PAT 498/598 (Winter 2025)

Music & AI

Lecture 4: Machine Learning Fundamentals

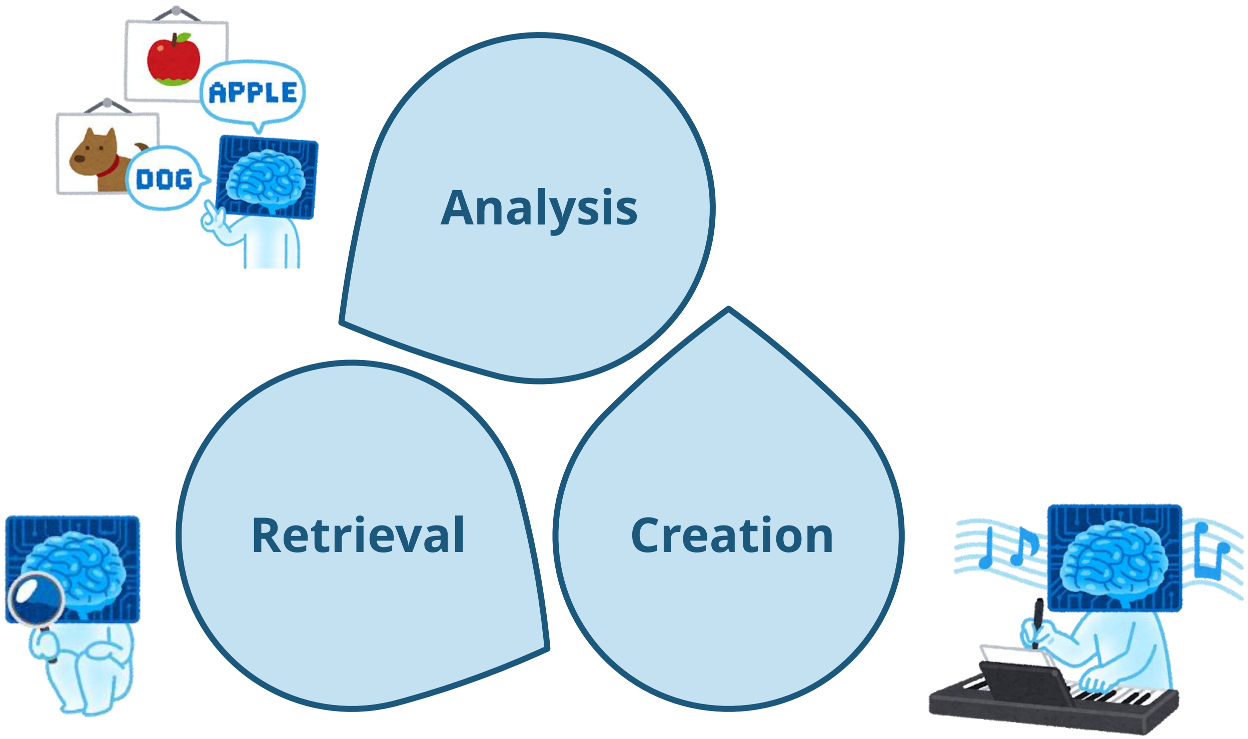
Instructor: Hao-Wen Dong

Homework 1: Real of Fake!?

- **Don't worry! You'll be graded by completion (instead of correctness)!**
- Instructions will be sent by **emails** and released on the **course website**
- Please submit you work to **Gradescope**
- Due at **11:59pm ET** on **January 22**
- Late submissions: **1 point deducted per day**

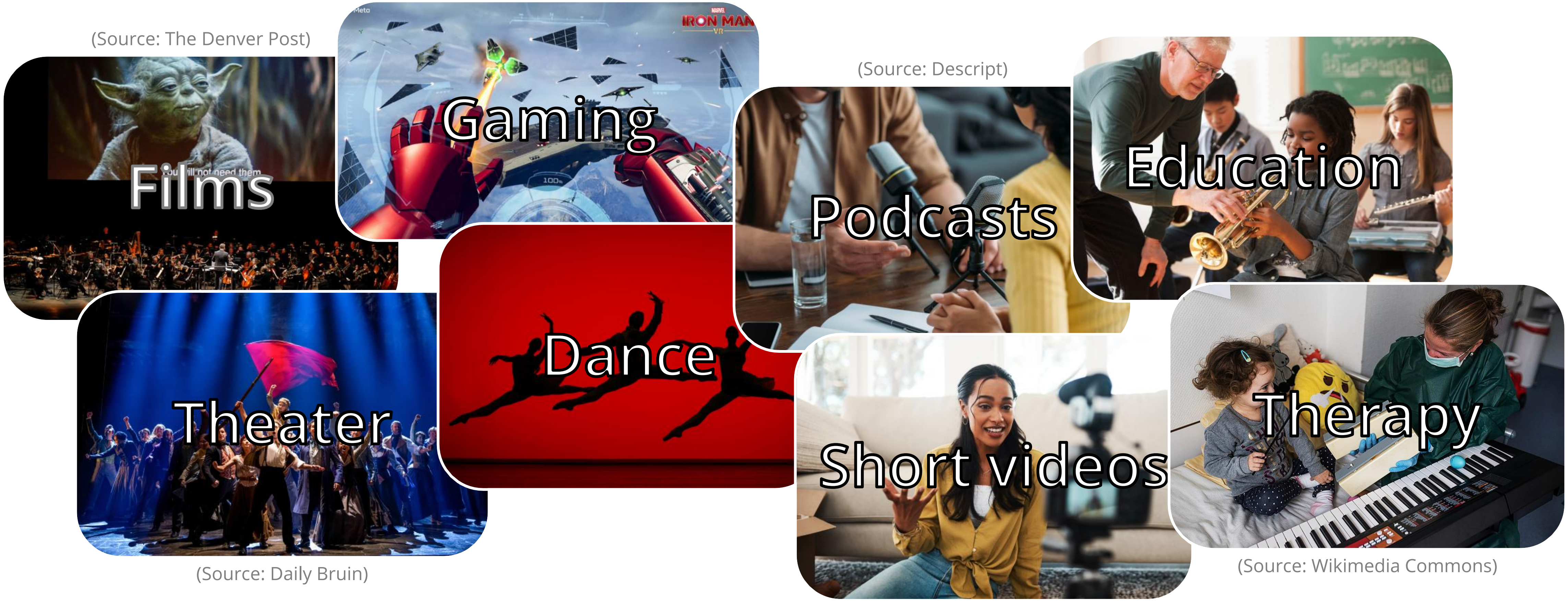
(Recap) Music Information Research (MIR)

- “Intelligent ways to analyze, retrieve and create music” (Yang 2018)



Yang, “Music Information Research,” SNHCC, TIGP, lecture notes, April 2018.

(Recap) Use Cases of AI for Music & Audio



(Source: UploadVR)

(Source: The Denver Post)

(Source: Descript)

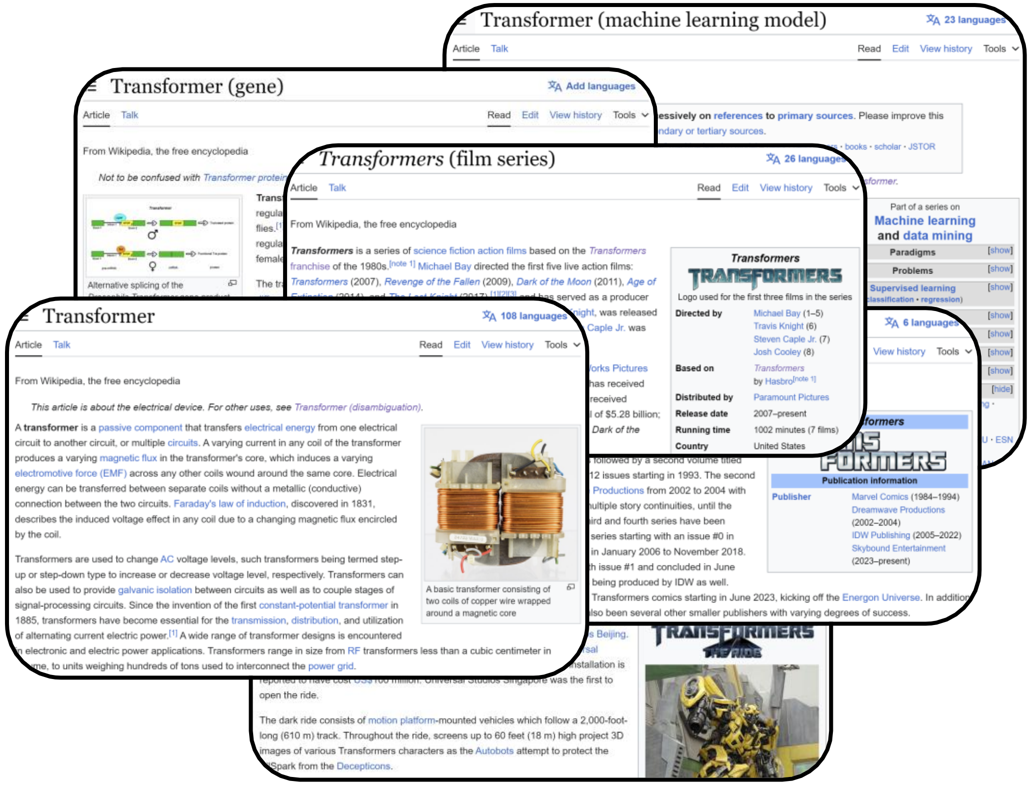
(Source: Daily Bruin)

(Source: Wikimedia Commons)

Universitaetsmedizin, CC BY-SA 4.0, via Wikimedia Commons
uploadvr.com/iron-man-vr-breaks-free-from-cords-load-screens-on-quest-2/
descript.com/blog/article/what-is-the-best-audio-interface-for-recording-a-podcast
denverpost.com/2019/08/02/colorado-symphony-movie-scores-harry-potter-star-wars/
dailybruin.com/2023/08/04/theater-review-the-musical-les-misrables-offers-stellar-displays-and-impassioned-vocals

(Recap) Language Models

- Predicting the next word given the past sequence of words


Transformer (machine learning model)
Transformer (gene)
Transformers (film series)

Transformer

From Wikipedia, the free encyclopedia

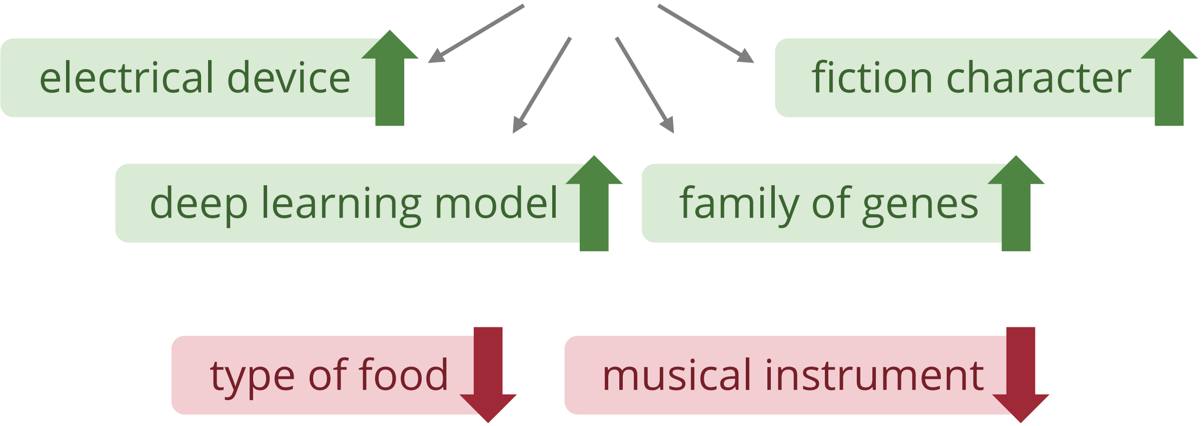
This article is about the electrical device. For other uses, see Transformer (disambiguation).

A **transformer** is a passive component that transfers electrical energy from one electrical circuit to another circuit, or multiple circuits. A varying current in any coil of the transformer produces a varying magnetic flux in the transformer's core, which induces a varying electromotive force (EMF) across any other coils wound around the same core. Electrical energy can be transferred between separate coils without a metallic (conductive) connection between the two circuits. Faraday's law of induction, discovered in 1831, describes the induced voltage effect in any coil due to a changing magnetic flux encircled by the coil.

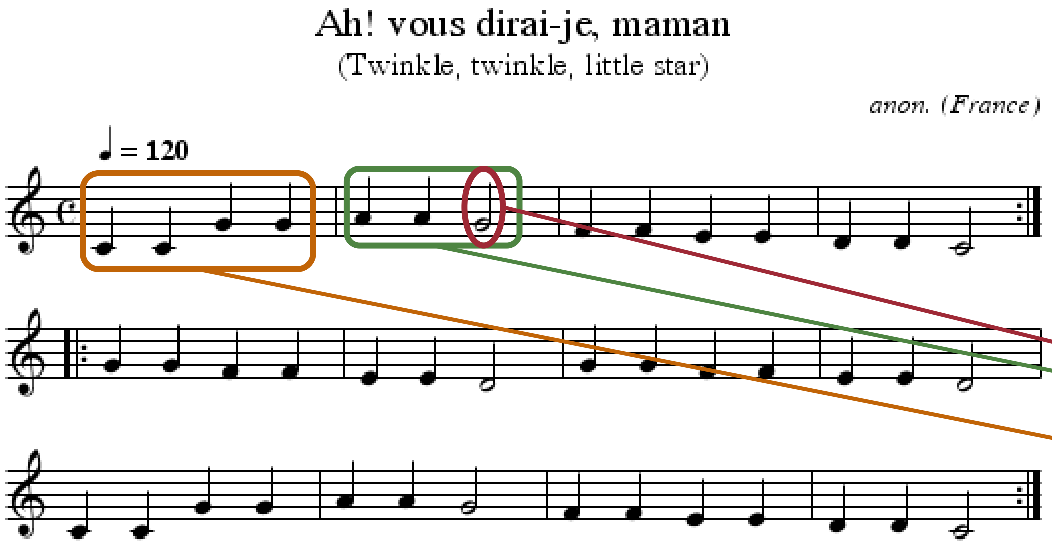
Transformers are used to change AC voltage levels, such transformers being termed step-up or step-down type to increase or decrease voltage level, respectively. Transformers can also be used to provide galvanic isolation between circuits as well as to couple stages of signal-processing circuits. Since the invention of the first constant-potential transformer in 1885, transformers have become essential for the transmission, distribution, and utilization of alternating current electric power.[1] A wide range of transformer designs is encountered in electronic and electric power applications. Transformers range in size from RF transformers less than a cubic centimeter in volume, to units weighing hundreds of tons used to interconnect the power grid.

A basic transformer consisting of two coils of copper wire wrapped around a magnetic core

A transformer is a ________


electrical device ⬆
fiction character ⬆
deep learning model ⬆
family of genes ⬆
type of food ⬇
musical instrument ⬇

(Recap) An Example of ABC Notation



```
X:571
T:Ah! vous dirai-je, maman
T:(Twinkle, twinkle, little star)
C:anon.
O:France
R:Nursery song
M:C
L:1/4
Q:120
K:C
CCGG|AAG2|FFEE|DDC2:|
|:GGFF|EED2|GGFF|EED2|
CCGG|AAG2|FFEE|DDC2:|
```

Metadata (X: to R:)

M:C Meter

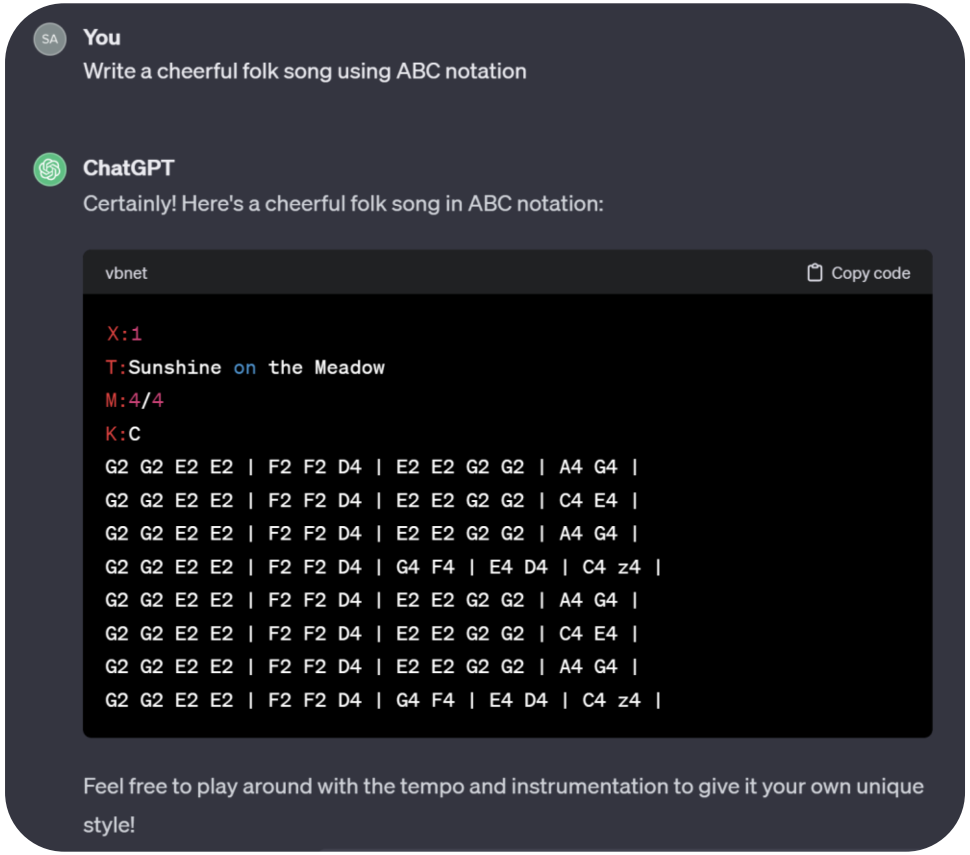
L:1/4 Unit note length (temporal resolution)

Q:120 Tempo

K:C Key

abcnotation.com/tunePage?a=trillian.mit.edu/~jc/music/abc/mirror/musicaviva.com/france/ah-vous-c/ah-vous-c-1/0000

(Recap) ChatGPT can Compose Songs in ABC Notation!



