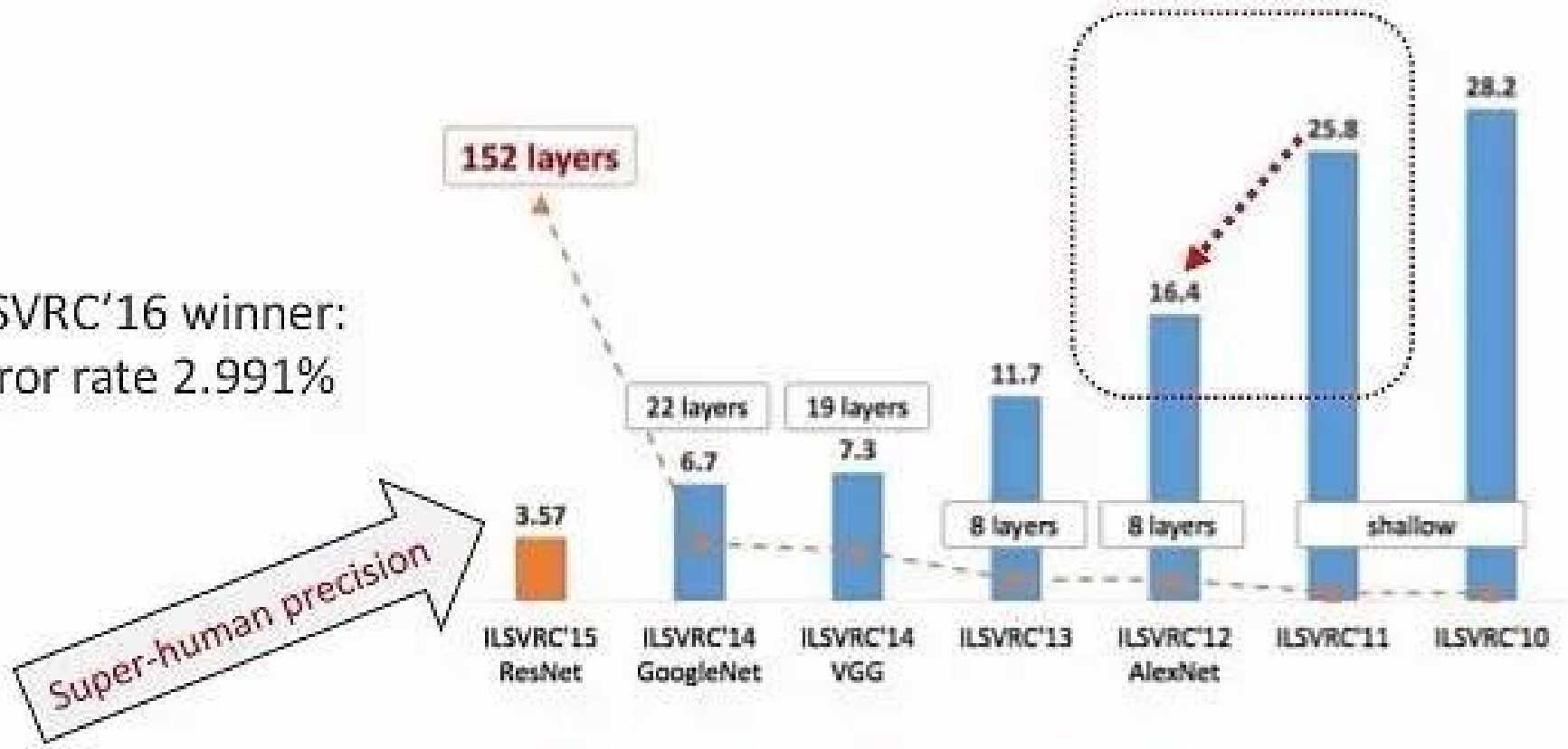

airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck



Credit: ML review

ImageNet Challenge

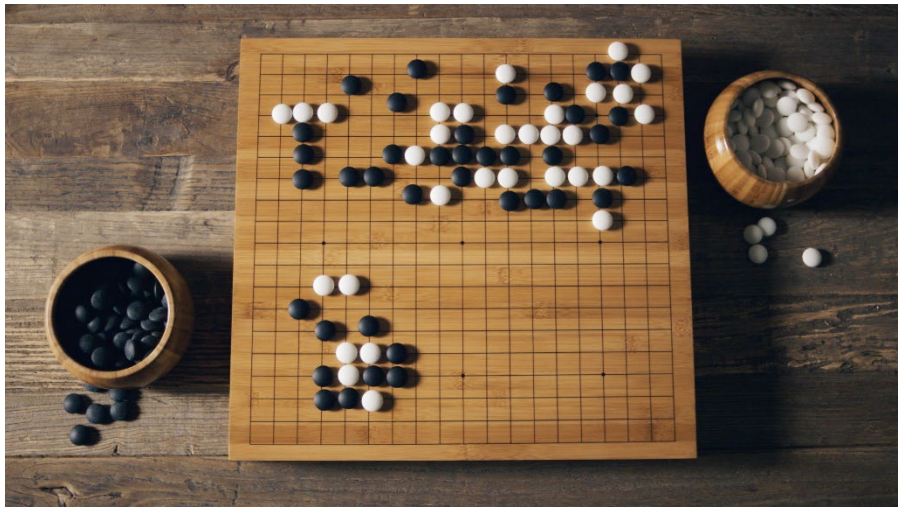
ILSVRC'16 winner:
Error rate 2.991%



* Human-level performance: 5.1%

Success Stories

- AlphaGo (2016) algorithm masters ancient game of Go
- **Deep-learning software defeats human professional for first time.**



