Northwestern Kellogg

Northwestern ENGINEERING

MBAi 448 I Winter 2026

Deep Learning

Alex Castrounis

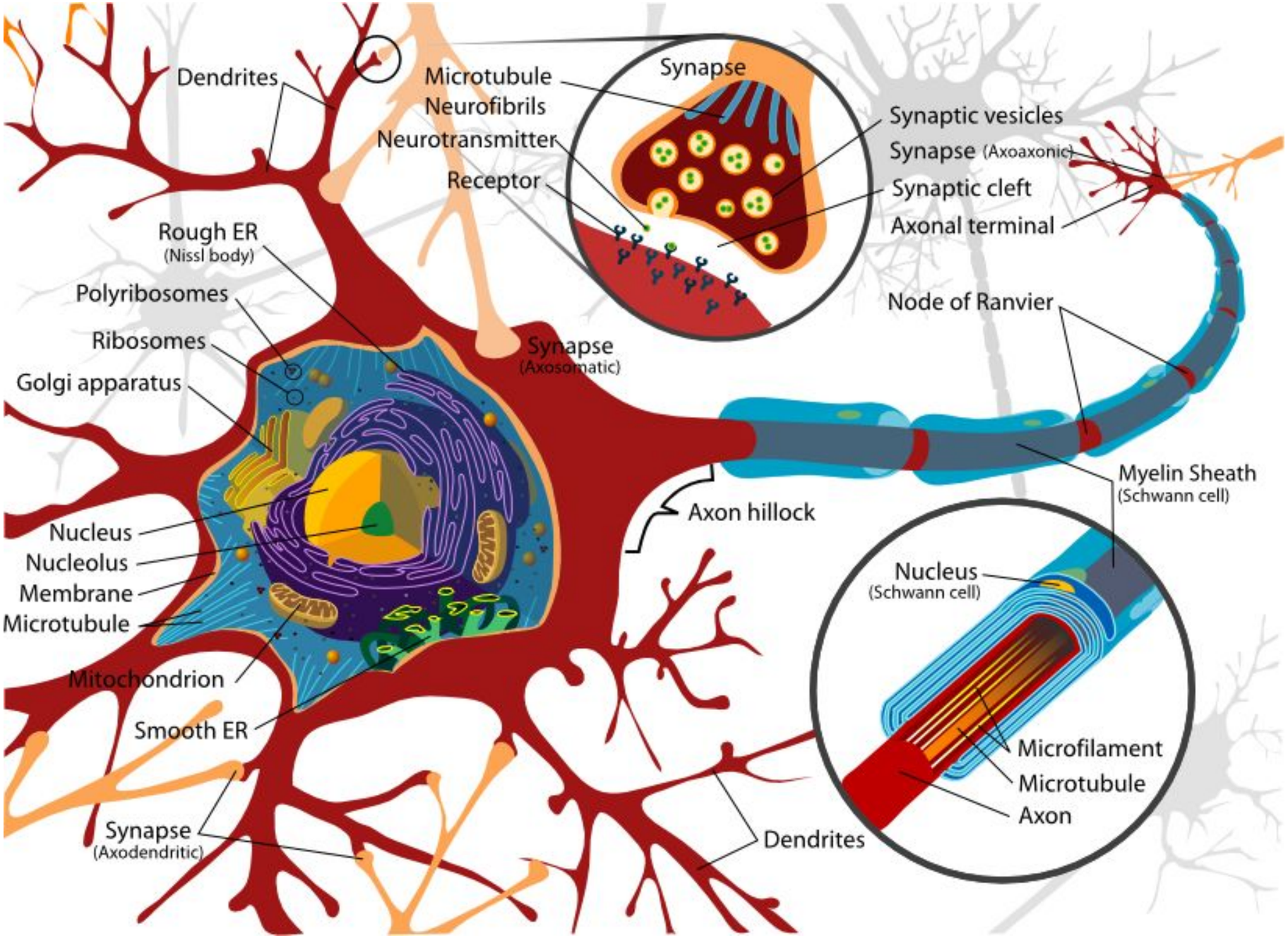
www.whyofai.com | linkedin.com/in/alexcastrounis