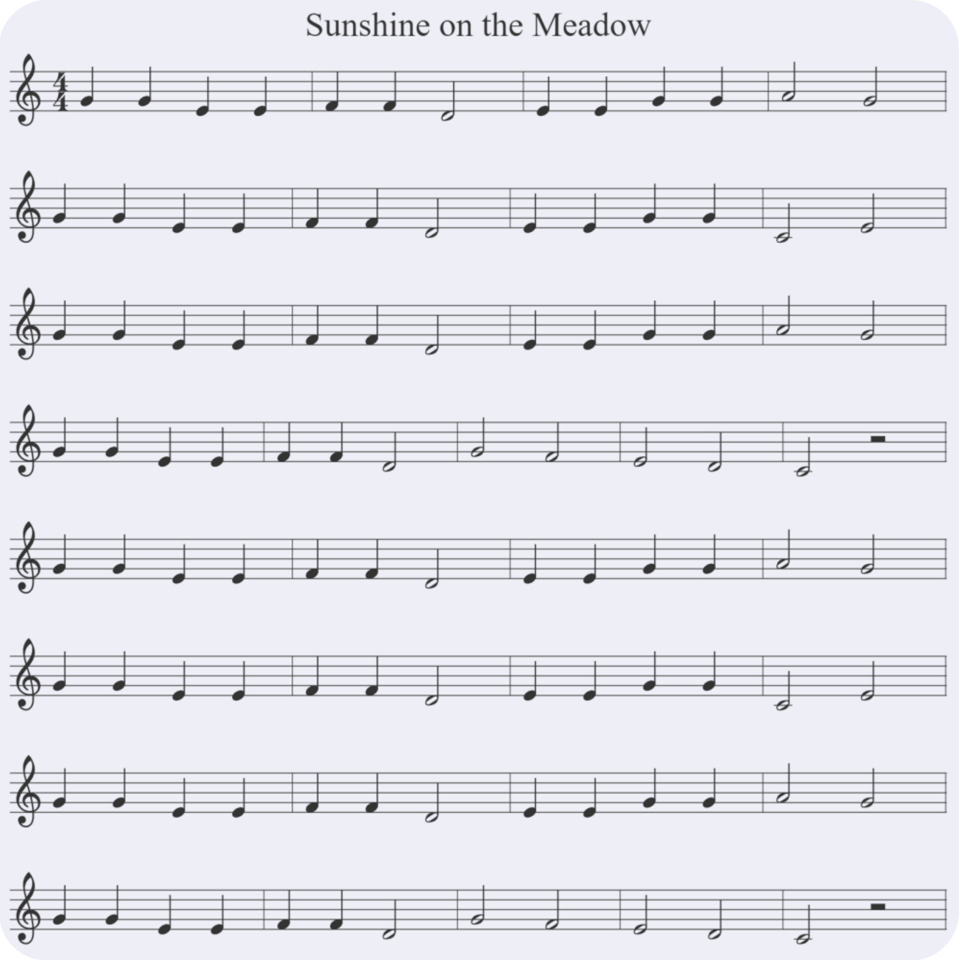
You
Write a cheerful folk song using ABC notation

ChatGPT
Certainly! Here's a cheerful folk song in ABC notation:

```
X:1
T:Sunshine on the Meadow
M:4/4
K:C
G2 G2 E2 E2 | F2 F2 D4 | E2 E2 G2 G2 | A4 G4 |
G2 G2 E2 E2 | F2 F2 D4 | E2 E2 G2 G2 | C4 E4 |
G2 G2 E2 E2 | F2 F2 D4 | E2 E2 G2 G2 | A4 G4 |
G2 G2 E2 E2 | F2 F2 D4 | G4 F4 | E4 D4 | C4 z4 |
G2 G2 E2 E2 | F2 F2 D4 | E2 E2 G2 G2 | A4 G4 |
G2 G2 E2 E2 | F2 F2 D4 | E2 E2 G2 G2 | C4 E4 |
G2 G2 E2 E2 | F2 F2 D4 | E2 E2 G2 G2 | A4 G4 |
G2 G2 E2 E2 | F2 F2 D4 | G4 F4 | E4 D4 | C4 z4 |
```

Feel free to play around with the tempo and instrumentation to give it your own unique style!

chat.openai.com



The sheet music is rendered with abc.rectanglered.com.

(Recap) A.I. Duet (Mann et al, 2016)

youtu.be/0ZE1bfPtvZo
experiments.withgoogle.com/ai/ai-duet/view

github.com/googlecreativelab/aiexperiments-ai-duet

(Recap) The Early Days

Musical Dice Game (1792)

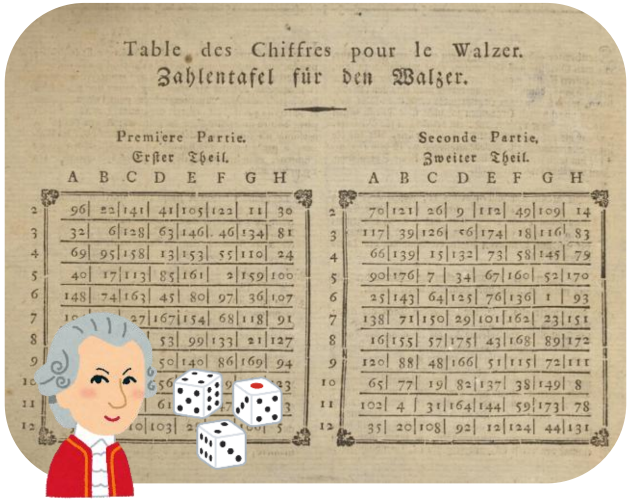


Table des Chiffres pour le Walzer.
Zahlentafel für den Walzer.

Premiere Partie.
Erſter Theil.

	A	B	C	D	E	F	G	H
2	96	22	141	41	105	122	11	30
3	32	6	128	63	146	46	134	81
4	69	95	158	13	153	55	110	24
5	40	17	113	85	161	2	159	100
6	148	74	163	45	80	97	36	107
7	[illegible]	[illegible]	27	167	154	68	118	91
8	[illegible]	[illegible]	[illegible]	53	99	133	21	127
9	[illegible]	[illegible]	[illegible]	50	140	86	169	94
10	[illegible]	[illegible]	[illegible]	56	[illegible]	[illegible]	[illegible]	[illegible]
11	[illegible]	[illegible]	[illegible]	61	[illegible]	[illegible]	[illegible]	[illegible]
12	[illegible]	[illegible]	[illegible]	103	[illegible]	[illegible]	[illegible]	5

Seconde Partie.
Zweiter Theil.

	A	B	C	D	E	F	G	H
2	70	121	26	9	112	49	109	14
3	117	39	126	56	174	18	116	83
4	66	139	15	132	73	58	145	79
5	90	176	7	34	67	160	52	170
6	25	143	64	125	76	136	1	93
7	138	71	150	29	101	162	23	151
8	16	155	57	175	43	168	89	172
9	120	88	48	166	51	115	72	111
10	65	77	19	82	137	38	149	8
11	102	4	31	164	144	59	173	78
12	35	20	108	92	12	124	44	131

(Source: gbrachetta)

gbrachetta.github.io/Musical-Dice/

→

ILLIAC Suite (1957)

(Source: Illinois Distributed Museum)

→

Emily Howell (2003)

(Source: The Guardian)

gbrachetta.github.io/Musical-Dice/
distributedmuseum.illinois.edu/exhibit/illiac-suite/
theguardian.com/technology/2010/jul/11/david-cope-computer-composer

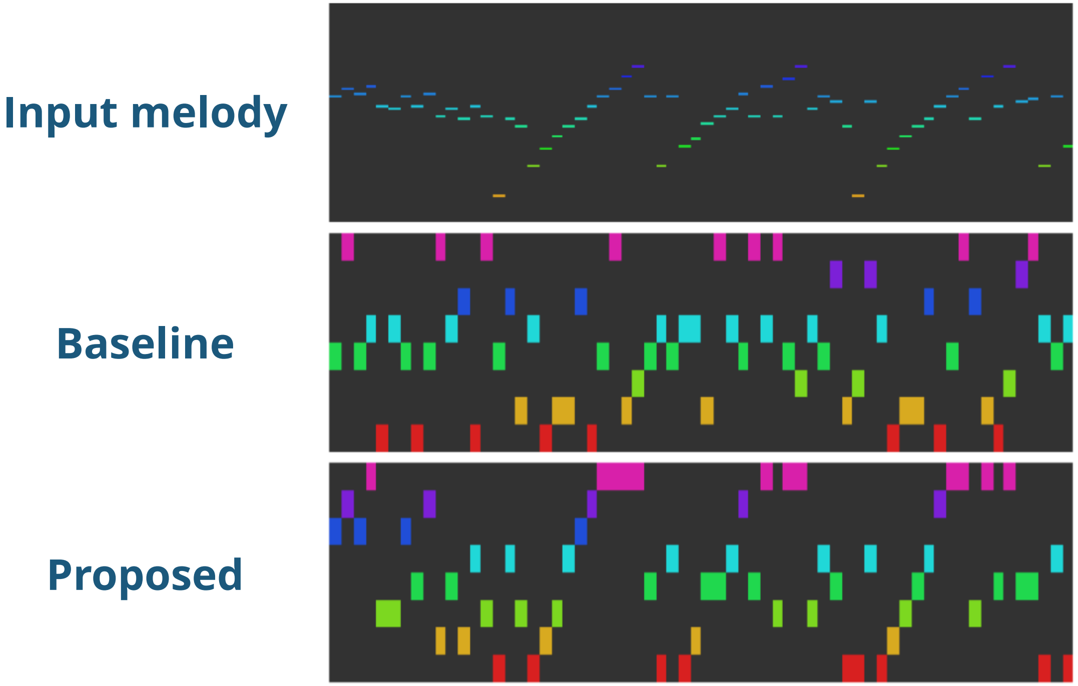
(Recap) Piano Genie (Donahue et al., 2018)

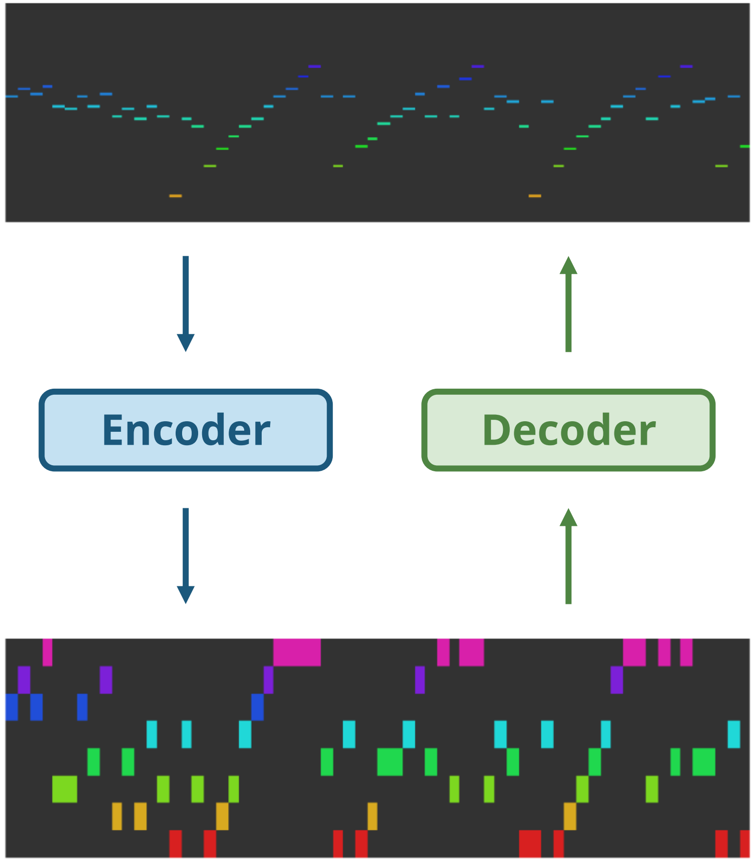
youtu.be/YRb0XAnUpIk & magenta.tensorflow.org/pianogenie

piano-genie.glitch.me/

Chris Donahue, Ian Simon, and Sander Dieleman, “Piano Genie,” IUI, 2019.

(Recap) Piano Genie (Donahue et al., 2018)




Encoder

Decoder

(Source: Donahue et al., 2019)

Chris Donahue, Ian Simon, and Sander Dieleman, “Piano Genie,” *IUI*, 2019.

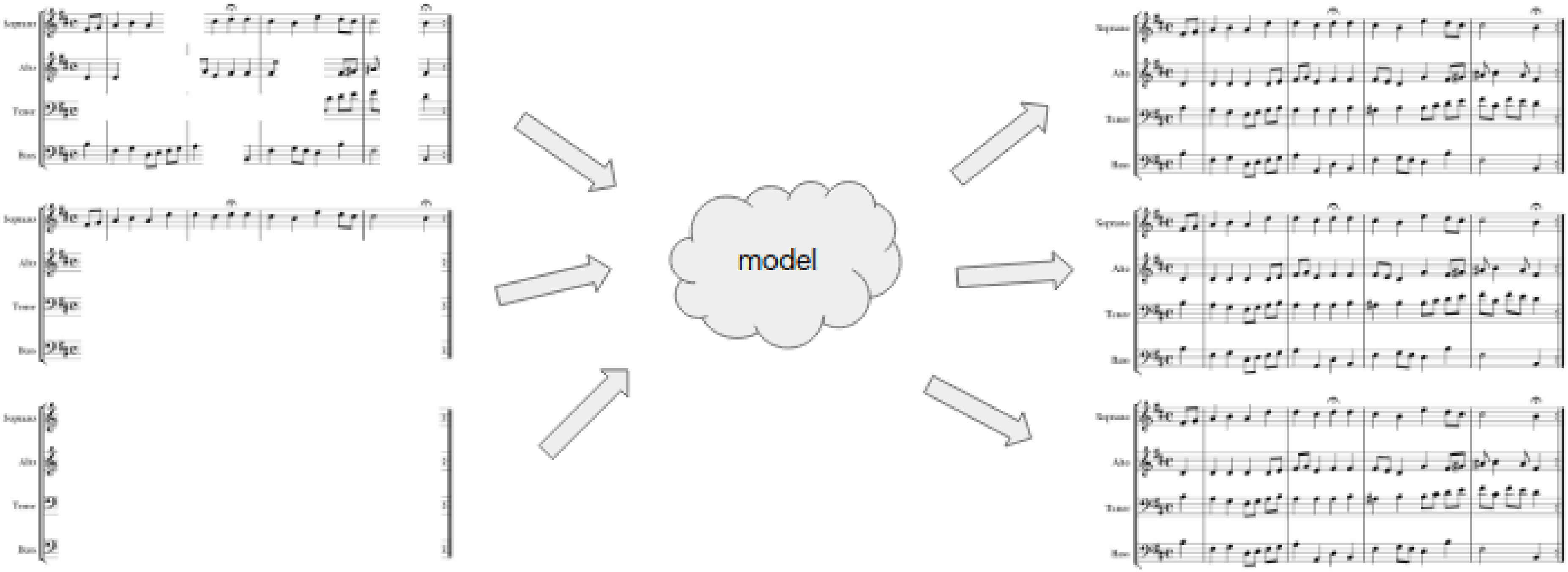
(Recap) JS Bach Doodle (2019)

youtu.be/XBfYPp6KF2g & magenta.tensorflow.org/coconet

doodles.google/doodle/celebrating-johann-sebastian-bach/

(Recap) Coconet (Huang et al., 2017)

- Based on Orderless NADE (Uria et al, 2014)



(Source: Huang et al., 2019)

Benigno Uria, Iain Murray, and Hugo Larochelle, “A Deep and Tractable Density Estimator,” *ICML*, 2014.
Cheng-Zhi Anna Huang, Tim Cooijmans, Adam Roberts, Aaron Courville, and Douglas Eck, “Counterpoint by Convolution,” *ISMIR*, 2017.
Cheng-Zhi Anna Huang, Tim Cooijmans, Monica Dinculescu, Adam Roberts, and Curtis Hawthorne, ”Coconet: the ML model behind today’s Bach Doodle,” *Magenta Blog*, 2019.

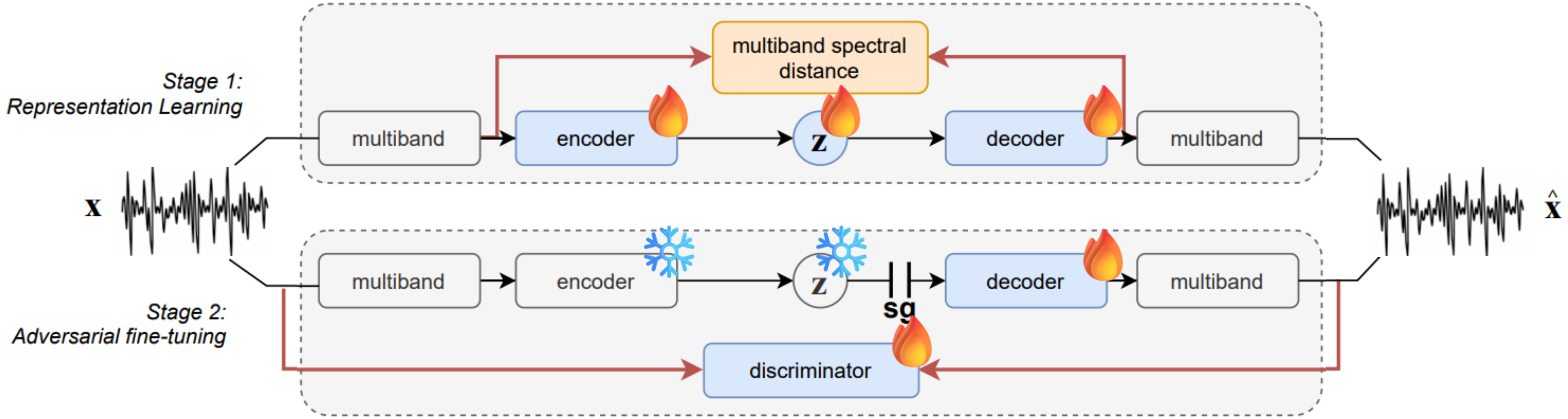
(Recap) RAVE (2022)