AlphaGo

- *March* 2016 : AlphaGo defeats Lee Sedol (9-dan)

  Ke Jie (world champion): "AlphaGo can't beat me"

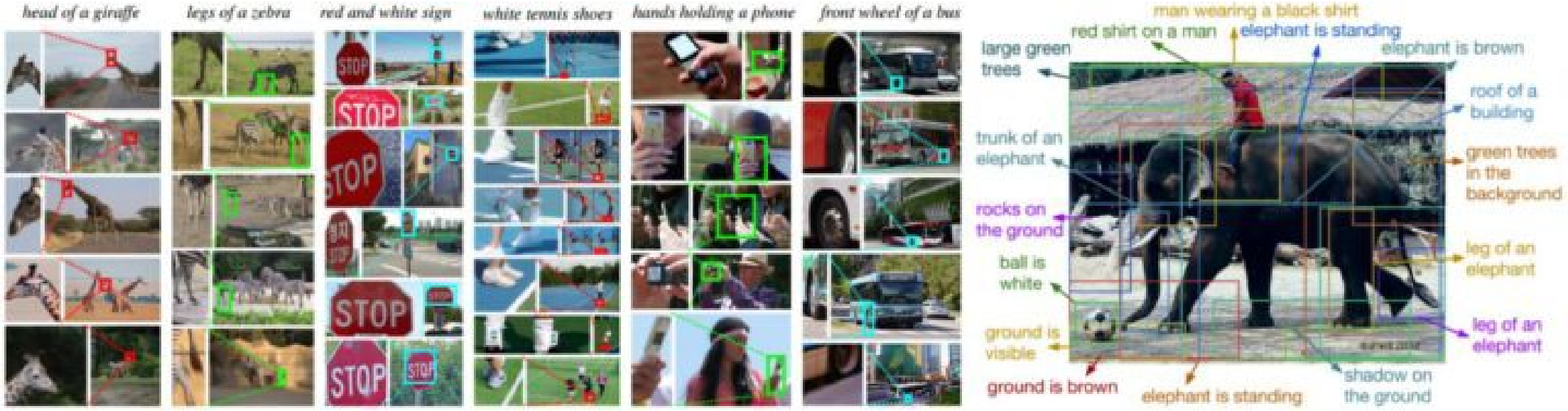
- *May* 2017 : AlphaGo defeats Ke Jie (world champion)

Recovering sound waves from the vibrations

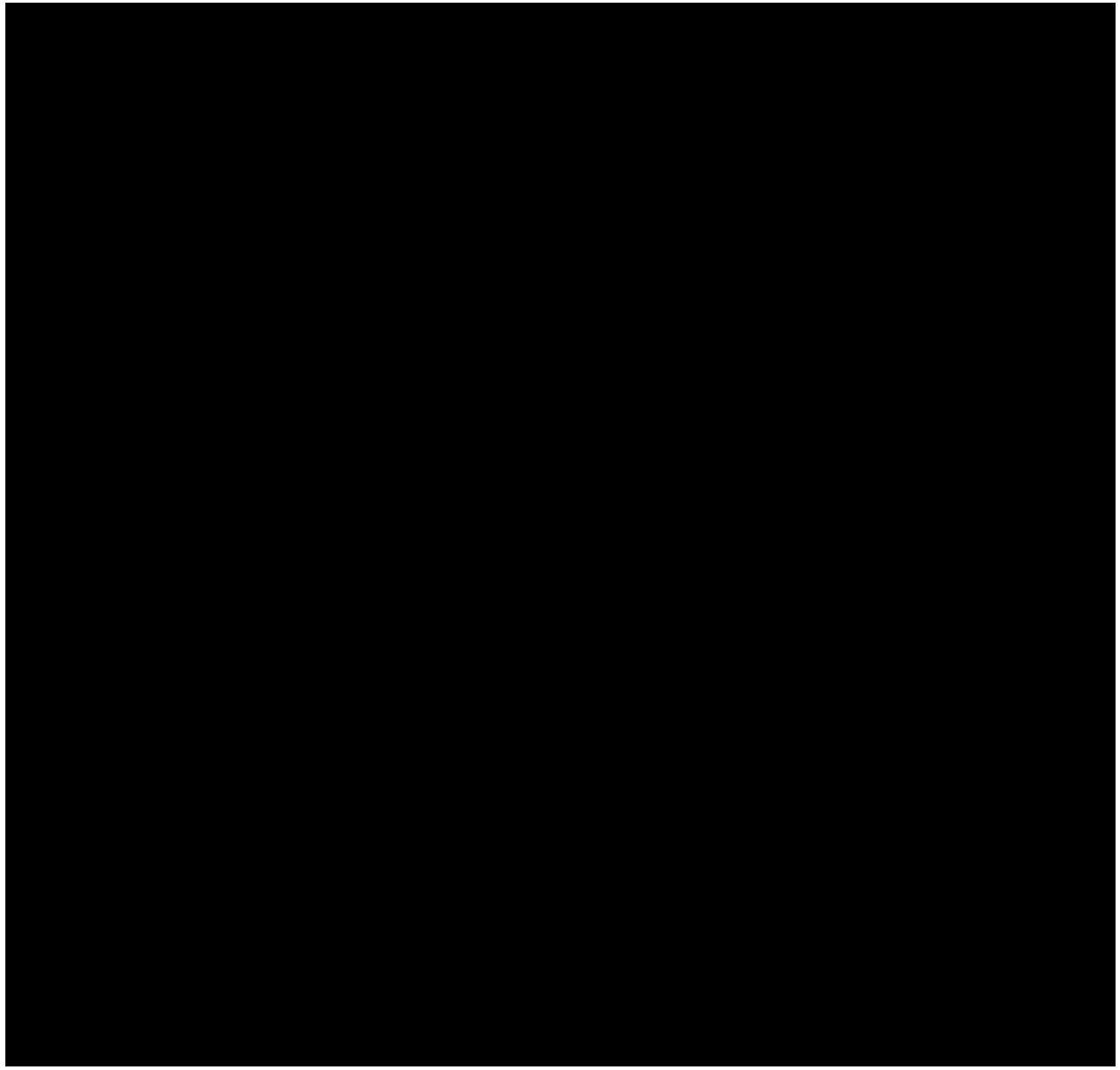
Davis, A., Rubinstein, M., Wadhwa, N., Mysore, G., Durand, F., and Freeman, W. T.(2014). The visual microphone: Passive recovery of sound from video. ACM Transac-tions on Graphics (Proc. SIGGRAPH), 33(4), 79:179:10.