Attendance

Enter the “Magic Word” in Canvas

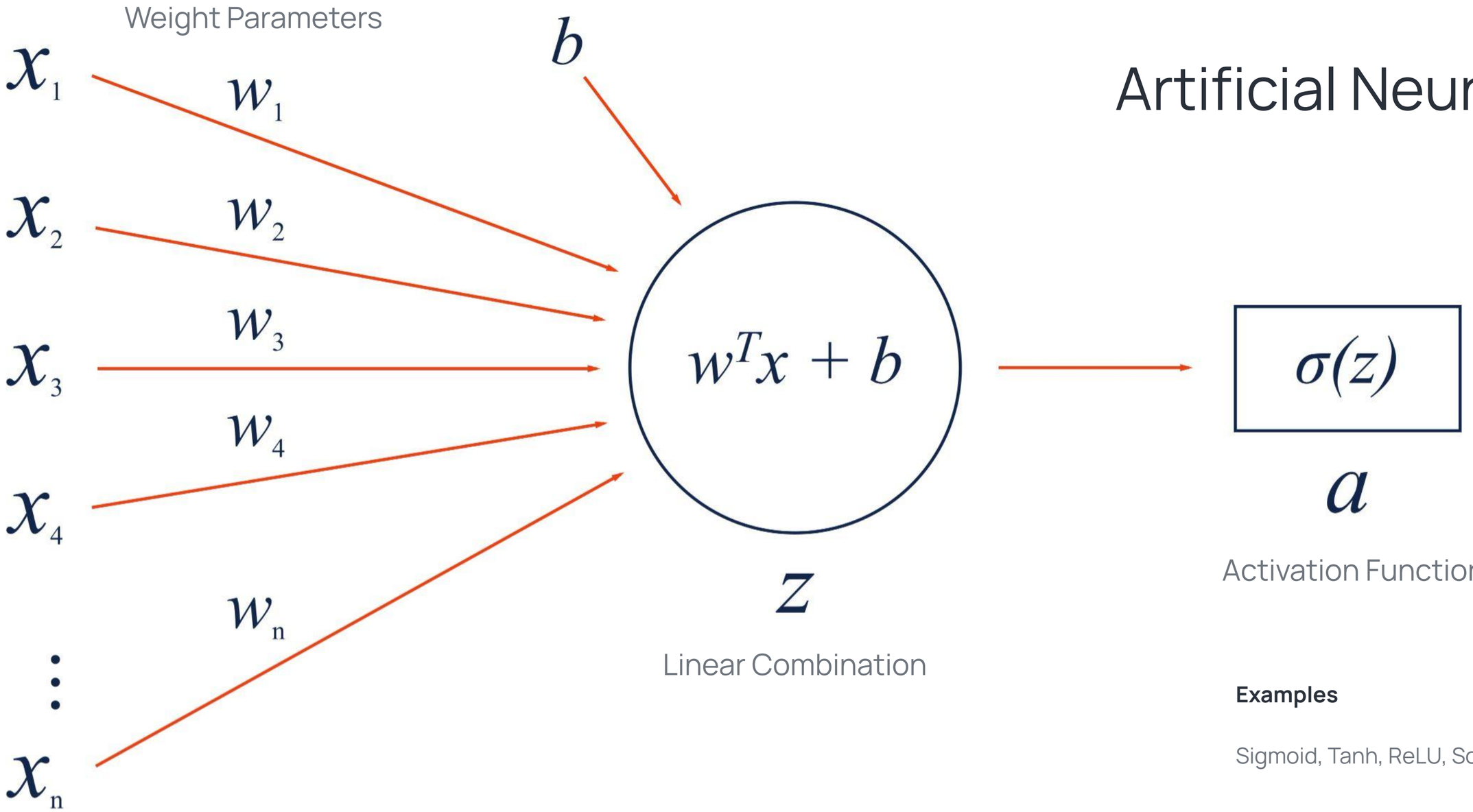
Do Not Share with Absent Students

Image via ChatGPT



By LadyofHats [Public domain], via Wikimedia Commons