youtu.be/jAIRf4nGgYI

github.com/acids-ircam/RAVE

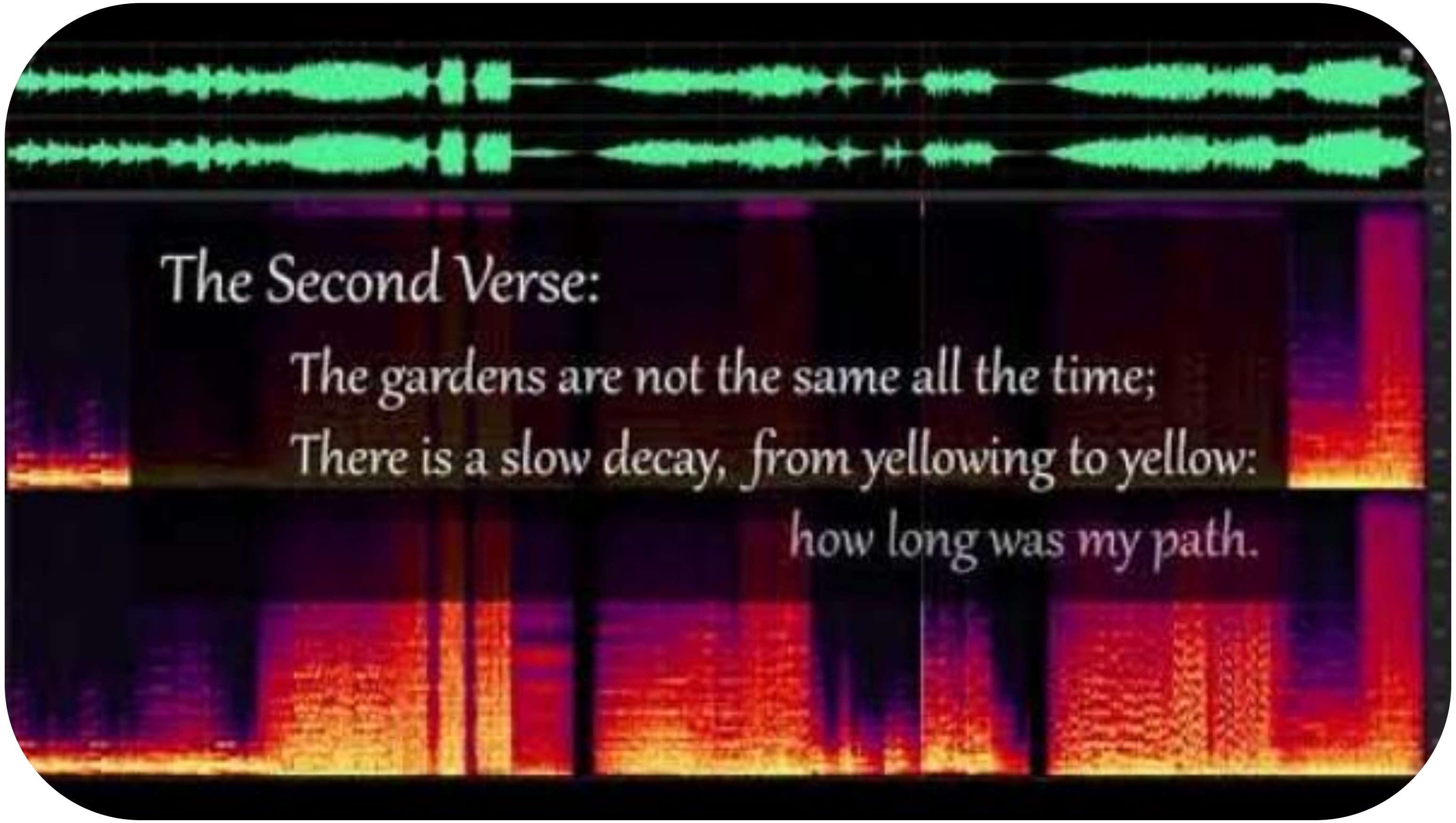
Antoine Caillon and Philippe Esling, "RAVE: A variational autoencoder for fast and high-quality neural audio synthesis," *arXiv preprint arXiv:2111.05011*, 2021.

(Recap) RAVE (2022)



Antoine Caillon and Philippe Esling, "RAVE: A variational autoencoder for fast and high-quality neural audio synthesis," *arXiv preprint arXiv:2111.05011*, 2021.

(Recap) MusicLDM (Chen et al., 2023)



The Second Verse:

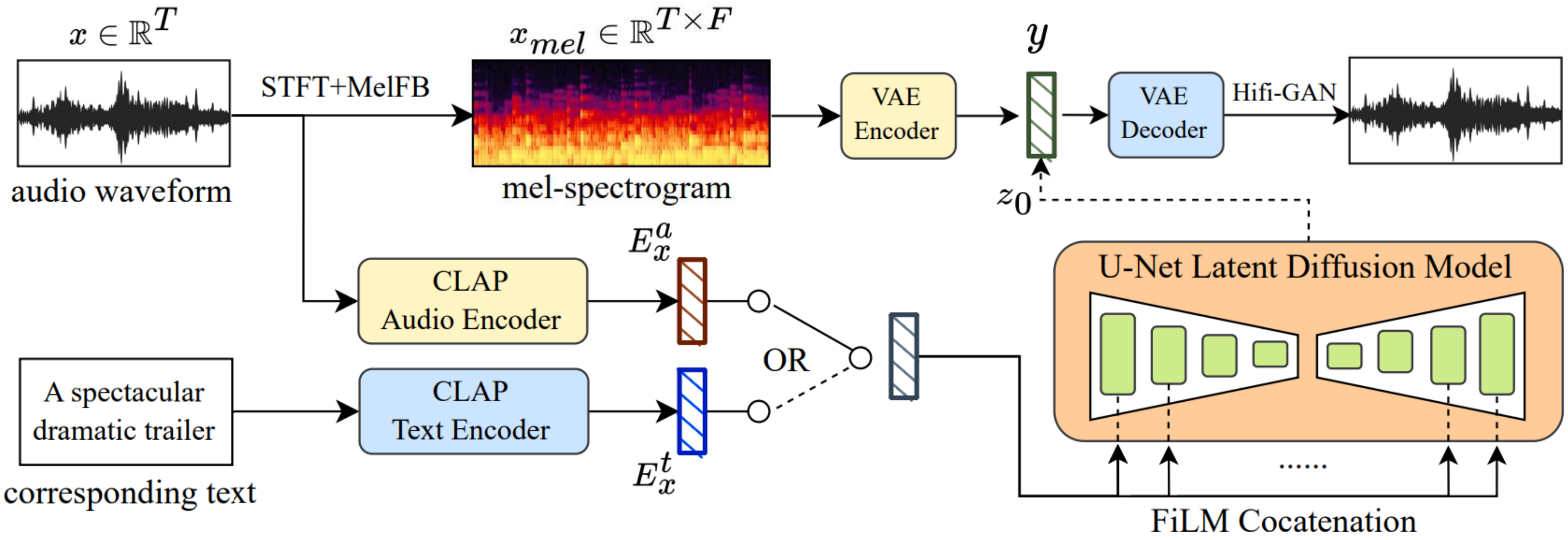
The gardens are not the same all the time;

There is a slow decay, from yellowing to yellow:

how long was my path.

youtu.be/DALv7ea6cv0

(Recap) MusicLDM (Chen et al., 2023)



(Source: Ke et al., 2023)

musicldm.github.io

Ke Chen, Yusong Wu, Haohe Liu, Marianna Nezhurina, Taylor Berg-Kirkpatrick, and Shlomo Dubnov, "MusicLDM: Enhancing Novelty in Text-to-Music Generation Using Beat-Synchronous Mixup Strategies," *ICASSP*, 2024.

What is Machine Learning?

(Recap) What is Artificial Intelligence?

> AI is the study of how to make computers **do things at which, at the moment, people are better**.
>
> – Elaine Rich and Kevin Knight, 1991

1997

(Source: Britannica)

→

2016

(Source: The Guardian)

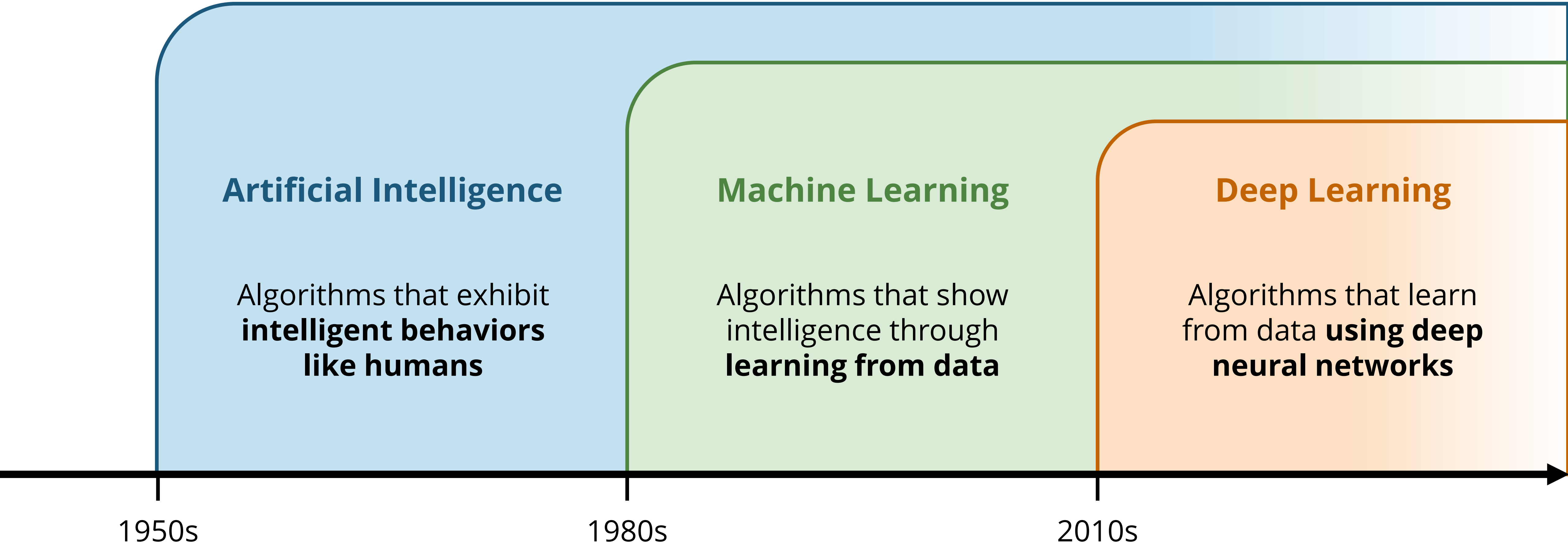
→

20??

(Source: SC2HL)

Elaine Rich and Kevin Knight, *Artificial Intelligence.* United Kingdom: McGraw-Hill, 1991.
britannica.com/topic/Deep-Blue
theguardian.com/technology/2016/mar/15/alphago-what-does-google-advanced-software-go-next
youtube.com/watch?v=PFMRDm_H9Sg

AI vs ML vs DL



Artificial Intelligence

Algorithms that exhibit **intelligent behaviors like humans**

Machine Learning

Algorithms that show intelligence through **learning from data**

Deep Learning

Algorithms that learn from data **using deep neural networks**

Building Blocks of Modern AI Systems

Data

Key difference from traditional AI

×

Model

×

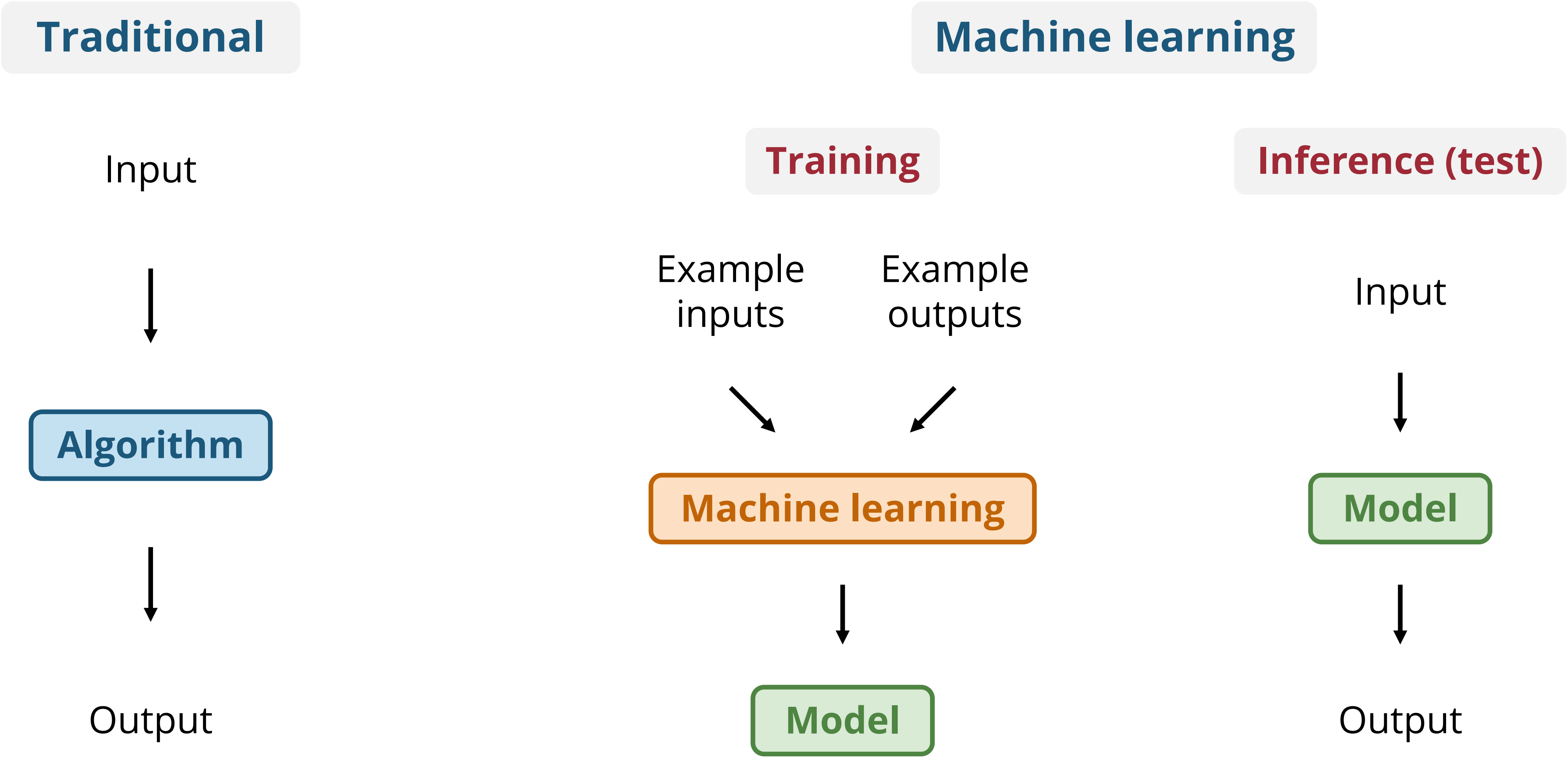
Use Case

What is Machine Learning?

> Machine Learning is a field of study that gives computers the ability to **learn without being explicitly programmed**.
>
> – Attributed to Arthur Samuel

> A computer program is said to learn from **experience E** with respect to some **class of tasks T** and **performance measure P** if ***its performance at tasks in T, as measured by P, improves with experience E***.
>
> – Tom M. Mitchell, 1997

Machine Learning



Traditional

Input

Algorithm

Output

Machine learning

Training

Example inputs

Example outputs

Machine learning

Model

Inference (test)

Input

Model

Output

Components of a Machine Learning Model

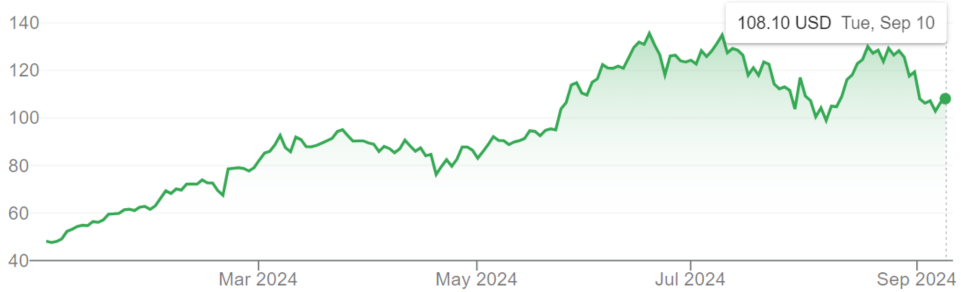
Improve on task T,
with respect to performance metric P,
based on experience E

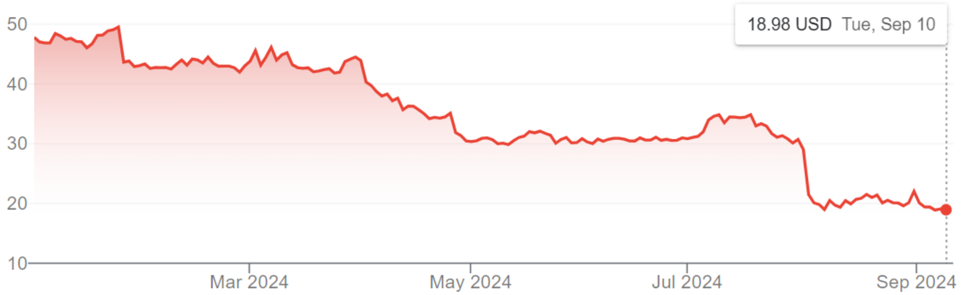
Task T	**Animal classification**
Performance metric P	Percentage of correct predictions
Experience E	Animal images with labels

Components of a Machine Learning Model

Improve on task T,
with respect to performance metric P,
based on experience E

• **Task T**	**Stock price prediction**
• **Performance metric P**	Difference between predicted and actual price
• **Experience E**	History stock price

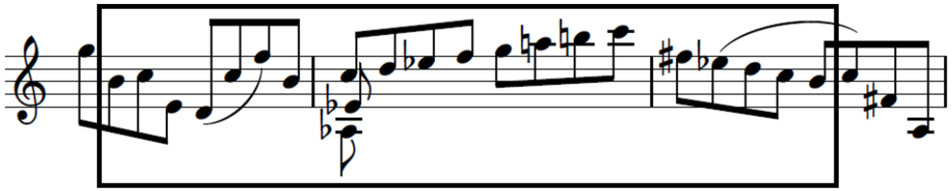




Components of a Machine Learning Model

Improve on task T,
with respect to performance metric P,
based on experience E

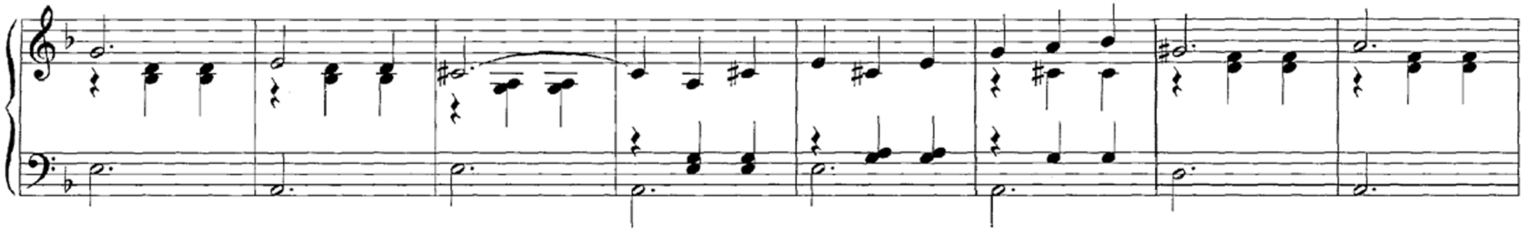
• **Task T**	**Violin transcription**
• **Performance metric P**	Percentage of correctly predicted notes
• **Experience E**	Recordings with sheet music