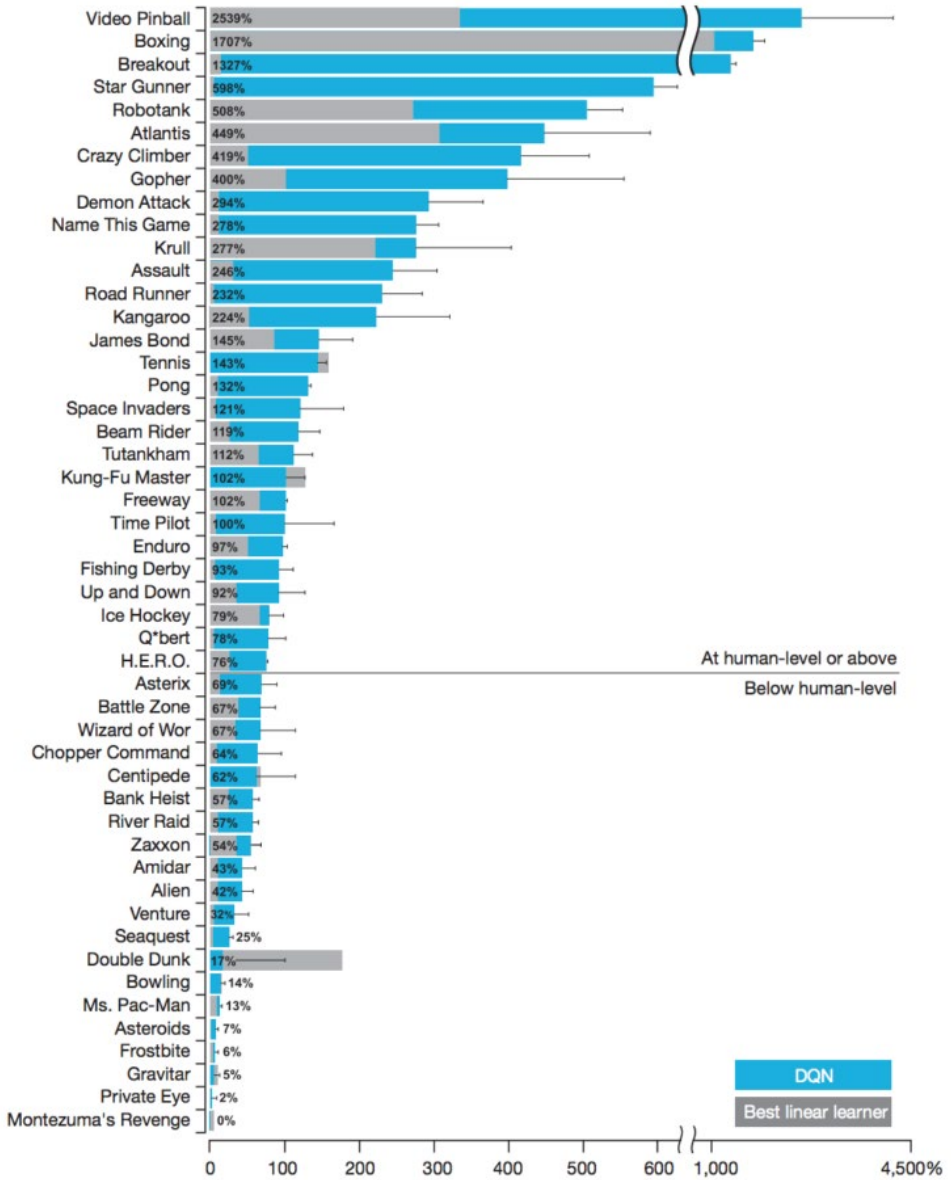
The visual microphone

Describing photos



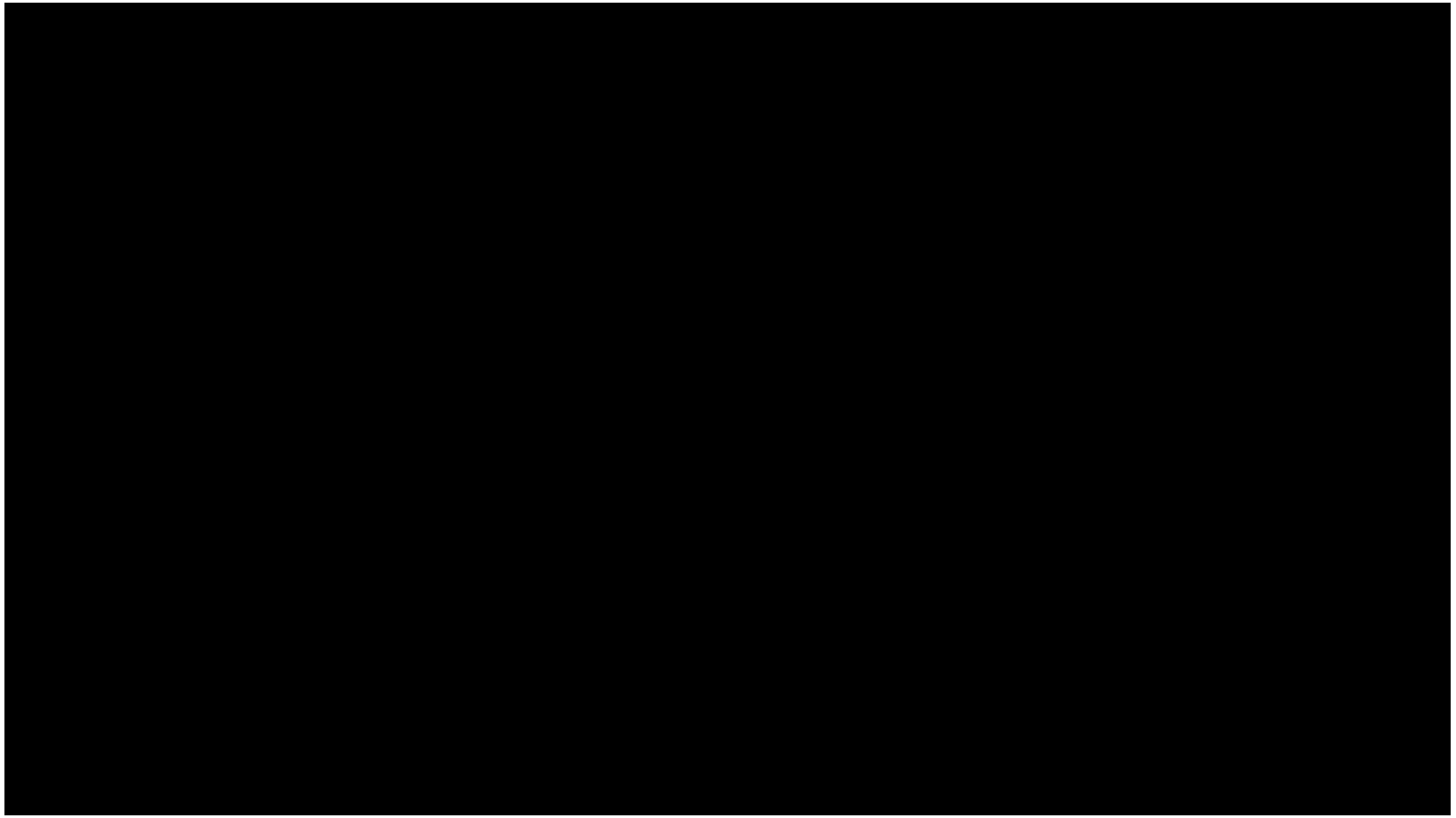
Winning Atari Breakout




Video Pinball 2539%
Boxing 1707%
Breakout 1327%
Star Gunner 598%
Robotank 508%
Atlantis 449%
Crazy Climber 419%
Gopher 400%
Demon Attack 294%
Name This Game 278%
Krull 277%
Assault 246%
Road Runner 232%
Kangaroo 224%
James Bond 145%
Tennis 143%
Pong 132%
Space Invaders 121%
Beam Rider 119%
Tutankham 112%
Kung-Fu Master 102%
Freeway 102%
Time Pilot 100%
Enduro 97%
Fishing Derby 93%
Up and Down 92%
Ice Hockey 79%
Q*bert 78%
H.E.R.O. 76%
Asterix 69%
Battle Zone 67%
Wizard of Wor 67%
Chopper Command 64%
Centipede 62%
Bank Heist 57%
River Raid 57%
Zaxxon 54%
Amidar 43%
Alien 42%
Venture 32%
Seaquest 25%
Double Dunk 17%
Bowling 14%
Ms. Pac-Man 13%
Asteroids 7%
Frostbite 6%
Gravitar 5%
Private Eye 2%
Montezuma's Revenge 0%
At human-level or above
Below human-level
DQN
Best linear learner
0
100
200
300
400
500
600
1,000
4,500%

Self-driving cars

LipNet



Generative Models

- Image Generation
- Data Augmentation
- Text Generation
- Music Generation
- Drug Discovery
- Speech Synthesis

Image Generation

https://thispersondoesnotexist.com/

ChatGPT - The Conversational AI

- Text Summarization
- Language Translation
- Code Generation
- Content Creation
- Customer Support
- Tutoring & Education
- General Q&A

Books written by AI

Alice and Sparkle Paperback – Jan. 14 2023

by Ammaar Reshi (Author), Chat GPT (Author), Mid Journey (Illustrator)

3.1 ★★★☆☆ 30 ratings

See all formats and editions

Get up to **$150 in rewards** with the Amazon.ca Rewards Mastercard. **No annual fee.**

Audiobook	Hardcover	**Paperback**
$0.00	$17.35 prime	**$11.95** prime
Free with your Audible trial	3 New from $17.35	3 New from $11.95

The children's book covered in The Washington Post, TIME Magazine, Le Figaro, NBC News, and media worldwide.