Artificial Neuron

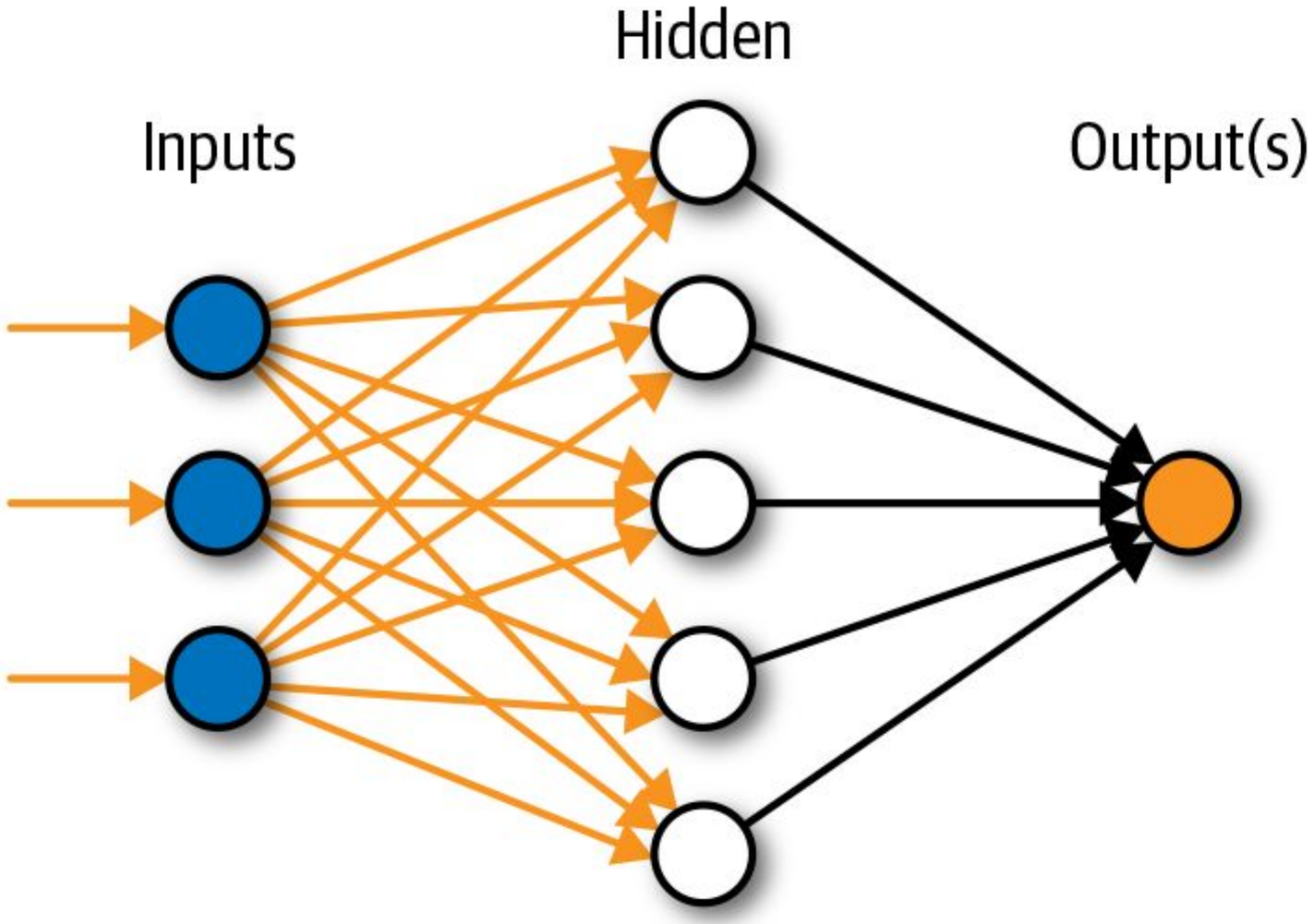


Examples

Sigmoid, Tanh, ReLU, Softmax, ...

X = Inputs / Data (e.g., image pixels, sequence of images, audio, text characters, words)