Components of a Machine Learning Model

Improve on task T,
with respect to performance metric P,
based on experience E

• **Task T**	**Beat tracking**
• **Performance metric P**	Average difference from actual timings
• **Experience E**	Recordings with beat timestamps



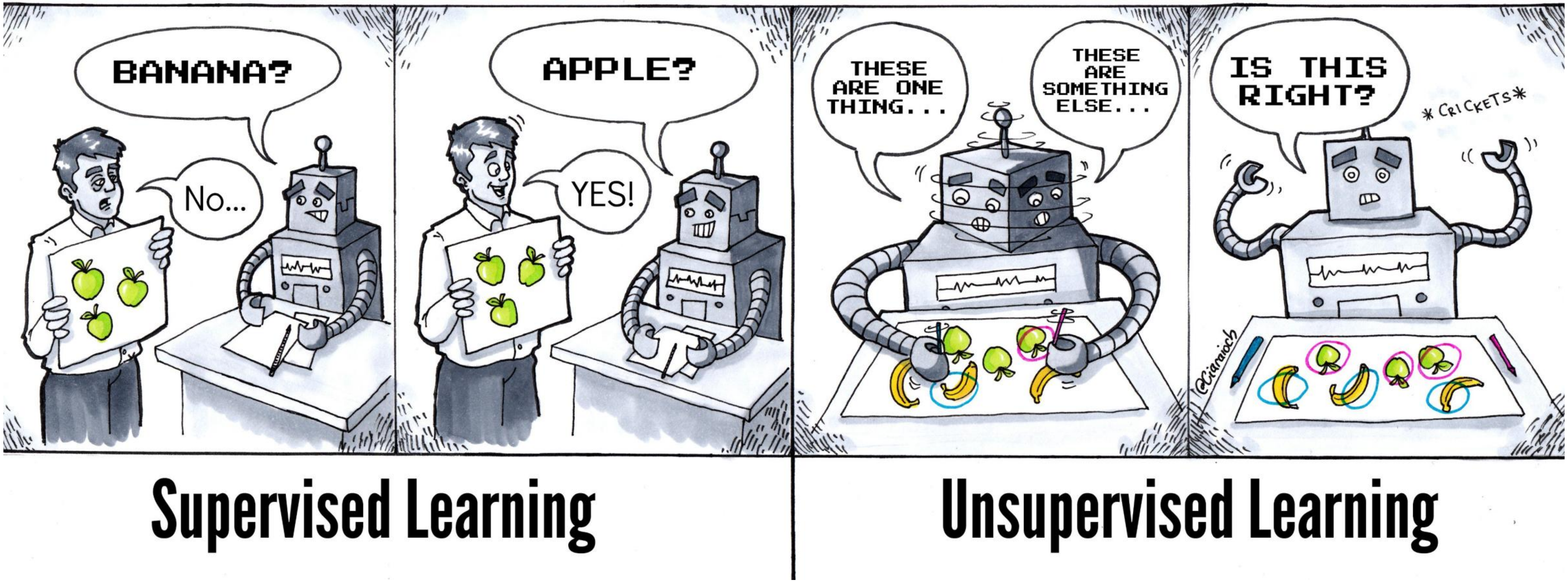
(Source: Müller)

audiolabs-erlangen.de/resources/MIR/FMP/C6/C6S3_BeatTracking.html
Meinard Müller, "Fundamentals of Music Processing – Using Python and Jupyter Notebooks," *Springer Verlag*, 2021.
Meinard Müller and Frank Zalkow, "FMP Notebooks: Educational Material for Teaching and Learning Fundamentals of Music Processing," *ISMIR*, 2019.

Input & Output of a Machine Learning Algorithm

- What is the **input** and **output** of a machine learning algorithm?
- **Input**: **Example inputs and outputs**
- **Output**: **A machine learning model**

Supervised vs Unsupervised Learning



(Source: Ciaraioch)

medium.com/@jervisaldanha/a-tale-of-two-learning-paradigms-unraveling-the-magic-of-supervised-and-unsupervised-learning-in-6acb8099c09a

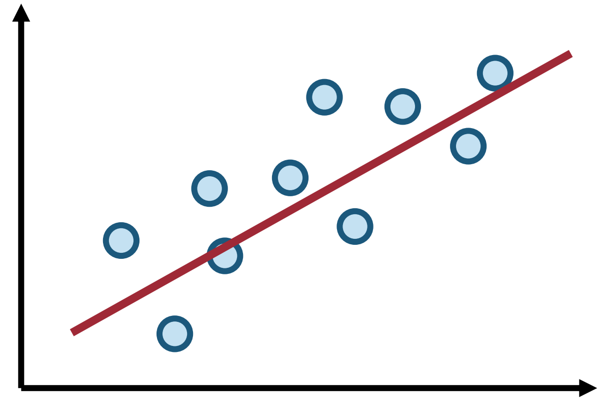
Types of Machine Learning

- **Supervised learning** Given **pairs of example inputs and outputs**
  - **Classification**: *discrete* outputs
  - **Regression**: *continuous* outputs
- **Unsupervised learning** Given ***only* example inputs**
  - **Self-supervised learning**
- **Semi-supervised learning** Given **example inputs** and **a few example outputs**
- **Reinforcement learning** Given **scalar rewards** for **a sequence of actions**

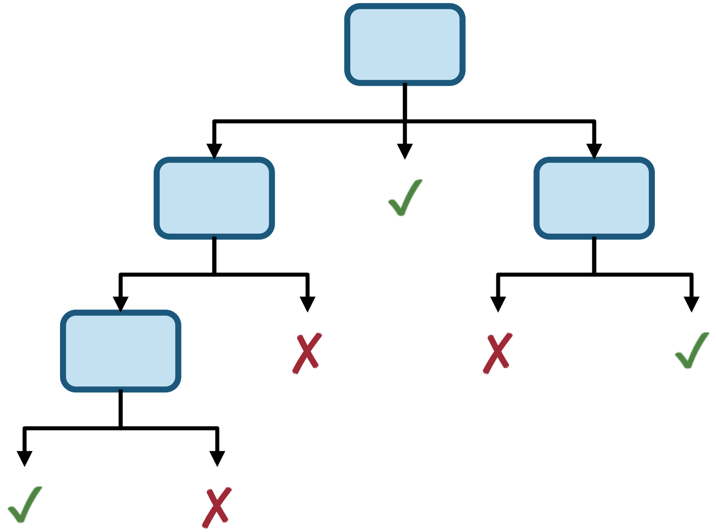
Examples of Machine Learning Algorithms

Examples of Machine Learning Algorithms

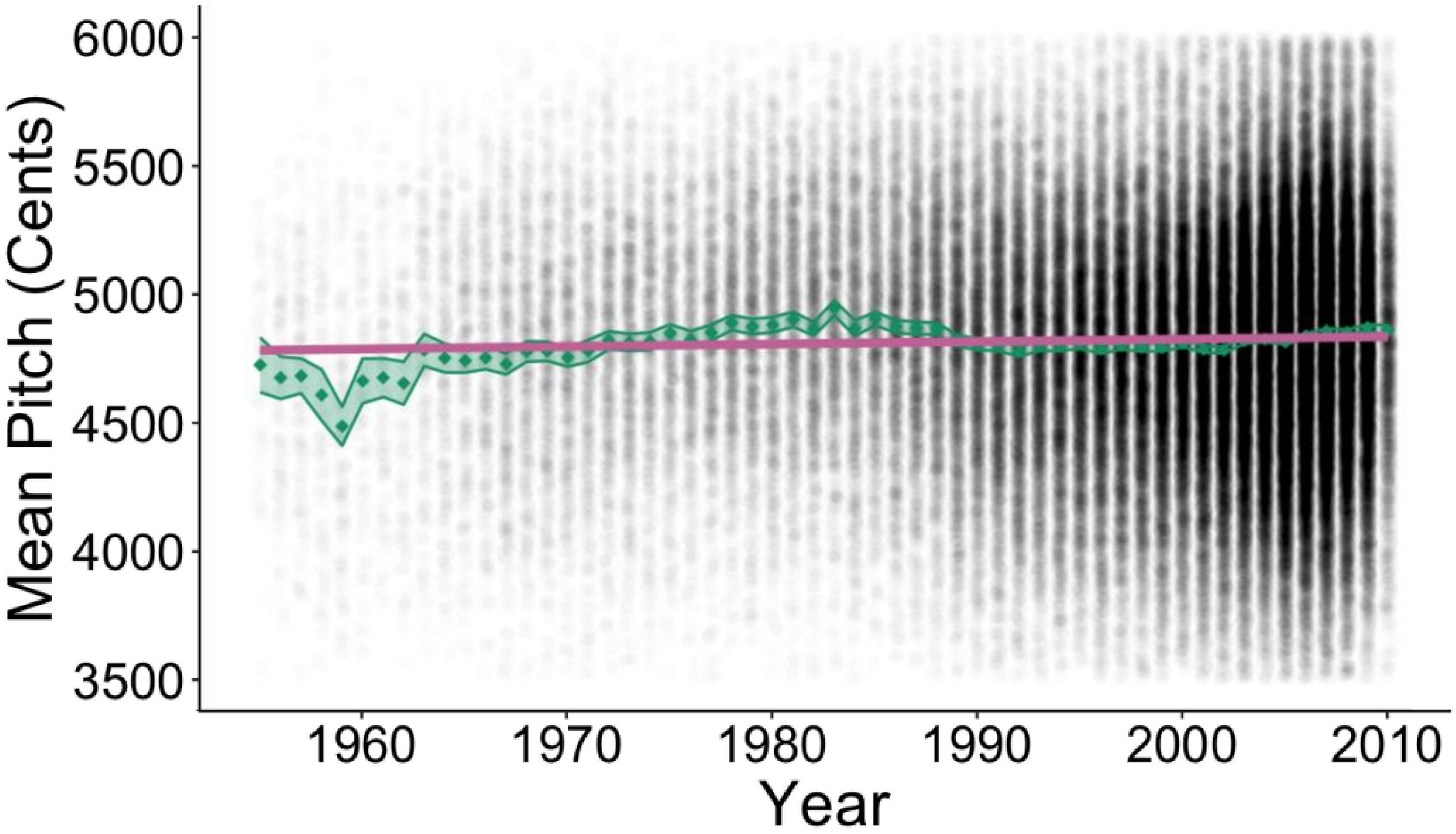
Linear regression



Decision tree



Example: Linear Regression

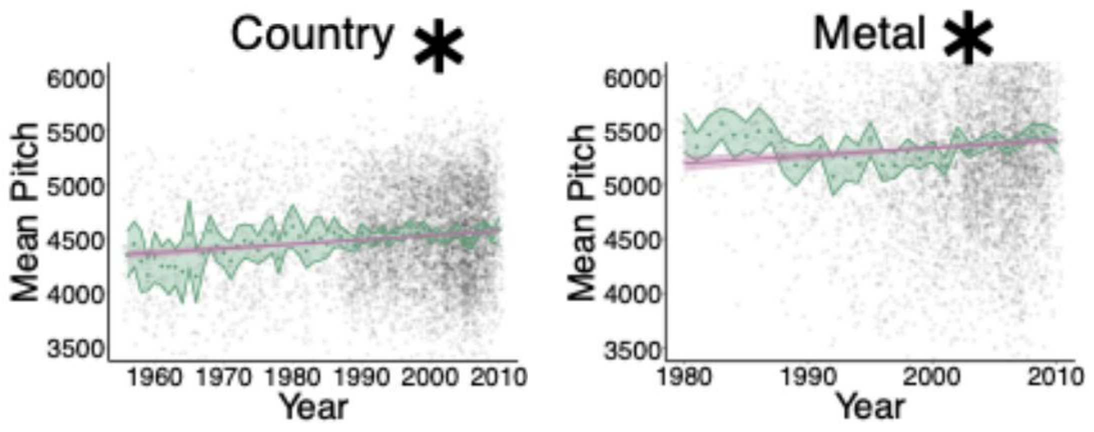


(Source: Georgieva et al., 2024)

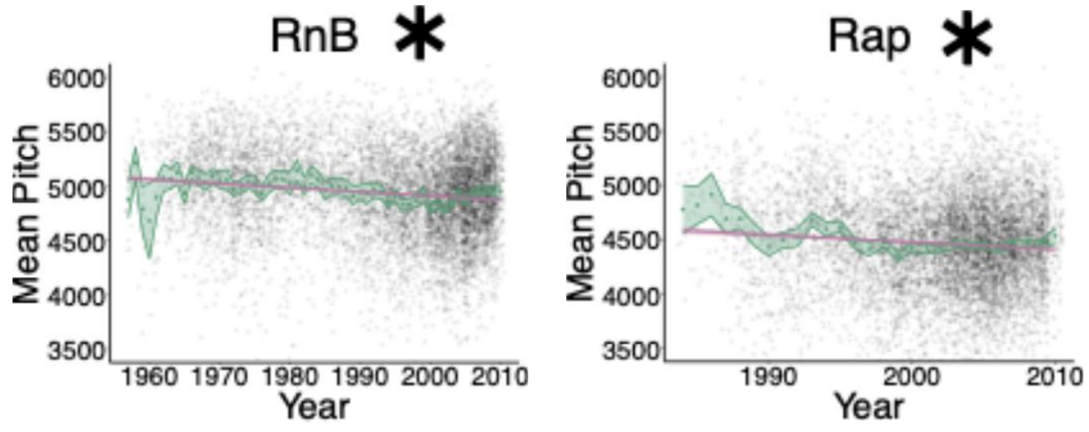
Elena Georgieva, Pablo Ripollés, and Brian McFee, "The Changing Sound of Music: An Exploratory Corpus Study of Vocal Trends Over Time," *ISMIR*, 2024.

Example: Linear Regression

Positive correlation



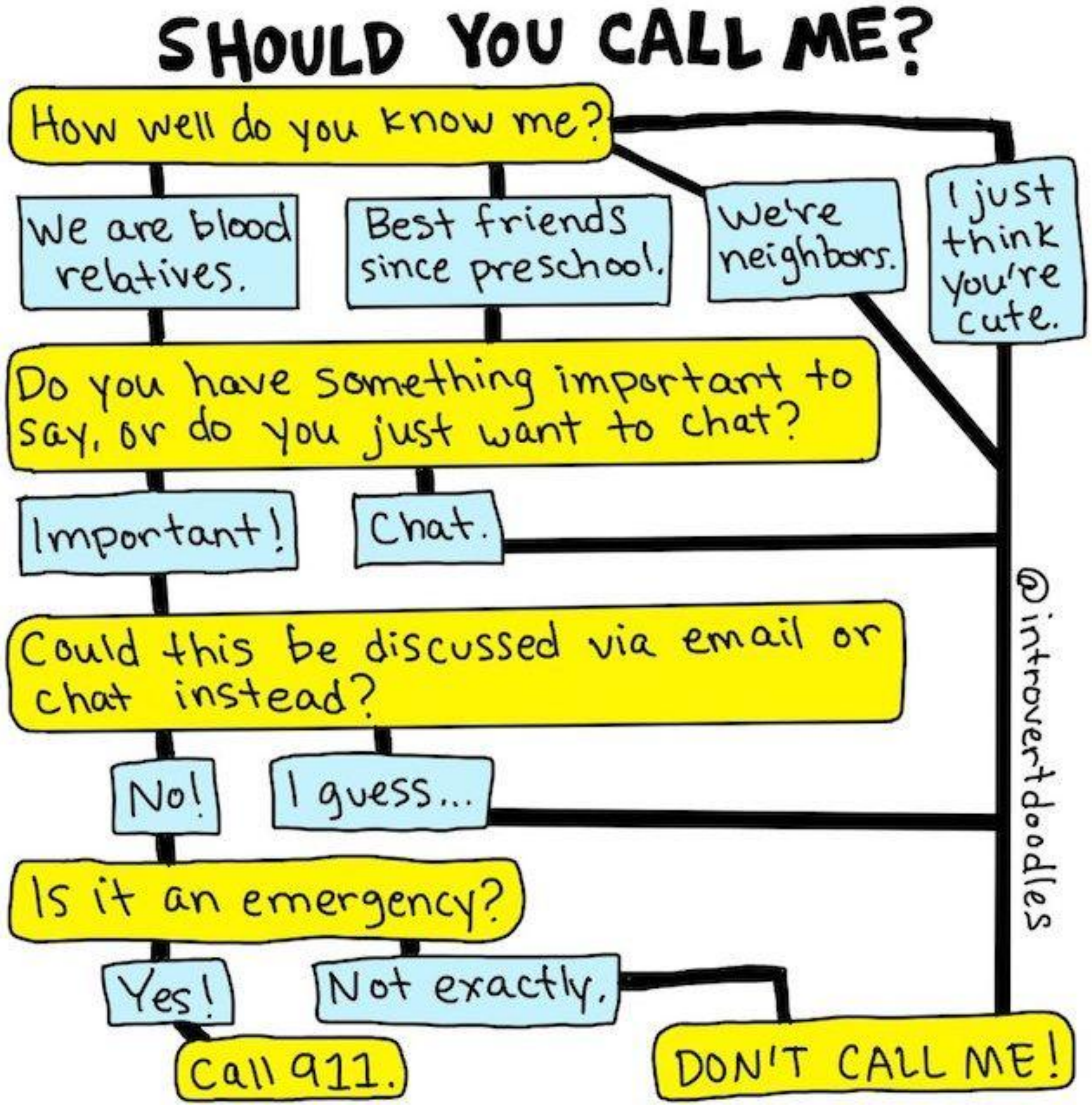
Negative correlation



(Source: Georgieva et al., 2024)

Elena Georgieva, Pablo Ripollés, and Brian McFee, "The Changing Sound of Music: An Exploratory Corpus Study of Vocal Trends Over Time," *ISMIR*, 2024.