This is a story about a young girl named Alice who discovers the magic of artificial intelligence. She creates her own AI, named Sparkle, and together they go on adventures and use their combined knowledge to make the world a better place. The story explores the incredible abilities of AI and the importance of using them for good. It is a tale of friendship and exploration, filled with magic and wonder.

Images from natural language descriptions

- Midjourney generat images from natural language descriptions.

- Similar to DALL-E and Stable Diffusion.

Image from text

Quran: Ad-Dukhan

Indeed, the righteous will be in a secure place; (51)

Within gardens and springs, (52) Wearing [garments of] fine silk and brocade, facing each other.

إِنَّ الْمُتَّقِينَ فِي مَقَامٍ أَمِينٍ ﴿٥١﴾ فِي
جَنَّاتٍ وَعُيُونٍ ﴿٥٢﴾ يَلْبَسُونَ مِن سُندُسٍ وَإِسْتَبْرَقٍ مُّتَقَابِلِينَ

Books written by AI

- The novella is the result of a collaboration between Stephen Marche and three artificial intelligence programs.

Books written by AI

- There were over 200 e-books in Amazon's Kindle store as of mid-February 2023 listing ChatGPT as an author or co-author

- AI-Generated Books of Nonsense - These are books that are created by AI and are often nonsensical. They have been found on Amazon's bestseller lists, but Amazon has been removing them due to copyright violations

What is Machine Learning

Classical Statistics

- Infer information from small data sets (Not enough data)

Machine Learning

- Infer information from large data sets (Too many data)

What is Machine Learning?

- Machine Learning is the ability to teach a computer without explicitly programming it
- Examples are used to train computers to perform tasks that would be difficult to program

Machine Learning

- *Supervised Learning* :
  Teacher tells learner what to remember
- *Reinforcement Learning* :
  Environment provides hints to learner
- *Unsupervised Learning* :
  Learner discovers on its own

Fundamental problems

- Classification
- Regression
- Clustering
- Dimensionality reduction (aka Feature extraction, Manifold learning)

Classification

Consider data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ where $\mathbf{x} \in \mathbb{R}^d$ and y_i **takes values in some finite set**.

Find a function f ,such that when we observe a new $\mathbf{x}$ we predict y to be $f(\mathbf{x})$.

Regression

Consider data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ where $\mathbf{x} \in \mathbb{R}^d$ and y_i **takes values in** $\mathbb{R}$.

Find a function f, such that when we observe a new $\mathbf{x}$ we predict y to be $f(\mathbf{x})$.

Clustering

Consider data $\{\mathbf{x}_i\}_{i=1}^n$ where $\mathbf{x} \in \mathbb{R}^d$.

Find a function f, when we observe a new $\mathbf{x}$ we predict y to be $f(\mathbf{x})$, such that for similar $\mathbf{x}$, y is the same.

What is Machine Learning?

- Machine Learning is just glorified "curve fitting"?

General Procedure

1. Model (Hypothesis class)

   $f(\mathbf{x}) \in \mathcal{F}$ (Hypothesis class)

2. Score Criterion

   Population: $S(f) = E_{\mathbf{x},y} L(y, f(\mathbf{x}))$

   Sample: $\frac{1}{n} \sum_{i=1}^{n} L(y_i, f(\mathbf{x}_i))$

3. Search Strategy

   $\hat{f} = arg \ \min_{f \in \mathcal{F}} S(f)$

Model
(Hypothesis class)

- Linear (Perceptron, SVM, Logistic regression)
- Directed graphical model
- Undirected graphical model
- Matrices

Score criterion

Cost for error: $L(y, F)$

$$L(y, F) = |\, y - F \,|,\ (y - F)^2, \qquad y \in R$$

$y \in \{-1, 1\}$:

$$L(y, F) = \log(1 + e^{-yF}) \quad \text{logistic reg.}$$

$$L(y, F) = \max(0, 1 - yF) \quad \text{SVM}$$

Many many more

Deep Structure as a Hypothesis class

- Feedforward Deep Networks
- Convolutional Networks
- Recurrent neural networks (RNN)

- Restricted Boltzmann Machines (RBM)

Tentative topics

- Feedforward Deep Networks
- Optimization and regularization for Training Deep Models
- Convolutional Networks
- Recurrent Neural Networks
- Seq2seq models
- Deep Generative Models (Moment Matching Networks)
- Generative Adversarial Networks (GANs)

Tentative topics

- Attention, Self-attention, Transformers, Performers
- BERT, GPT
- Auto-Encoders
- Variational Autoencoders
- Deep Reinforcement Learning
- Graph Neural Networks
- Diffusion models
- Large Language Models (LLMs)

Materials and Resources

- **Textbook**

Required

- No textbook is required for this course.