Artificial Neural Network



Shallow Learning = 1 hidden layer | Deep Learning = 2 or more hidden layers

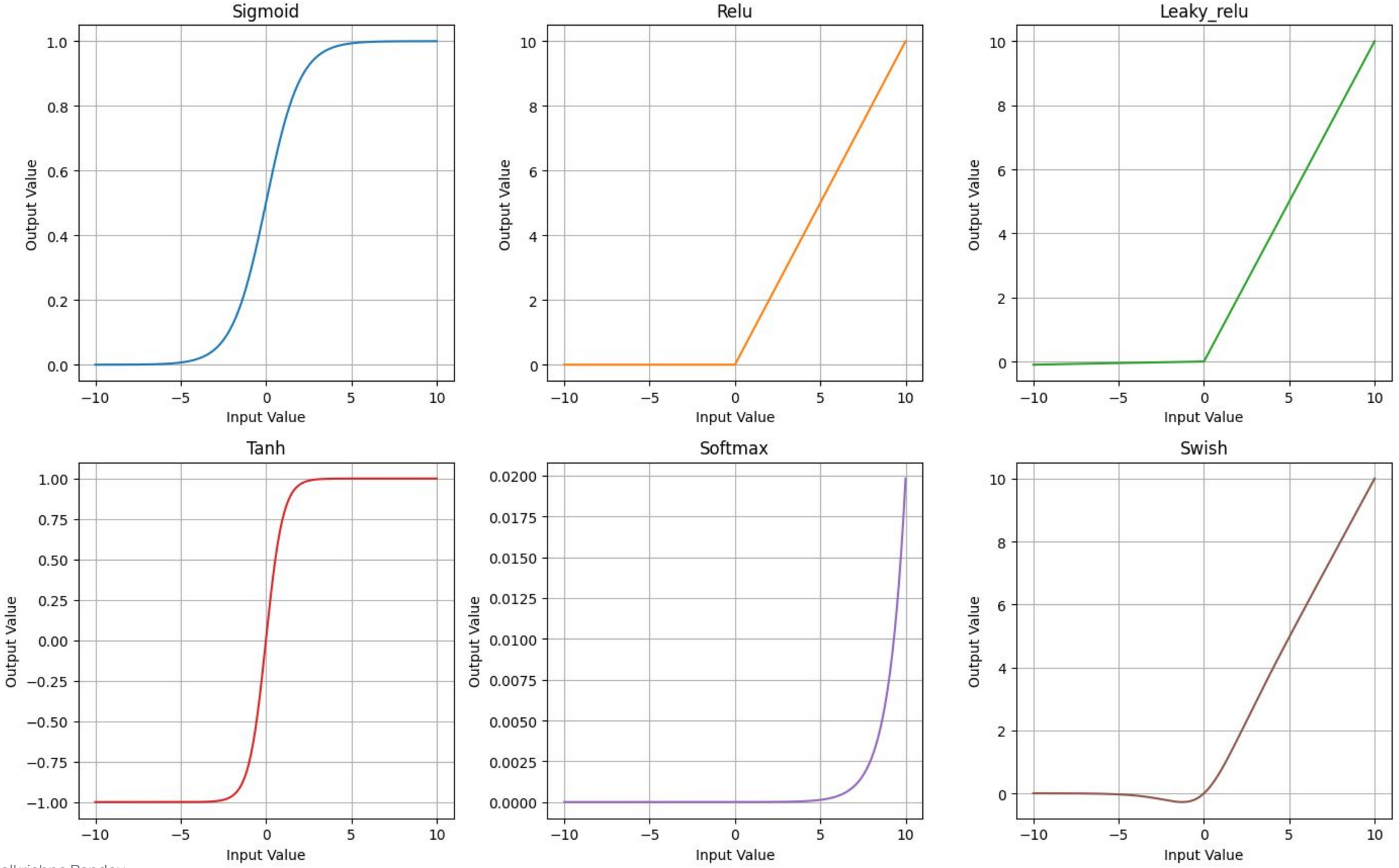


Image Credit: Balkrishna Pandey

Relevant Concepts and Terms

- **Training loss**: Diff between actual and predicted values
- **Backpropagation**: The algorithm that implements gradient descent in neural networks.
- **Iteration**: A single update of a model's parameters—the model's weights and biases—during training.
- **Batch**: The set of examples used in one training iteration. The batch size determines the number of examples in a batch.
- **Epoch**: A full training pass over the entire training set such that each example has been processed once.
- **Representation learning**: Automatically learning useful features from raw data (in the form of feature representations and hierarchies in latent space)

Definitions via Google's Machine Learning Glossary

Activity

From Pixels to Decisions

Use case: Factory camera inspects glass beverage bottles for cracks

Goal: Explain (1) a **feature hierarchy** and (2) **activation choices** using the graphs

Output options:

- Option 1: OK vs Crack (binary)
- Option 2: OK vs Crack vs Flag for human inspection (multi-class)

***** Don’t use any devices for assistance**
***** Follow guidance on activity steps**

Activity

Activation functions:

- tanh (−1 to +1)
- ReLU (0+ for positive inputs)
- sigmoid (0 to 1)
- softmax (probabilities sum to 1 across classes)

What does each layer learn?