Example: Decision Trees



(Source: Introvert Doodles)

Toy Example: Animal Classification

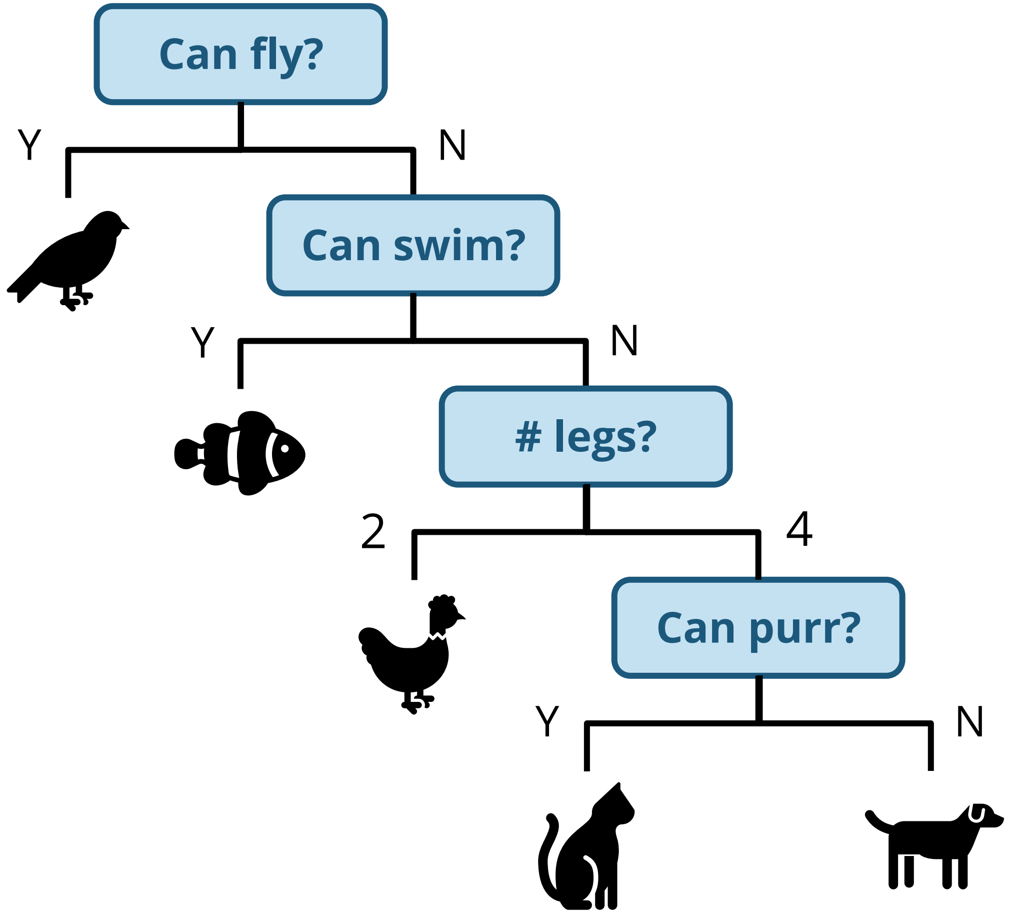
- Suppose we have the following dataset

	Can fly?	Can swim?	# of legs	Can purr?
🐓	N	N	2	N
🐈	N	N	4	Y
🐟	N	Y	0	N
🐦	Y	N	2	N
🐕	N	N	4	N

Features

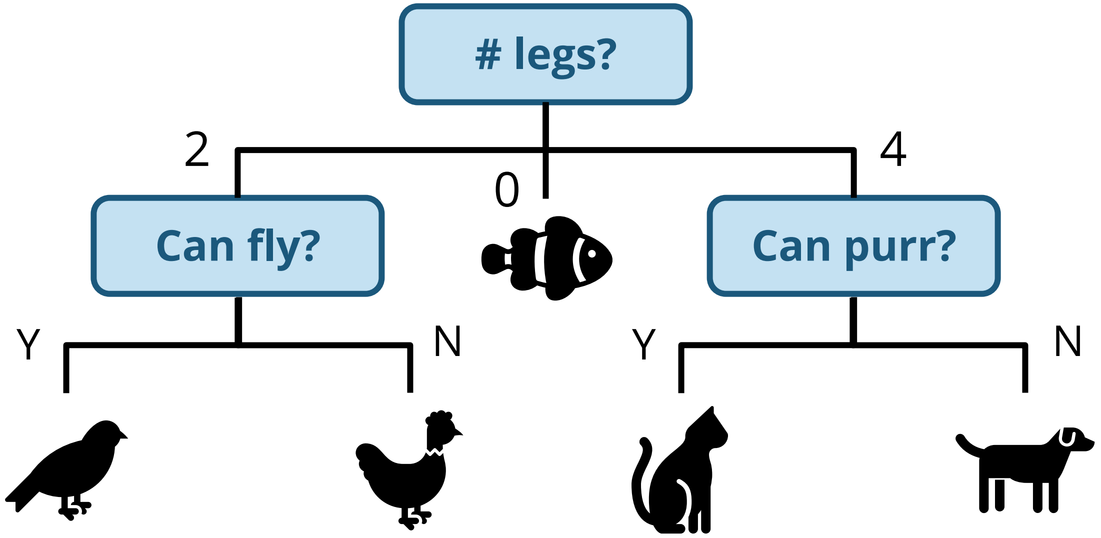
Building a Decision Tree

	Can fly?	Can swim?	# of legs	Can purr?
🐔	N	N	2	N
🐈	N	N	4	Y
🐟	N	Y	0	N
🐦	Y	N	2	N
🐕	N	N	4	N

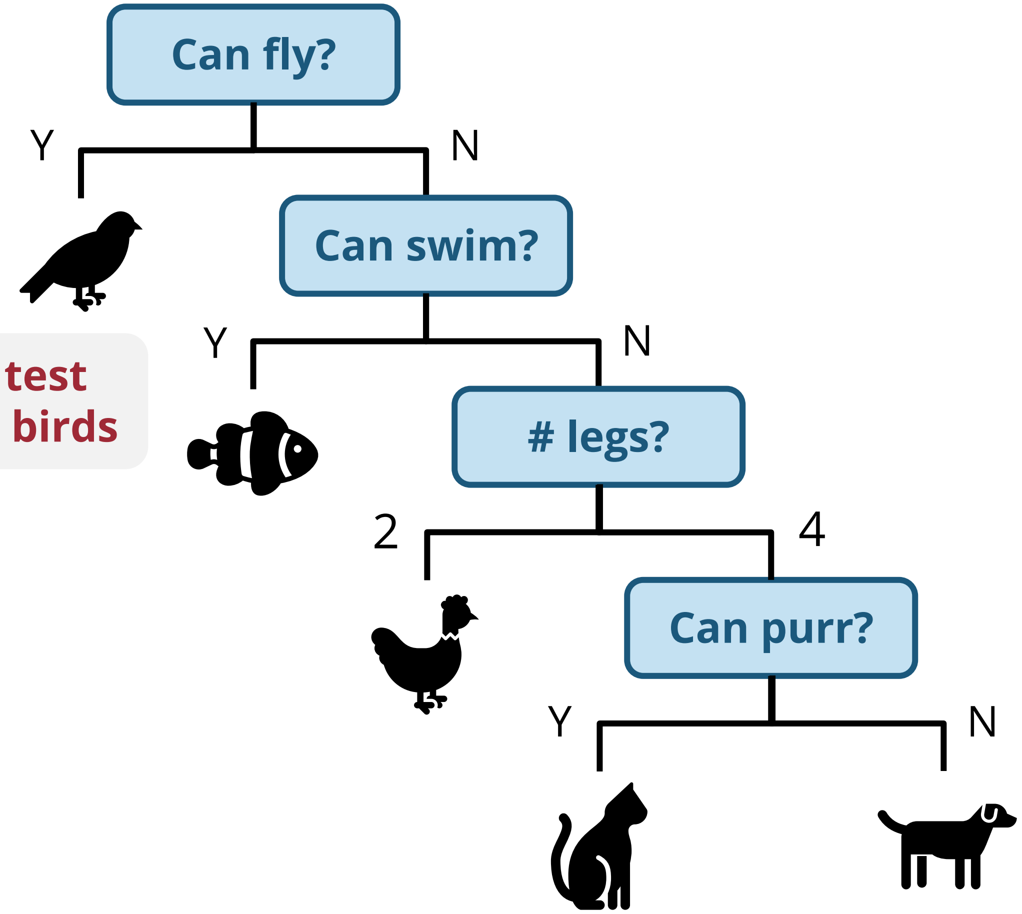


Building a Decision Tree

	Can fly?	Can swim?	# of legs	Can purr?
🐔	N	N	2	N
🐈	N	N	4	Y
🐟	N	Y	0	N
🐦	Y	N	2	N
🐕	N	N	4	N

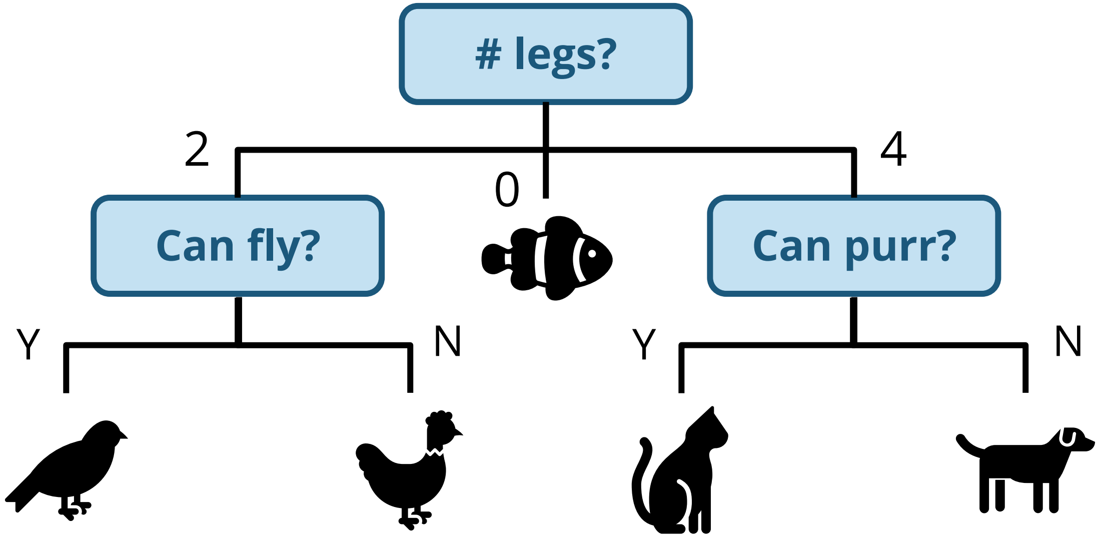


Which one is Better?



1 test
for birds

4 tests for dogs and cats



2 tests maximum for any animal type

Decision Tree!?

What is Deep Learning?

Components of a Machine Learning Model

Optimization

Defining inputs & outputs

Improve on task T,

with respect to performance metric P,

Loss function (objective function)

based on experience E

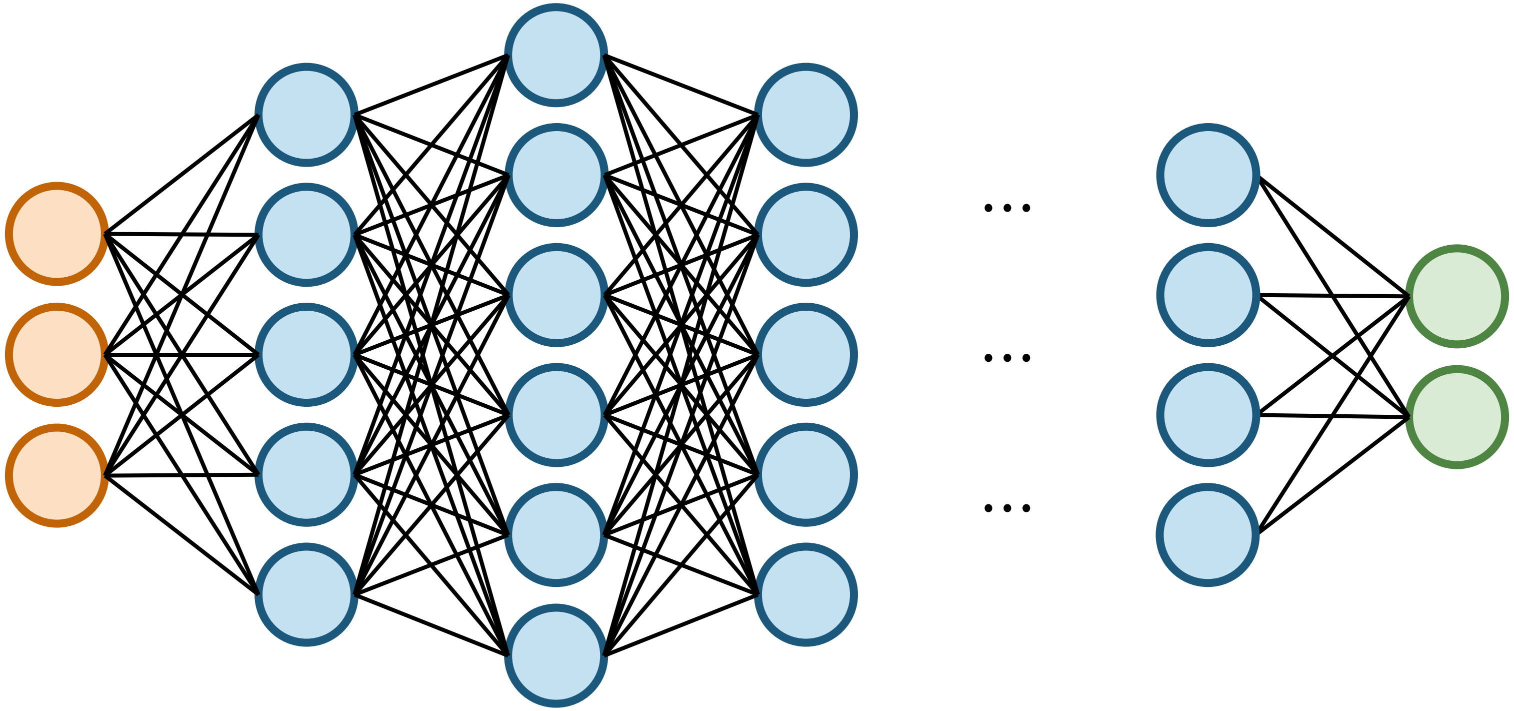
Training data

Deep learning is almost the same as machine learning by this definition!

What's special about deep learning?

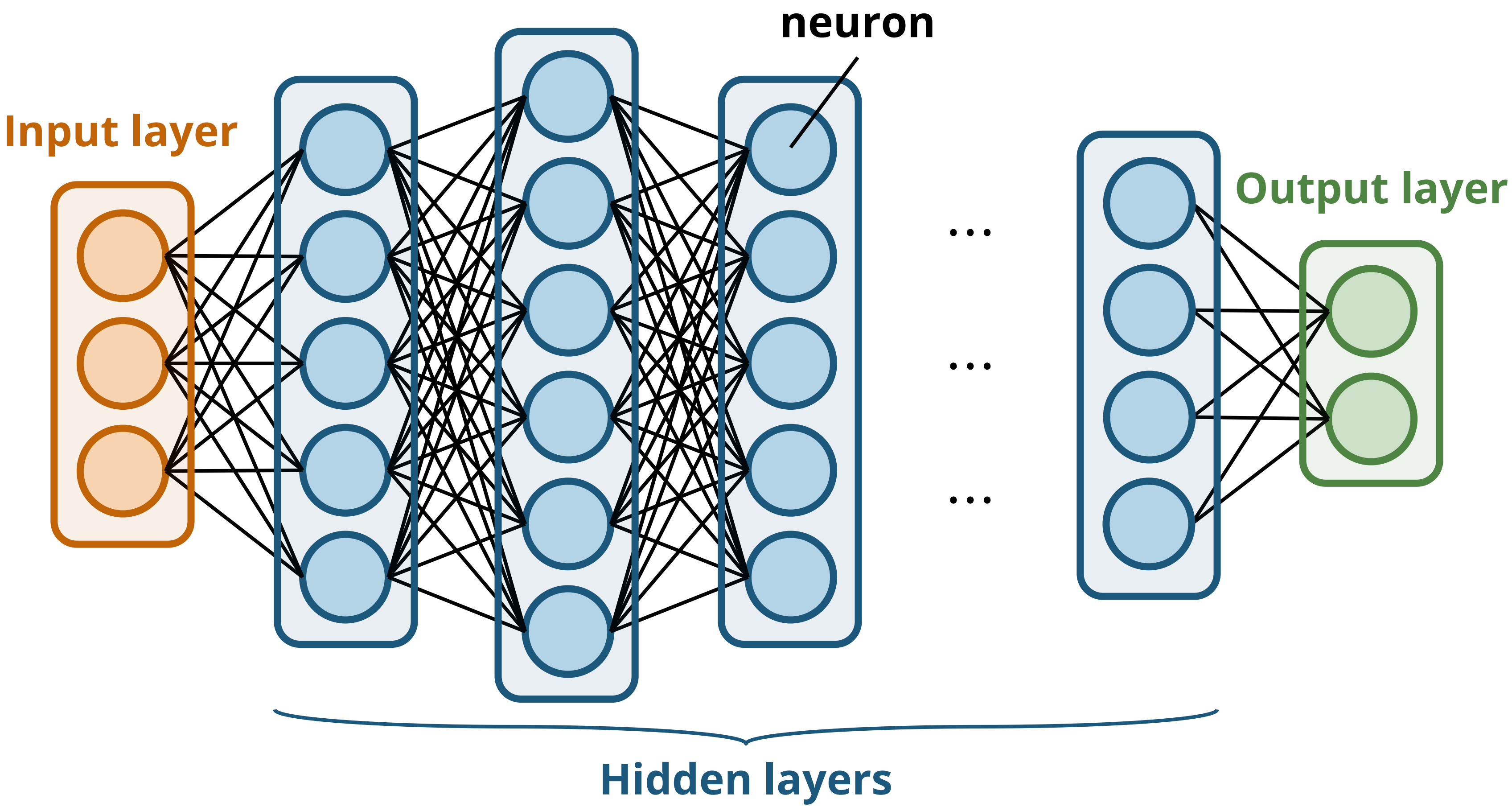
What is Deep Learning?