Recommended

- Deep Learning by Ian Goodfellow and Yoshua Bengio and Aaron Courville
- The book is online https://www.deeplearningbook.org/

Data Challenge - Overview

- Individual Kaggle Competition
- Submit on LEARN: Code & Brief Report
- WatIAM ID for Leaderboard
- Questions: Piazza under data_challenge
- Create Kaggle Account
- Submit Predictions on Kaggle
- Submit Code & Report on LEARN
- Zip File: First_Last_Number.zip

Data Challenge - Overview

- Clean and Readable Code
- Necessary Comments
- Include All Training Parameters
- Seed Set in Code
- Code: .py or .ipynb
- One File per Assignment
- Email Confirmation Receipt

Data Challenge - Overview

- Public & Private Leaderboard Scores
- Minimum & Maximum Thresholds
- Grading Scheme:

$$\frac{(avg - \min)}{(\max - \min)}$$

- No More Than 100%, No Less Than 0%

Paper Presentation (10%)

- Group presentation
- Use Piazza to Find Partner
- Choose Paper from Listed Conferences
- Latest Conference & Deep Learning Specific
- Register on Designated Spreadsheet
- Check Course Schedule for Dates
- Mandatory Attendance & Peer Review

Final Project: **Exploring Novel Ideas**

- **Exciting Opportunity**: Venture into new, unique ideas.
- **Risk & Reward**: Success is not guaranteed, but the journey is valuable.
- **Evaluation Criteria**:
  - Quality of Idea
  - Logical Progression
- **Grade Determinants**: Innovation, Effort, and Soundness of Approach.
- **Spirit of Research**: Explore uncharted territories and contribute new insights.

Final Project : **Applying Existing Algorithms**

- **Strategic Approach**: Utilize well-known algorithms for specific problems.
- **Avoid Triviality**: Aim for complexity and meaningful challenges.
- **Balance**: Don't just pair known tools with known problems.
- **Demonstrate Skills**: Solve nontrivial problems or face implementation challenges.
- **Project Goals**:
  - Meaningful Exploration
  - Creative Application
  - Problem-Solving
- **Outcome**: Showcase your ability to innovate and advance knowledge.

Project Report Guidelines

- **Cover Page**: Accurately filled with essential project details.
- **Report Length**: Max 8 pages, 12-point font, single-column format. Excludes references.
- **References**: Cite relevant external sources.
- **Submission**: Merge elements into a single PDF file.

Course Communication via Piazza

- Discuss Course Materials
- Ask Questions Publicly
- Learn from Others
- Avoid Redundancy
- Announcements & Clarifications
- Regular Reading Expected
- Would you like to proceed with something else?

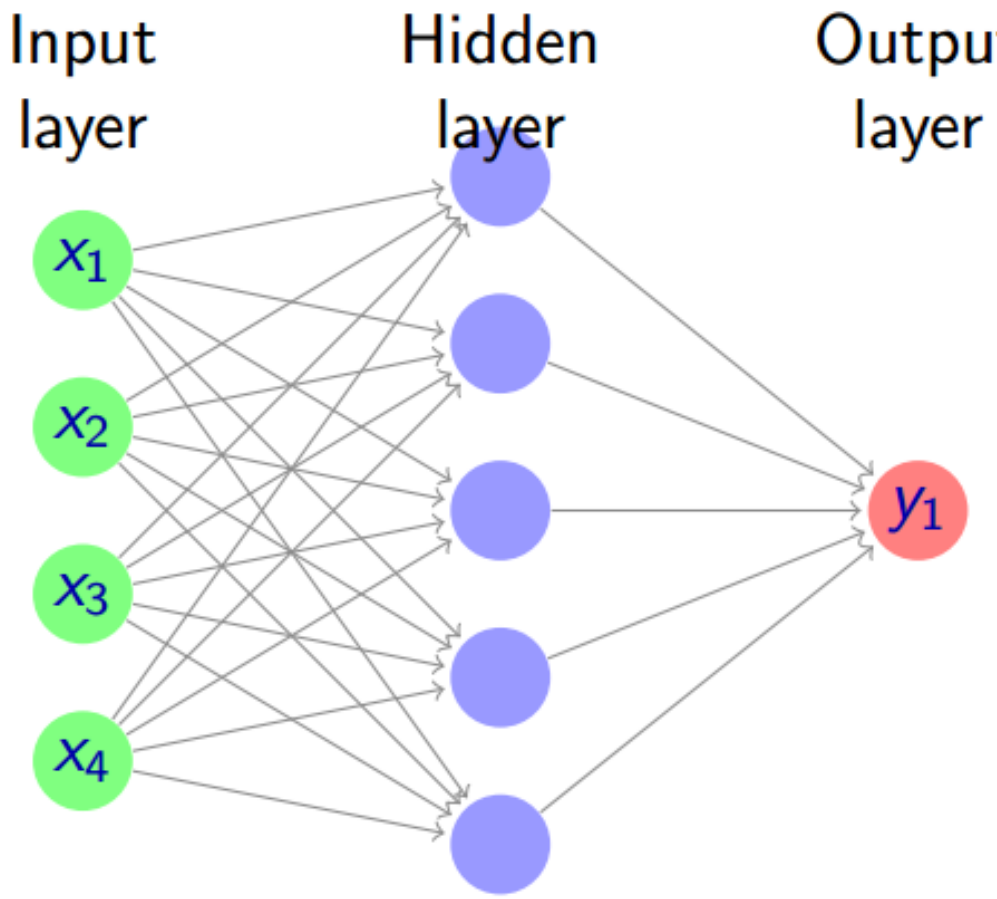
Antirequisite

We will cover

- Perceptron
- Regularization
- Feedforward Neural Networks
- Convolutional Neural Networks

In STAT 441/841 CM 763

Feedforward Neural Network



History, McCulloch and Pitts network

1943

The first model of a neuron was invented by McCulloch (physiologists) and Pitts (logician).