Layer 1 learns: _____
Layer 2 learns: _____
Layer 3 learns: _____

Choose activations

Hidden layers:
Output (binary):
Output (3-way):

***** Don't use any devices for assistance**
***** Follow guidance on activity steps**

Activity

Discussion

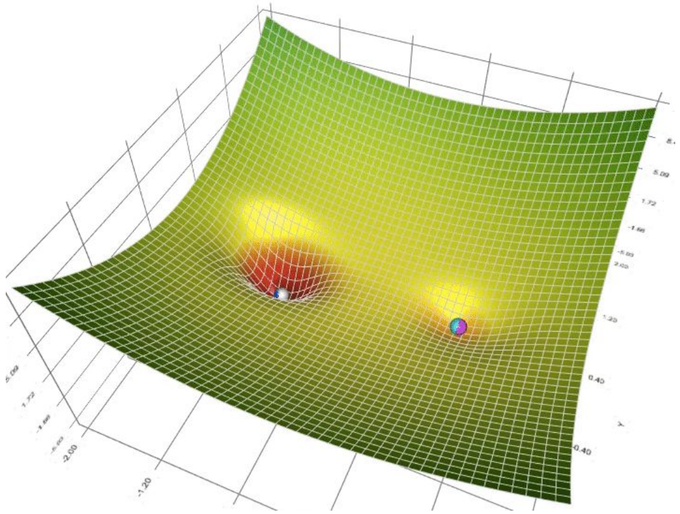
- What does your network learn early → middle → late?
- Why did you choose specific hidden layer activation functions?
- Why did you choose specific output layer activation functions?
- When would you prefer “Flag for inspection” instead of forcing a confident label?
- Looking at the activation graphs, which functions have big “flat” regions where the output barely changes, and what happens to the learning signal as it passes back through many layers like that?
- Looking at the activation graphs, which functions can produce very large outputs or very steep changes, and what happens to the learning signal as it passes back through many layers like that?

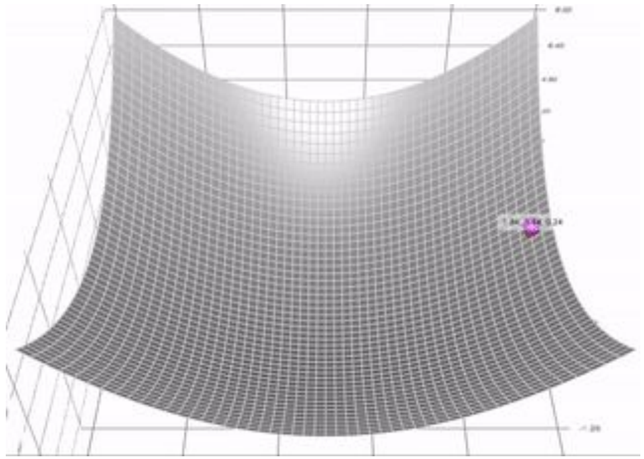
***** Don’t use any devices for assistance**
***** Follow guidance on activity steps**

$$\frac{dh}{dx} = \frac{dh}{du} \cdot \frac{du}{dx}$$

The Chain Rule of Calculus via Machine Learning Mastery

Optimization





$$l \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N}(y_i - (wx_i + b))^2$$

Goal = Minimize MSE (our cost function)

$$\frac{\partial l}{\partial w} = \frac{1}{N}\sum_{i=1}^{N} -2x_i(y_i - (wx_i + b));$$

$$\frac{\partial l}{\partial b} = \frac{1}{N}\sum_{i=1}^{N} -2(y_i - (wx_i + b)).$$

Parameter Partial Derivatives

$$w \leftarrow w - \alpha\frac{\partial l}{\partial w};$$

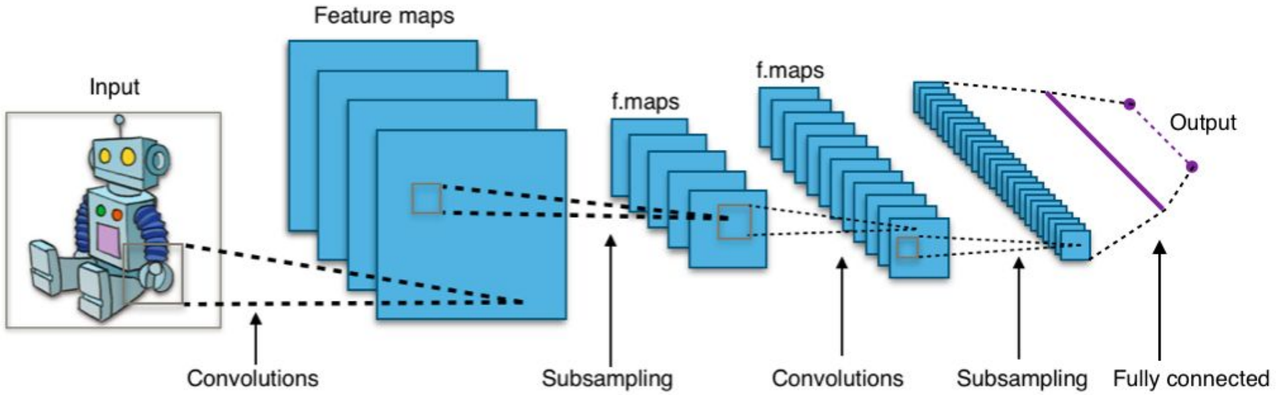
$$b \leftarrow b - \alpha\frac{\partial l}{\partial b}.$$