- A type of machine learning that uses **deep neural networks**

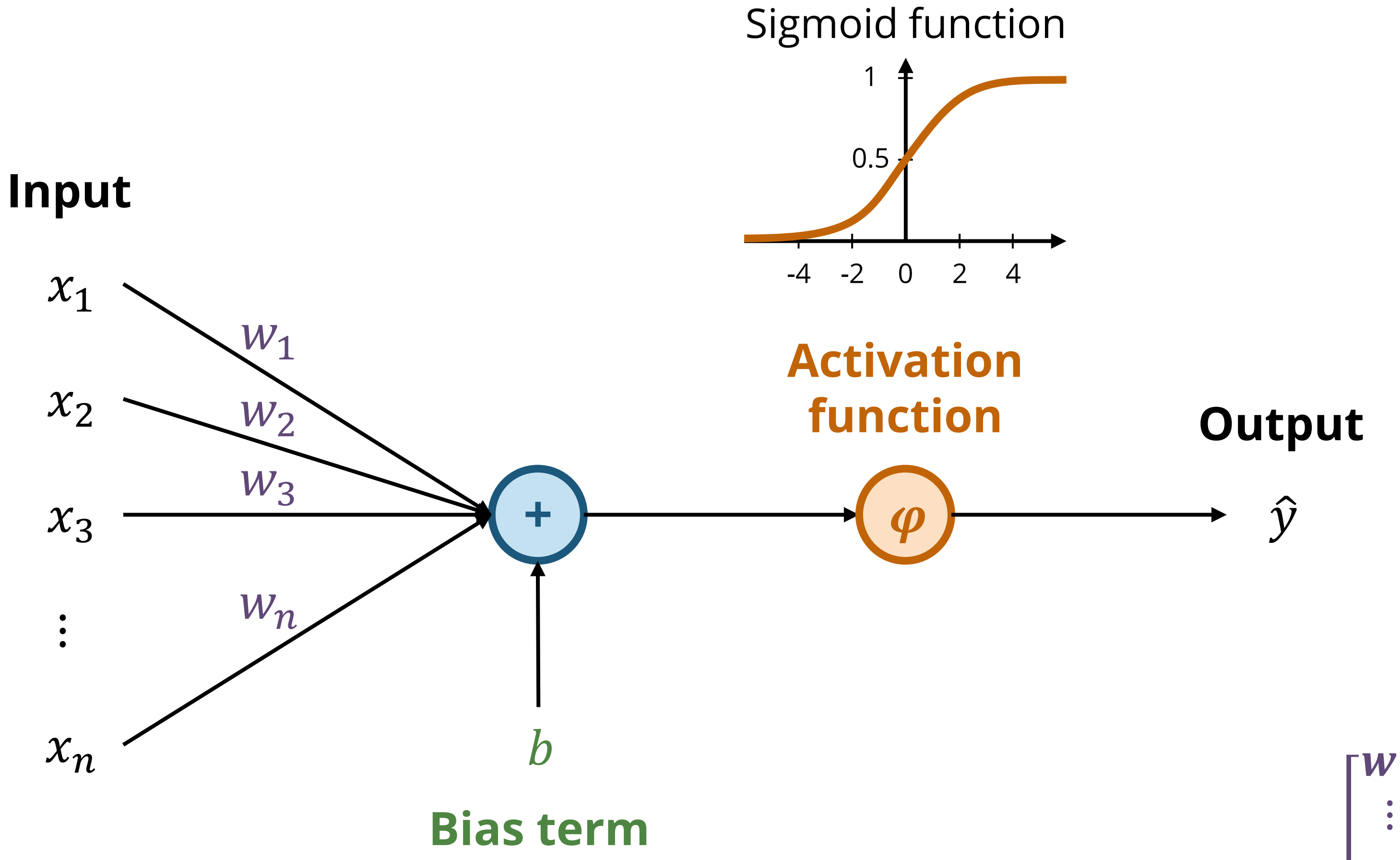


What is Deep Learning?

- A type of machine learning that uses **deep neural networks**



Inside a Neuron


Sigmoid function

Activation function

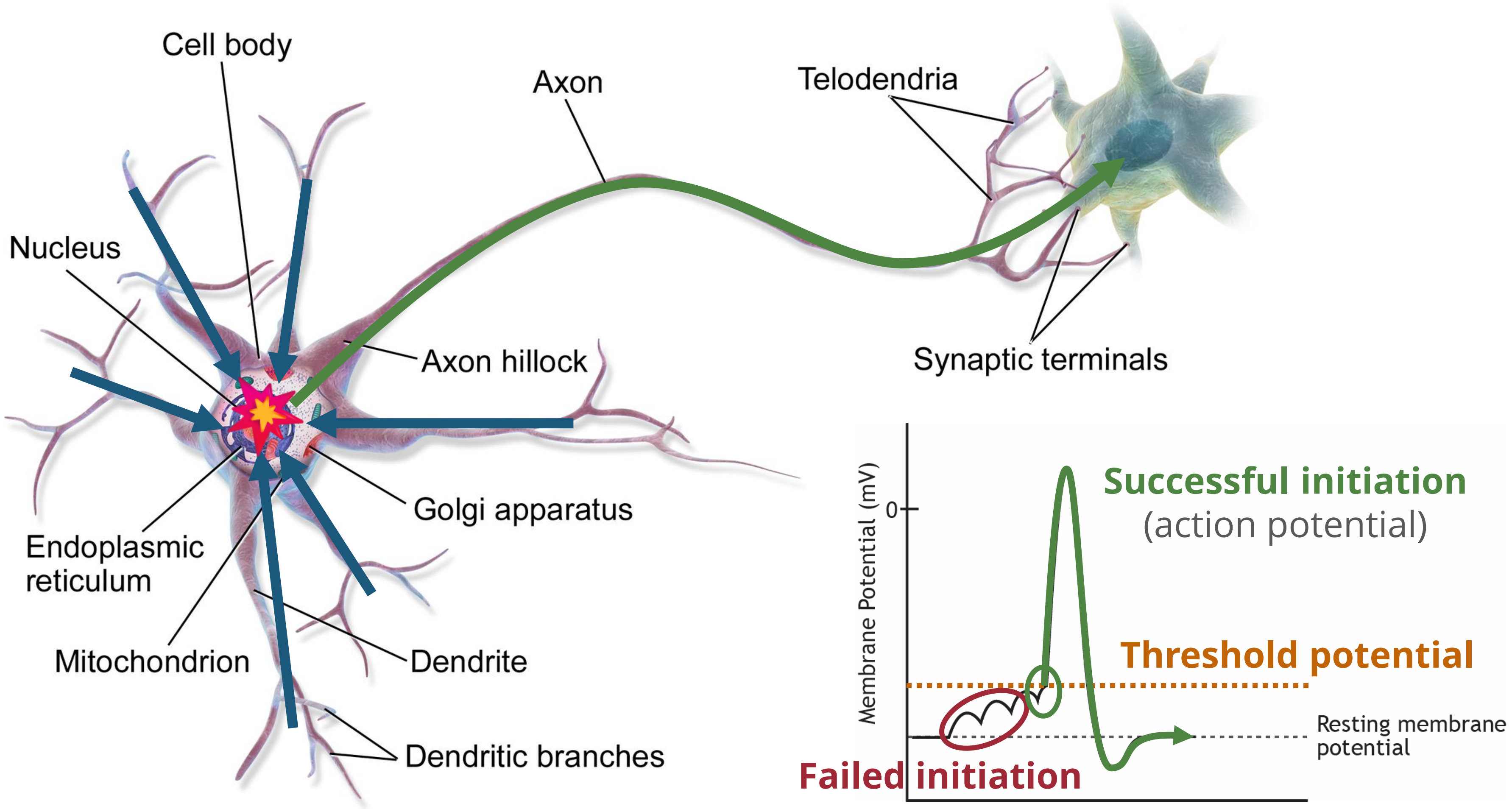
Input

Output

Bias term

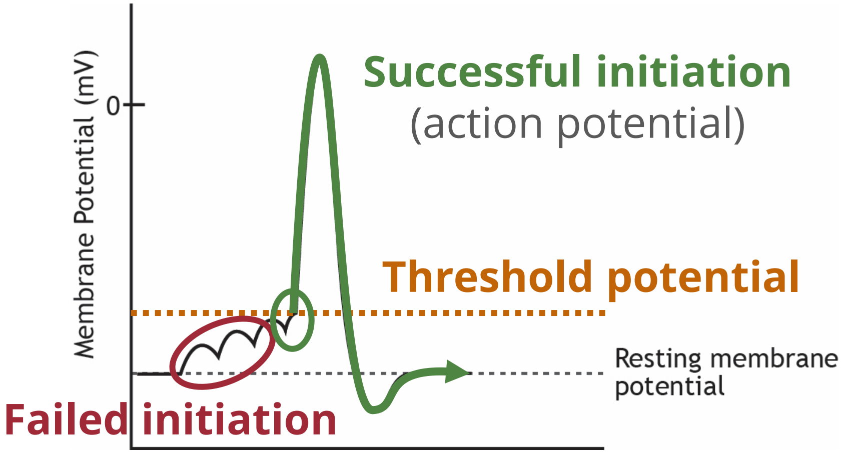
$$\hat{\boldsymbol{y}} = \varphi(w_1x_1 + w_2x_2 + \cdots + w_nx_n + b) = \varphi\left(\sum_{i=1}^{n} w_ix_i + b\right) = \varphi(\overset{\begin{bmatrix} \boldsymbol{w_1} \\ \vdots \\ \boldsymbol{w_n} \end{bmatrix}}{\mathbf{w}} \cdot \overset{\begin{bmatrix} \boldsymbol{x_1} \\ \vdots \\ \boldsymbol{x_n} \end{bmatrix}}{\mathbf{x}} + b)$$

Human Neuron

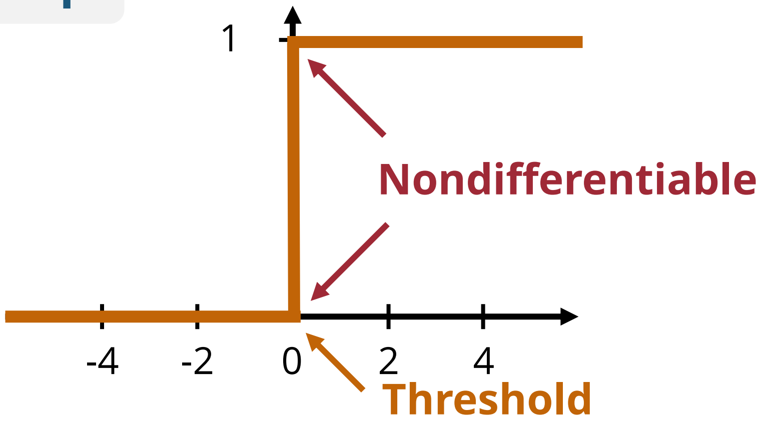


Casey Henley, Introduction to Neuroscience

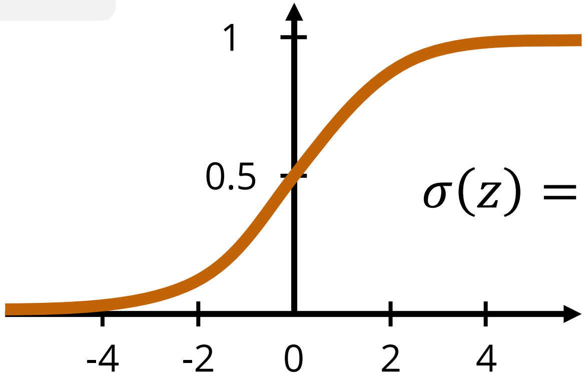
Why Sigmoid?



Unit step



Sigmoid



$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Casey Henley, Introduction to Neuroscience

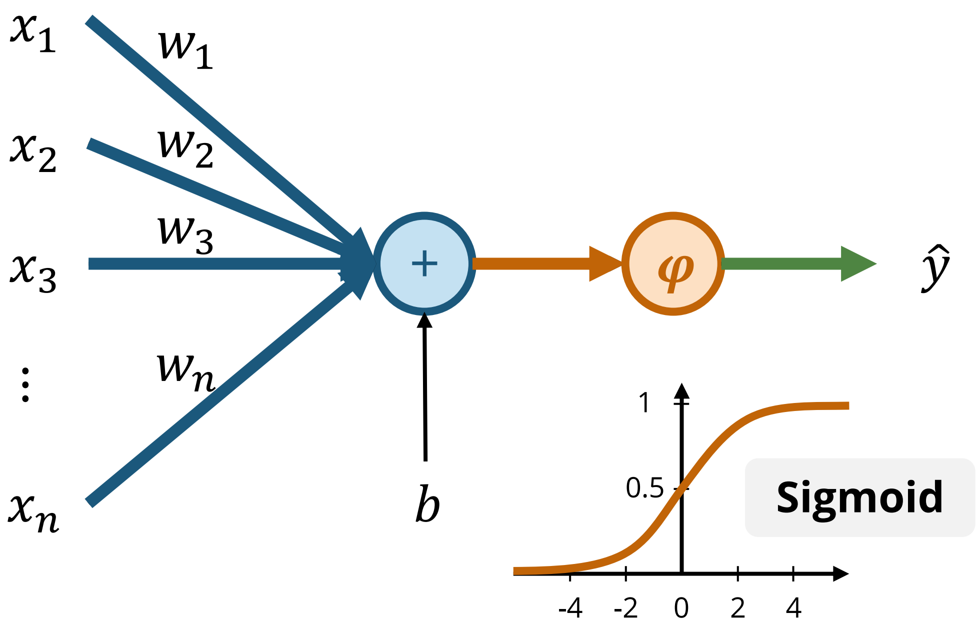
Why Bias Term?

- Allow nonzero outputs when all inputs are zero

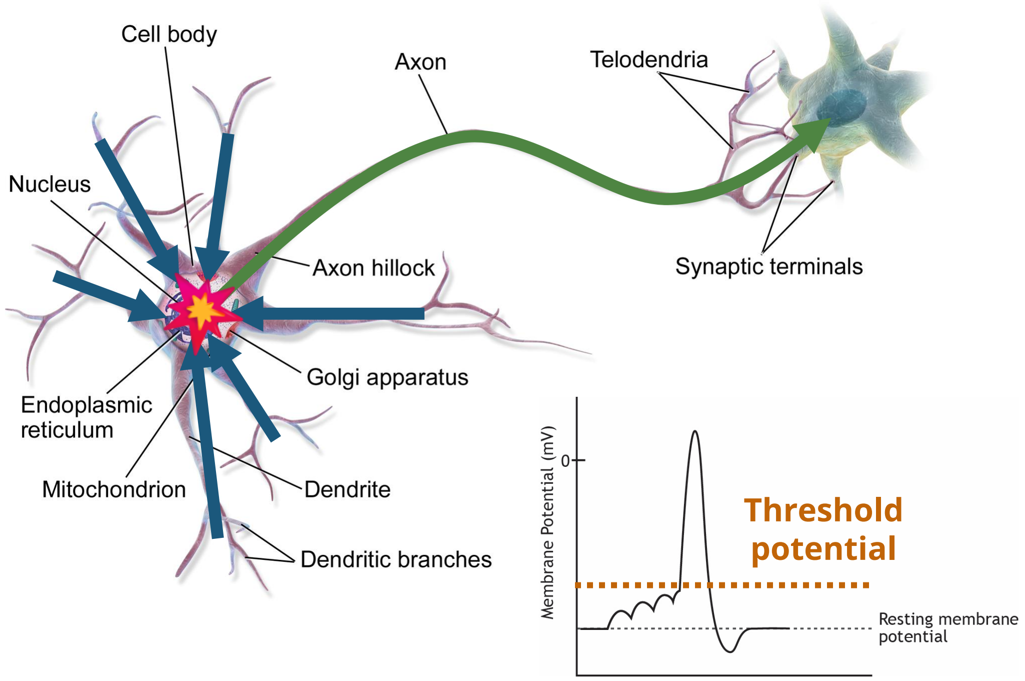
$$\hat{\boldsymbol{y}} = \boldsymbol{\varphi}(w_1 \cancelto{0}{x_1} + w_2 \cancelto{0}{x_2} + \cdots + w_n \cancelto{0}{x_n} + b) = \boldsymbol{\varphi}(b)$$