The model had two inputs and a single output.

A neuron would not activate if only one of the inputs was active.

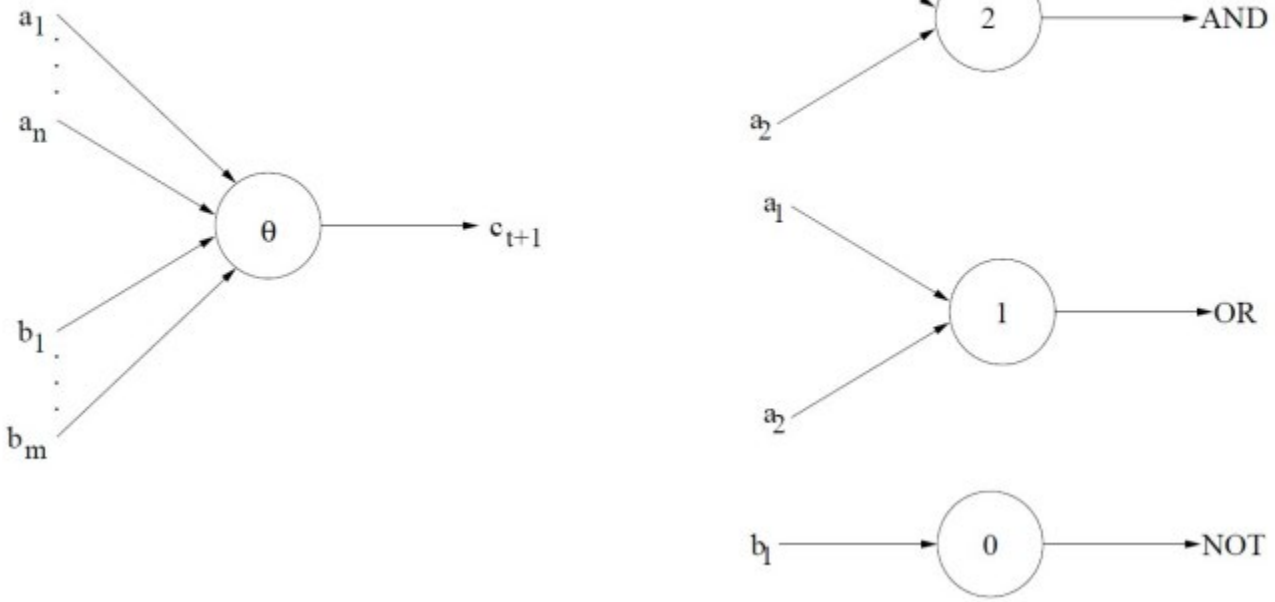
The weights for each input were equal, and the output was binary.

Until the inputs summed up to a certain threshold level, the output would remain zero.

The McCulloch and Pitts' neuron has become known today as a logic circuit.

History, McCulloch and Pitts network (MPN)