Artificial vs Human Neuron

Artificial neuron



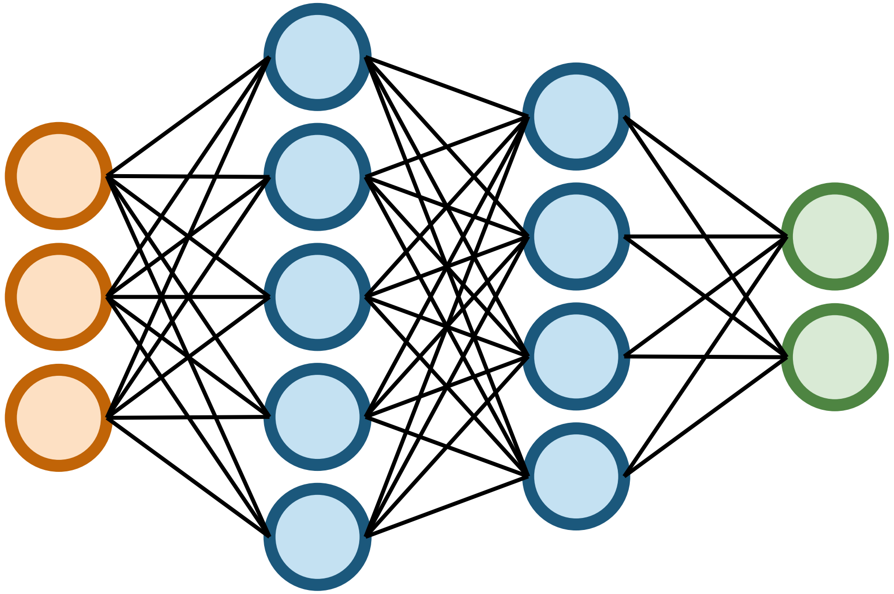
Human neuron

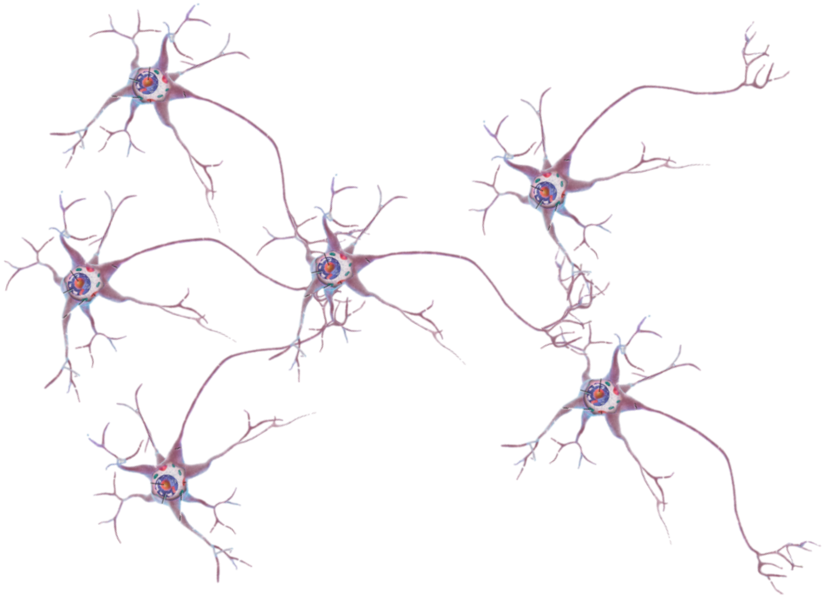


Casey Henley, Introduction to Neuroscience

Artificial Neural Networks

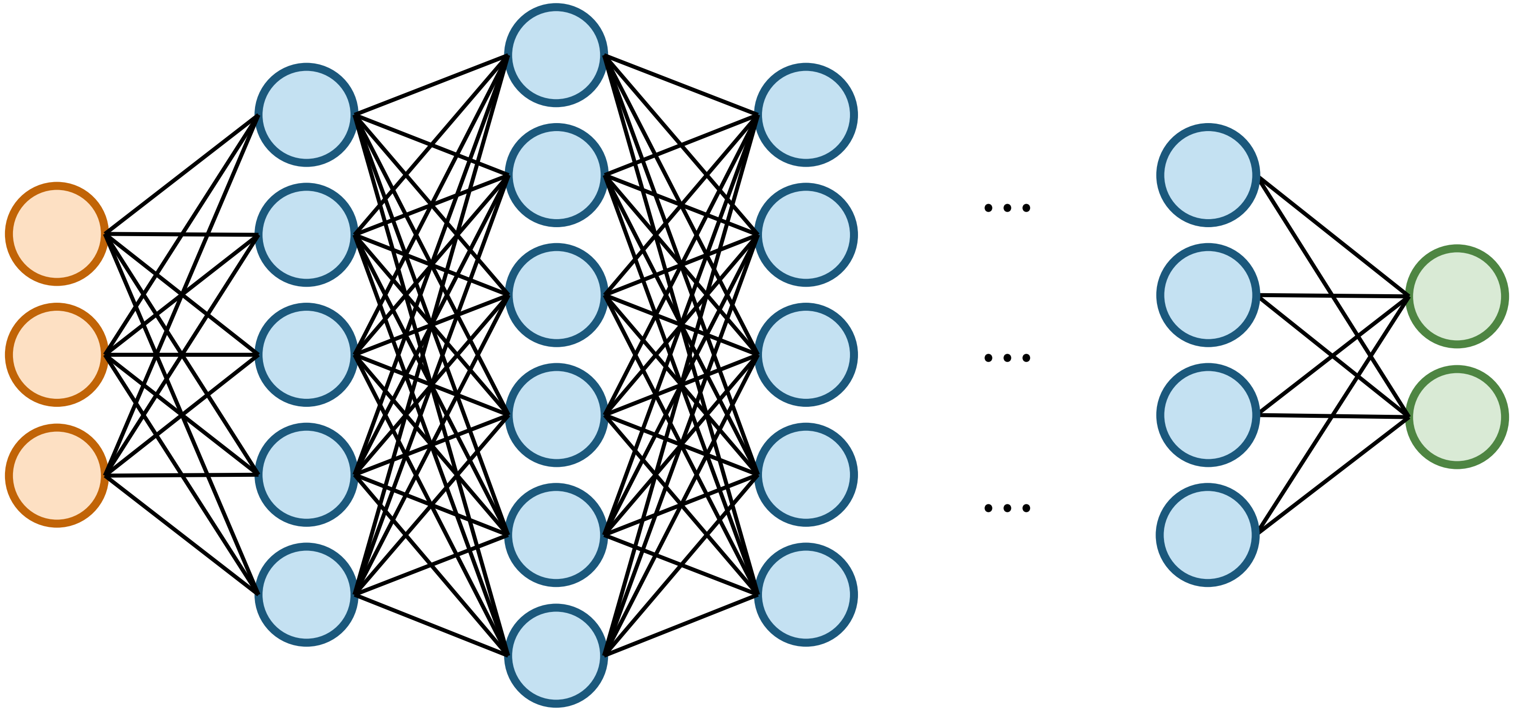
- Although inspired by human neural networks, artificial neural networks nowadays *do not work like human brains*
  - Lacking **functional hierarchy**, **high-level feedback loops**, **memory module**, etc.
  - Human brains work more like **spiking neural networks** → Efficiency!





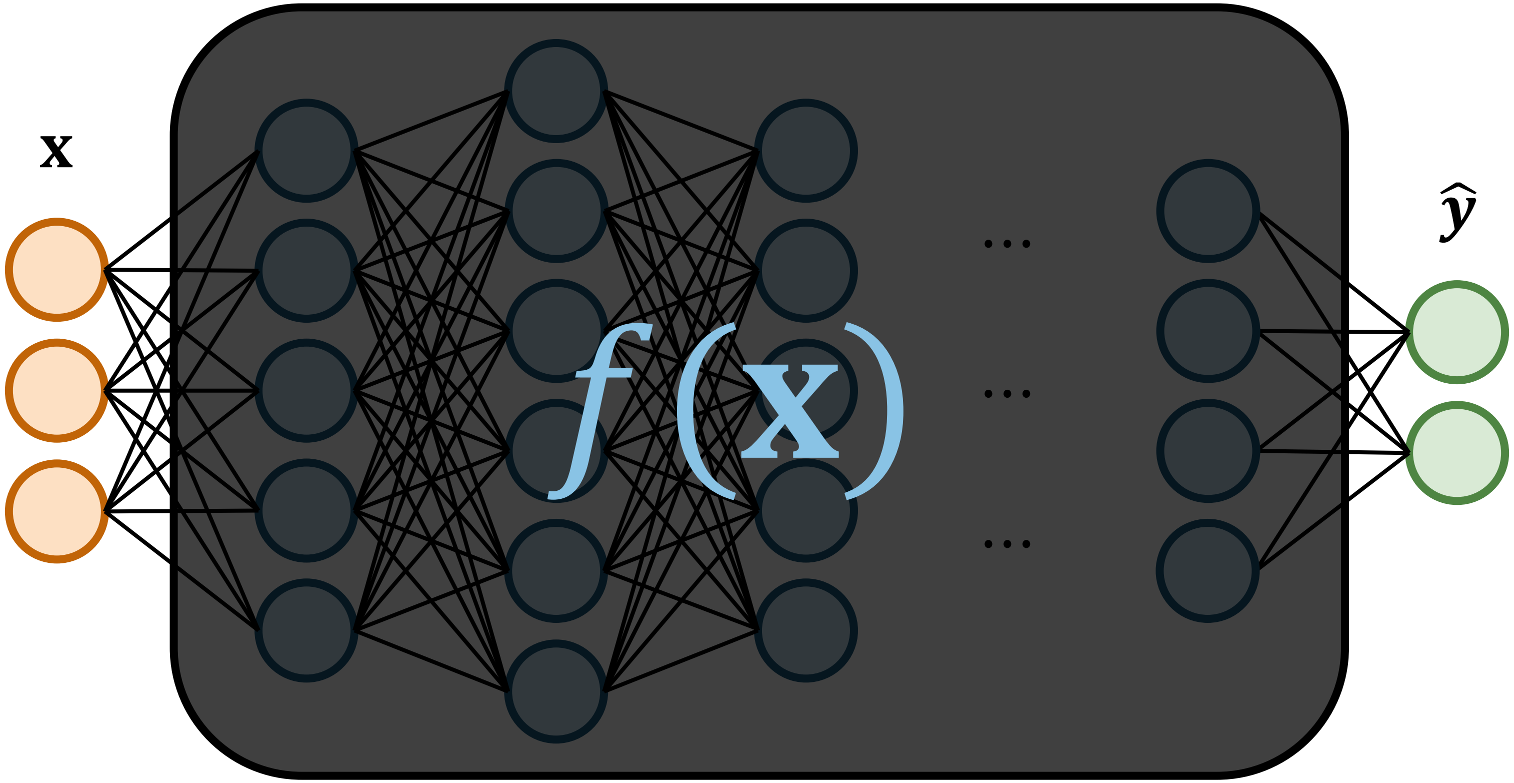
Fully Connected Feedforward Network

- Most basic form of deep neural networks



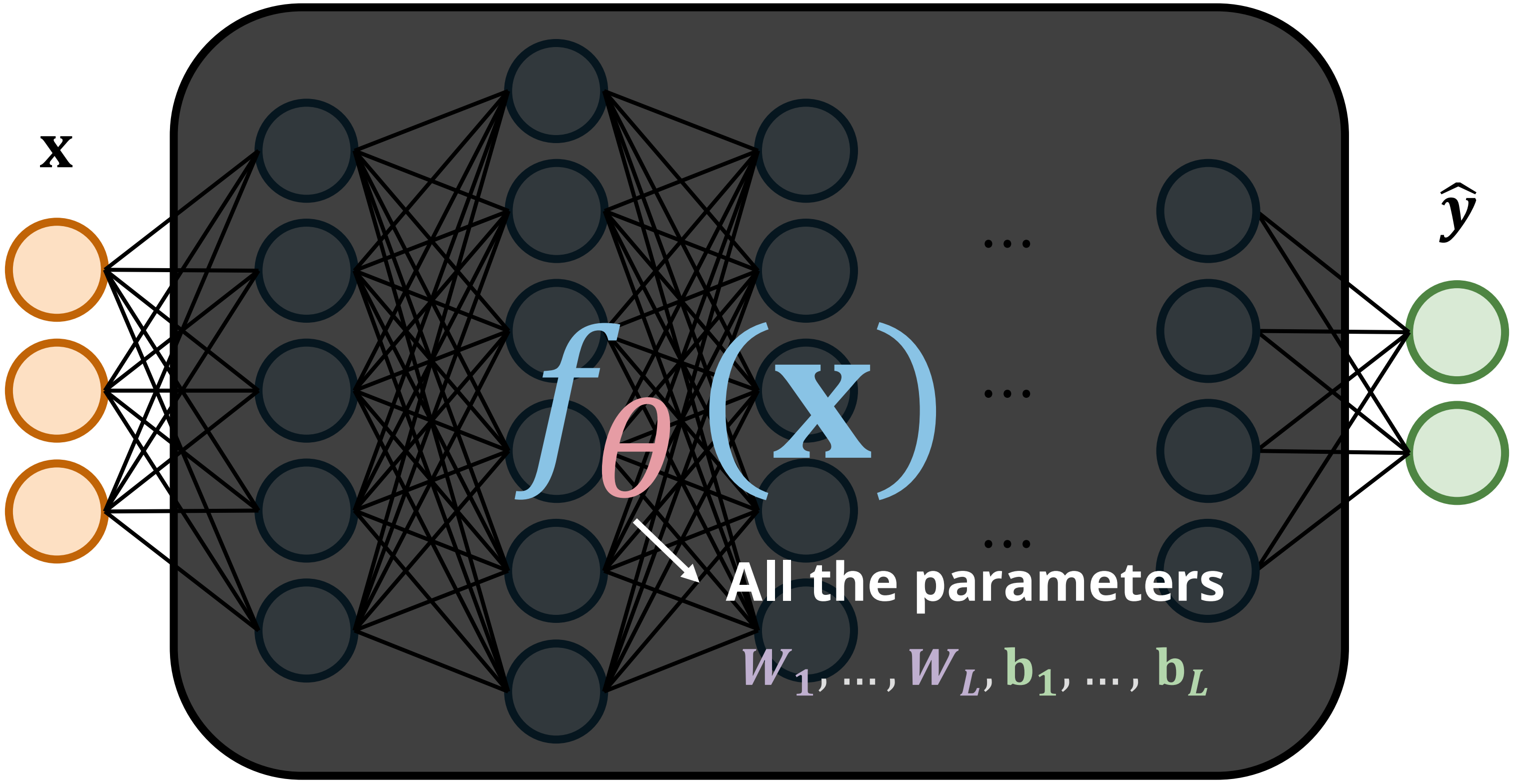
Neural Networks are Parameterized Functions

- A neural network represents **a set of functions**



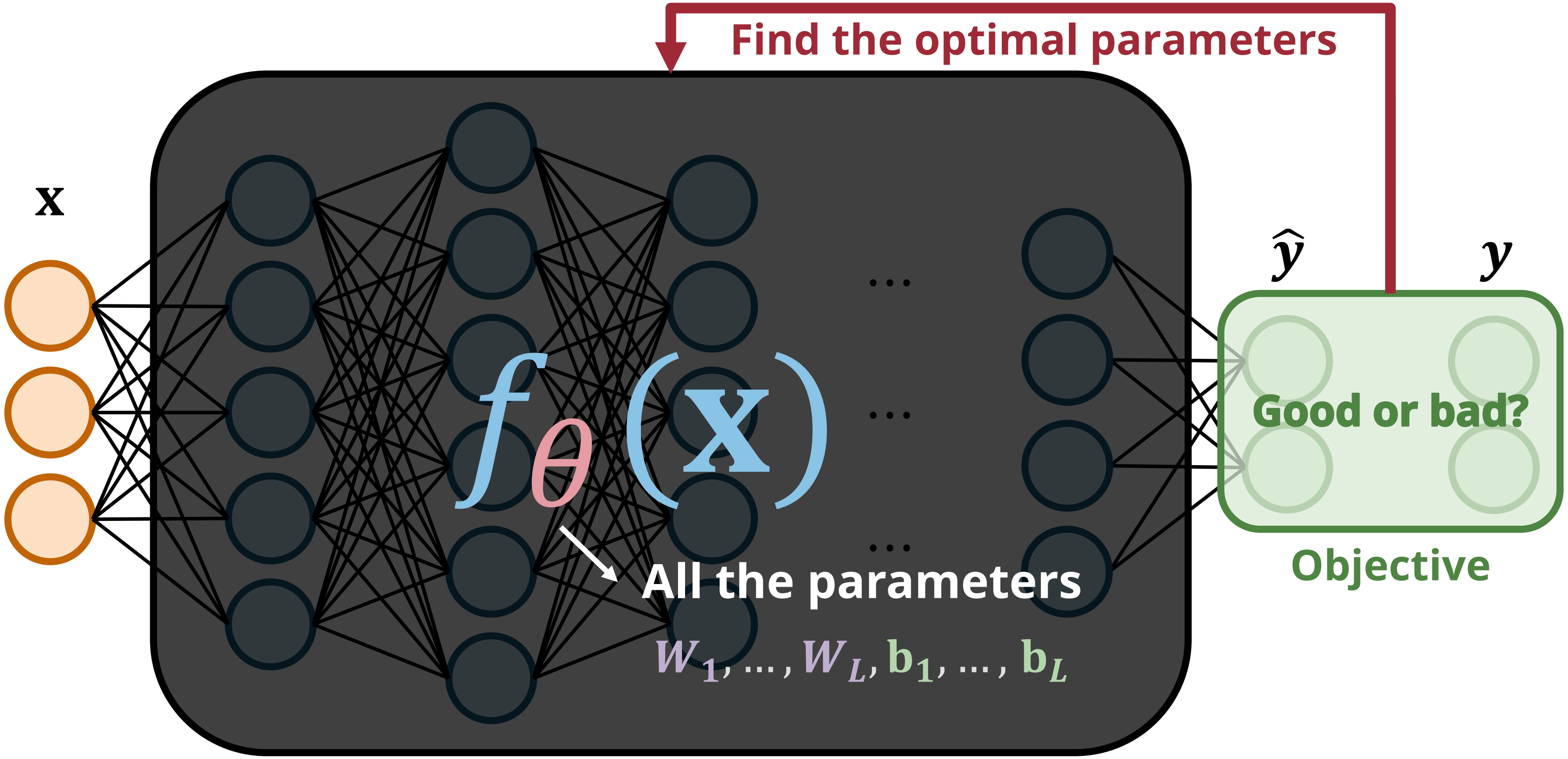
Neural Networks are Parameterized Functions

- A neural network represents **a set of functions**

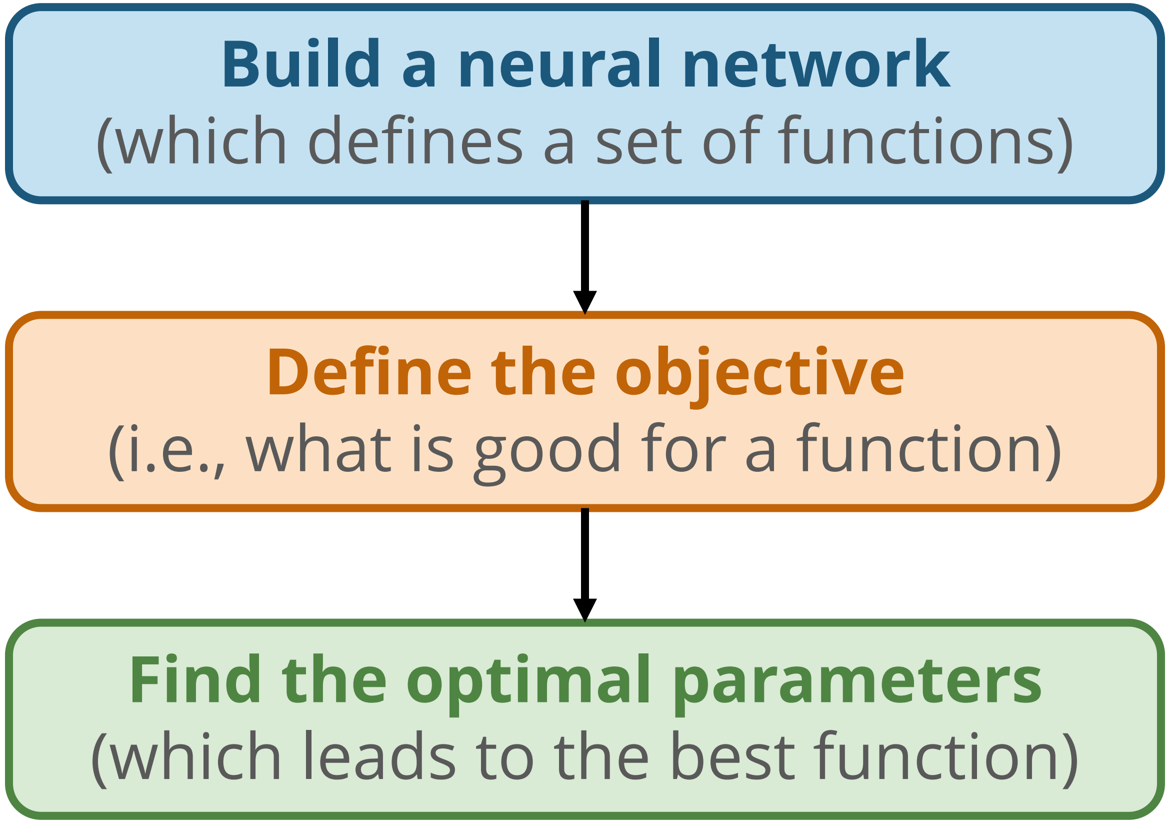


Neural Networks are Parameterized Functions

- A neural network represents **a set of functions**



Training a Neural Network



Build a neural network
(which defines a set of functions)

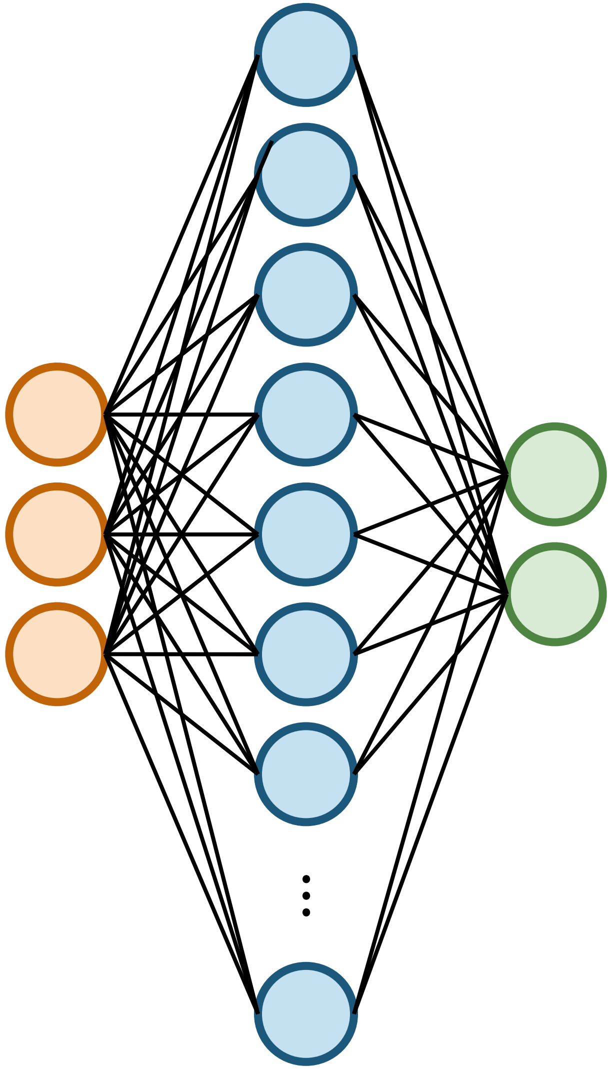
Define the objective
(i.e., what is good for a function)

Find the optimal parameters
(which leads to the best function)

Expressiveness of Neural Networks

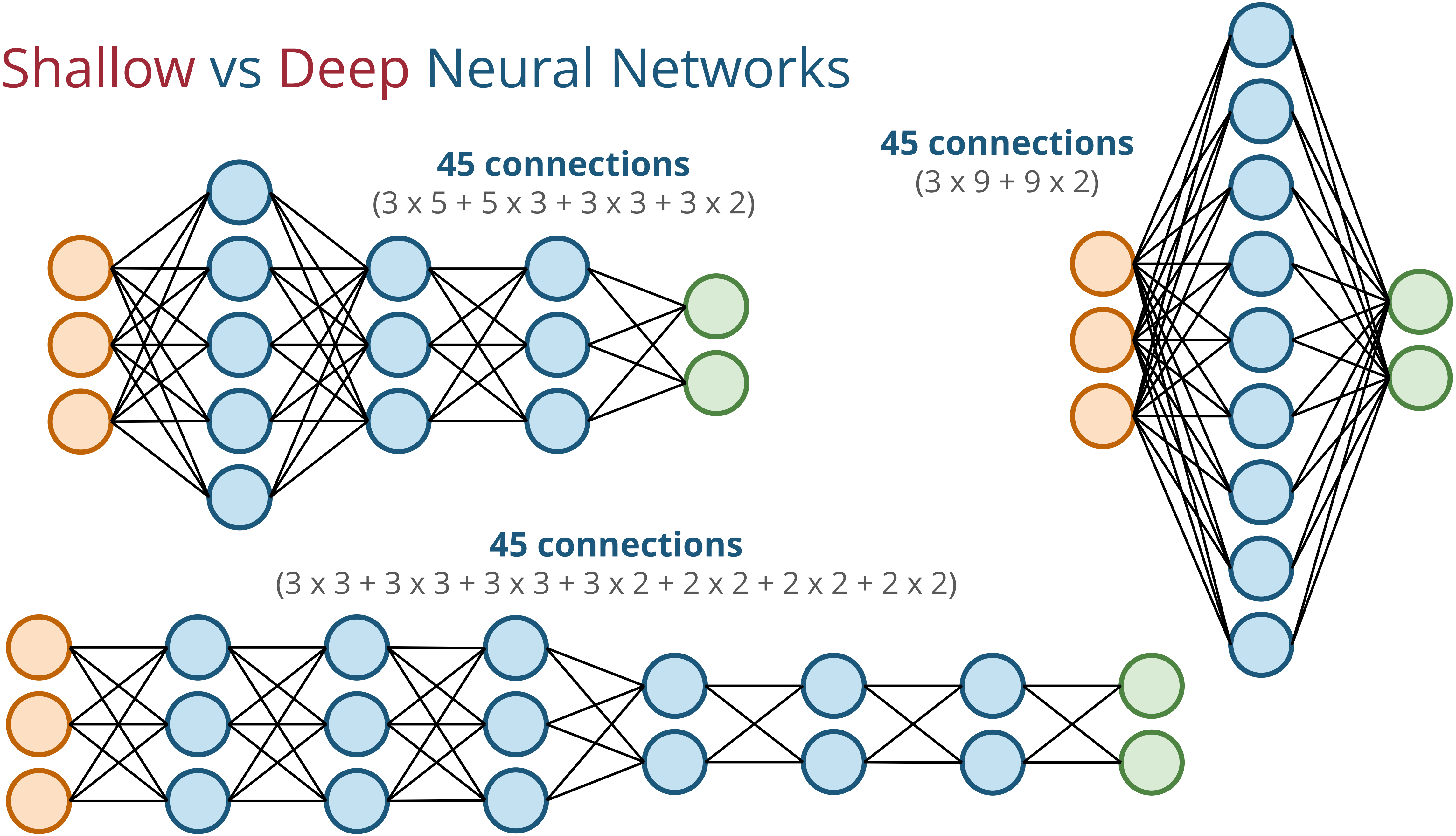
Universal Approximation Theorem

- A neural network with **one hidden layer** can **approximate any continuous function** given **sufficient hidden neurons** and **appropriate activation functions**
  - Sigmoid, ReLUs are good activation functions



Then why do we want to go deep?

Shallow vs Deep Neural Networks



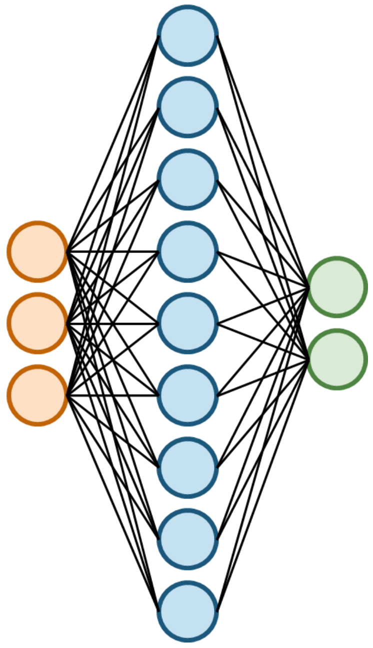
45 connections
(3 x 5 + 5 x 3 + 3 x 3 + 3 x 2)

45 connections
(3 x 9 + 9 x 2)

45 connections
(3 x 3 + 3 x 3 + 3 x 3 + 3 x 2 + 2 x 2 + 2 x 2 + 2 x 2)

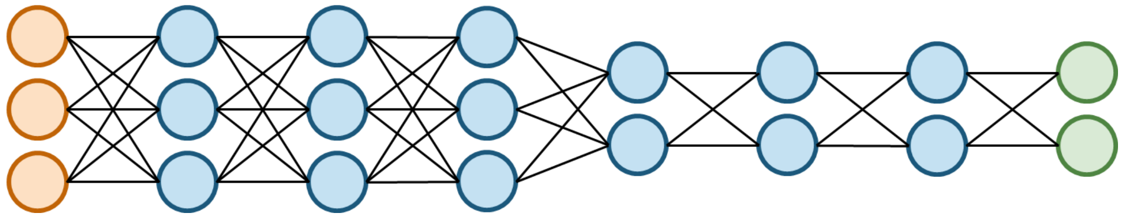
Shallow vs Deep Neural Networks – In Practice

Shallow neural nets



Less expressive
(less parameter efficient)

Deep neural nets

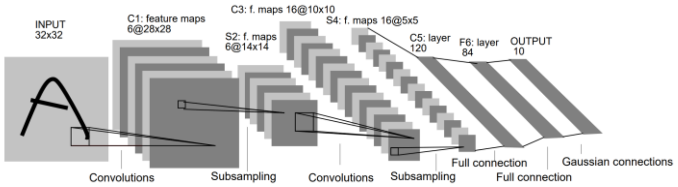


More expressive
(more parameter efficient)

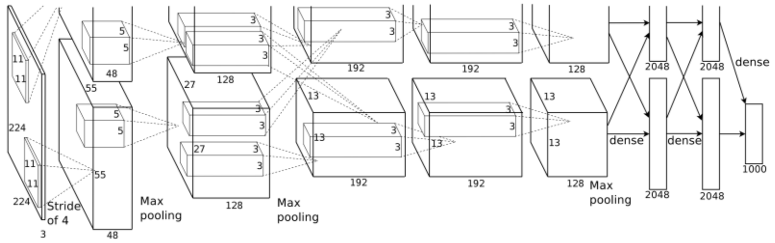
How Deep is Deep Enough?

- **Deeper is not always better**
  - Actual number of parameters
  - Optimization difficulties
  - Data size
  - Inductive bias of the model

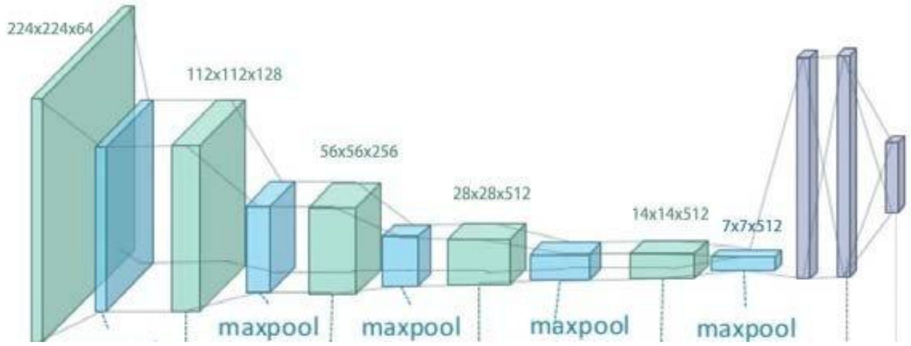
LeNet
(1998)



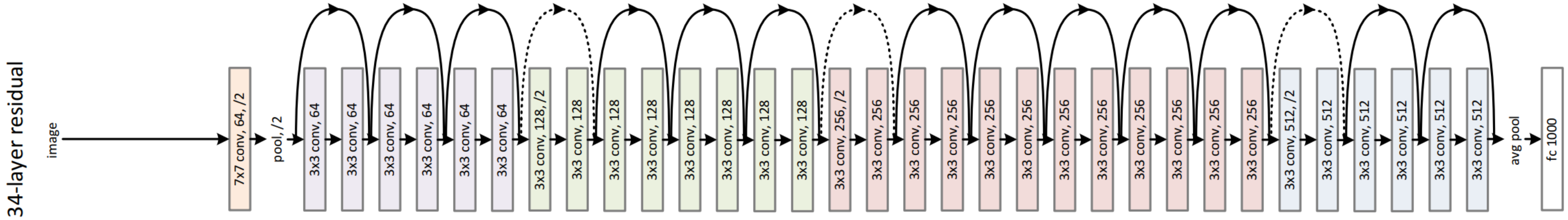
AlexNet
(2012)



VGG-19
(2015)

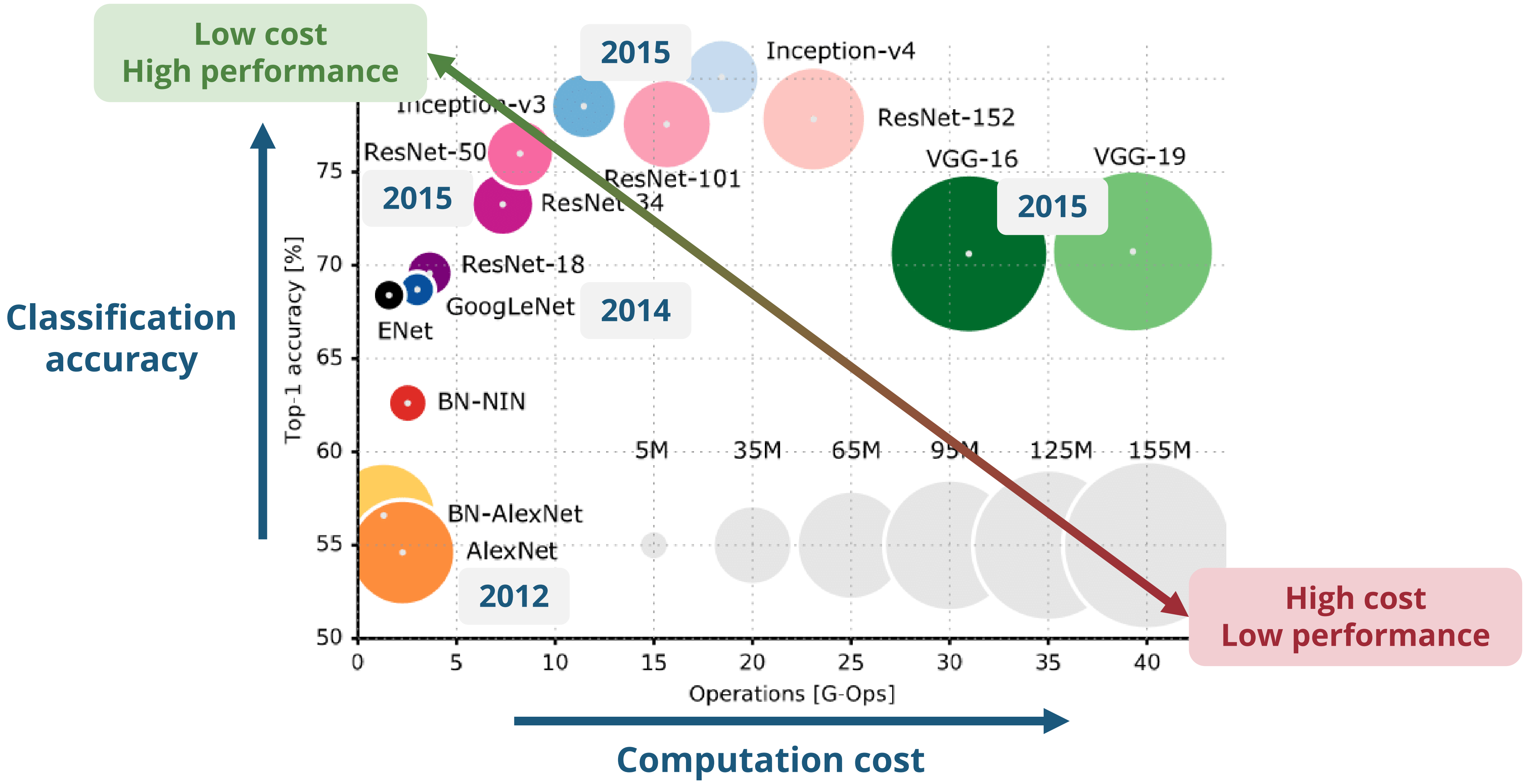


ResNet
(2015)

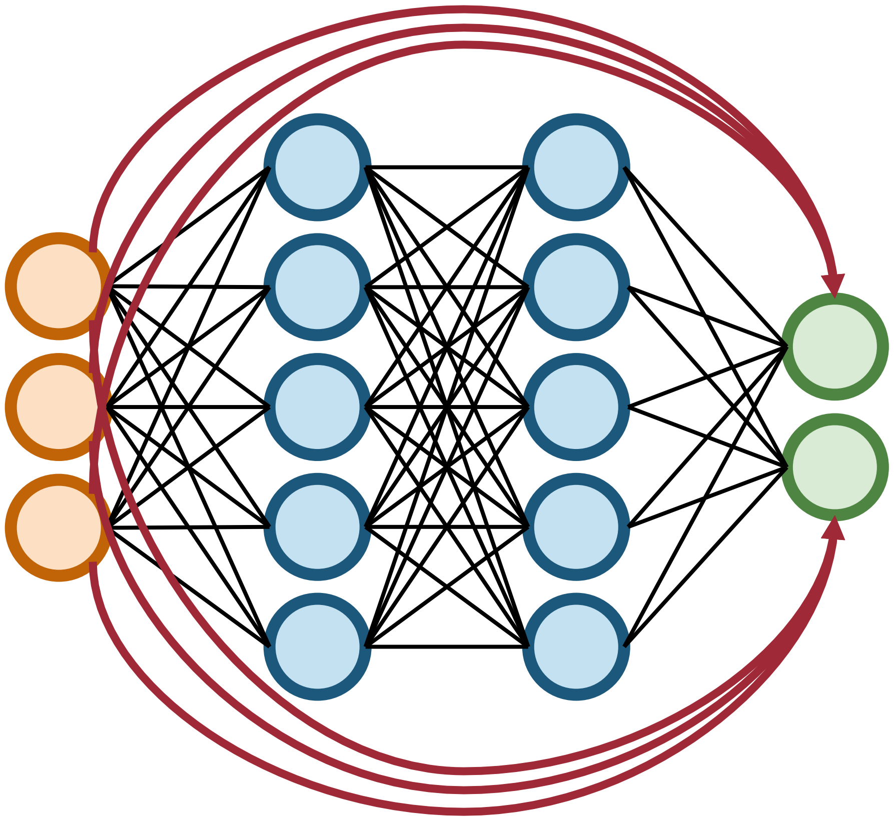


Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, 1998.
Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *NeurIPS*, 2012.
Karen Simonyan and Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ICLR*, 2015.
Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," *CVPR*, 2016.

Computation Cost vs Classification Accuracy

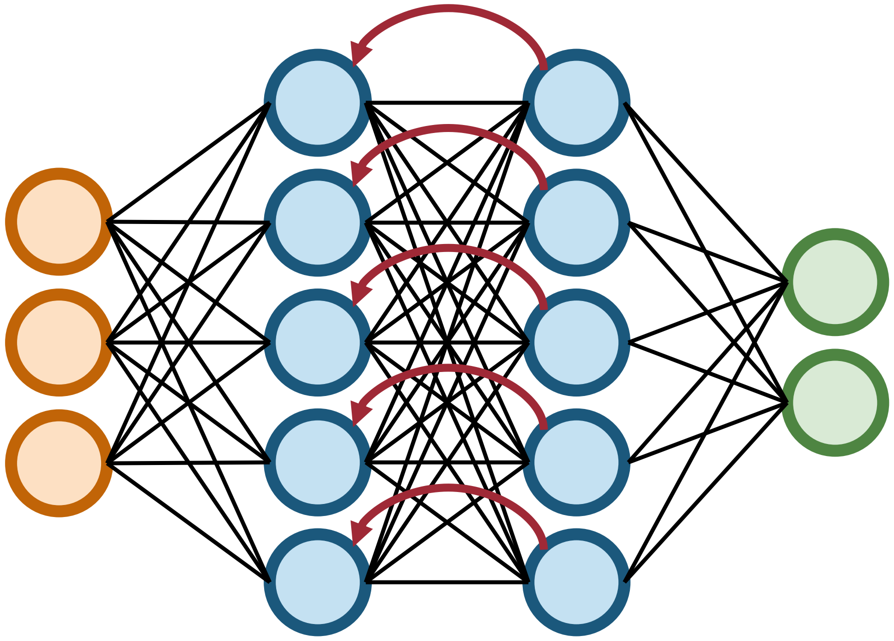


Neural Networks are NOT always Layer-by-Layer



Skip connections

Used in ResNets, U-Nets, diffusion models



Feedback loops

Used in RNNs, LSTMs, GRUs