1943

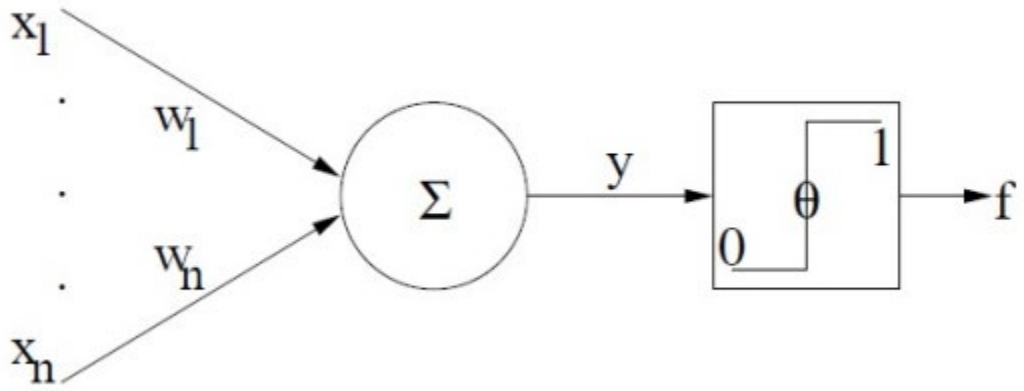


logic functions can be modeled by a network of MP-neurons

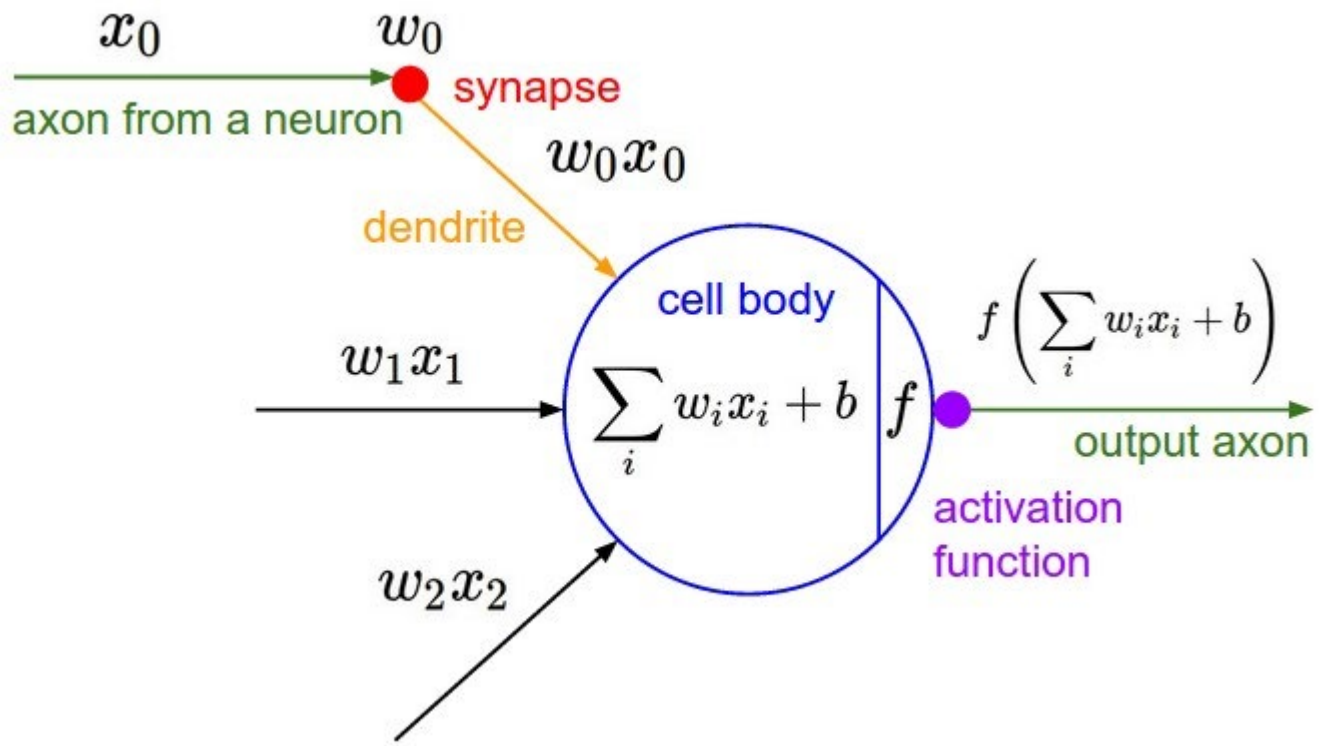
History, Perceptron

1958

The perceptron was developed by Rosenblatt (physiologist).



Credit:



Crediit: http://cs231n.github.io/convolutional-networks/

Perceptron, the dream

1958

Rosenblatt randomly connected the perceptrons and changed the weights in order to achieve "learning."

Based on Rosenblatt's statements in a press confrence in 1958, The New York Times reported the perceptron to be 'the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.'

History, optimization

1960

Widrow and Hoff proposed a method for adjusting the weights. They introduced a gradient search method based on minimizing the error squared (Least Mean Squares).

In the 1960's, there were many articles promising robots that could think.

It seems there was a general belief that perceptrons could solve any problem.

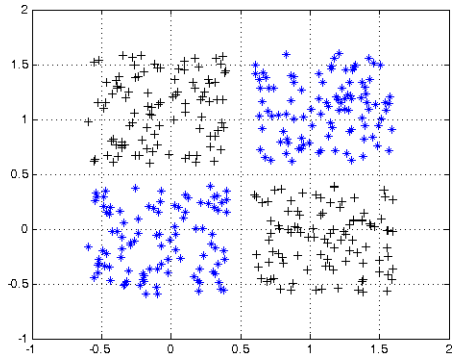
History, shattered dream

1969

Minsky and Papert published their book *Perceptrons*. The book shows that perceptrons could only solve linearly separable problems.

They showed that it is not possible for perseptron to learn an XOR function.

After Perceptrons was published, researchers lost interest in perceptron and neural networks.



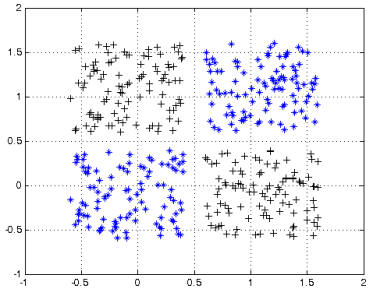
History, optimization

1969

Arthur E. Bryson and Yu-Chi Ho described proposed Backpropagation as a multi-stage dynamic system optimization method. (Bryson, A.E.; W.F. Denham; S.E. Dreyfus. Optimal programming problems with inequality constraints. I: Necessary conditions for extremal solutions. AIAA J. 1, 11 (1963) 2544-2550)

1972

Stephen Grossberg proposed networks capable of learning XOR function.



History

1974

Backpropagation was reinvented / applied in the context of neural networks by Paul Werbos, David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams.

Back propagation allowed perceptrons to be trained in a multilayer configuration.

Deep Learning

Returned again in the 2010s, now able to train much larger networks using huge modern computing power such as GPUs.