Parameter Updates per Epoch

Images credit: Towards Data Science The & 100-Page ML Book by Andriy Burkov

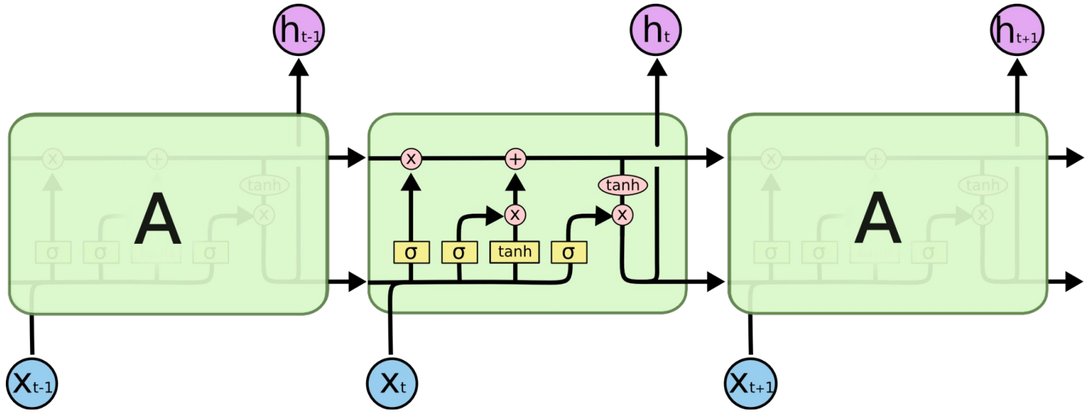
DL Architectures Visualized

CNNs



By Aphex34 - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=45679374

RNNs



Shown: Long short-term memory (LSTM)

https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Transformers

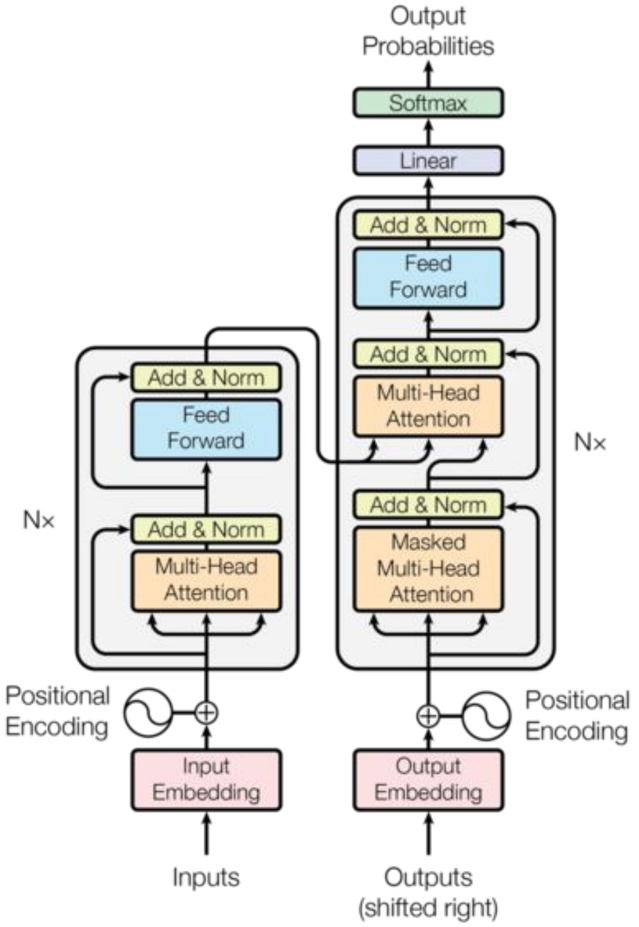


Image credit: Hugging Face

“Learning” Features and Feature Hierarchies

aka Feature Extraction

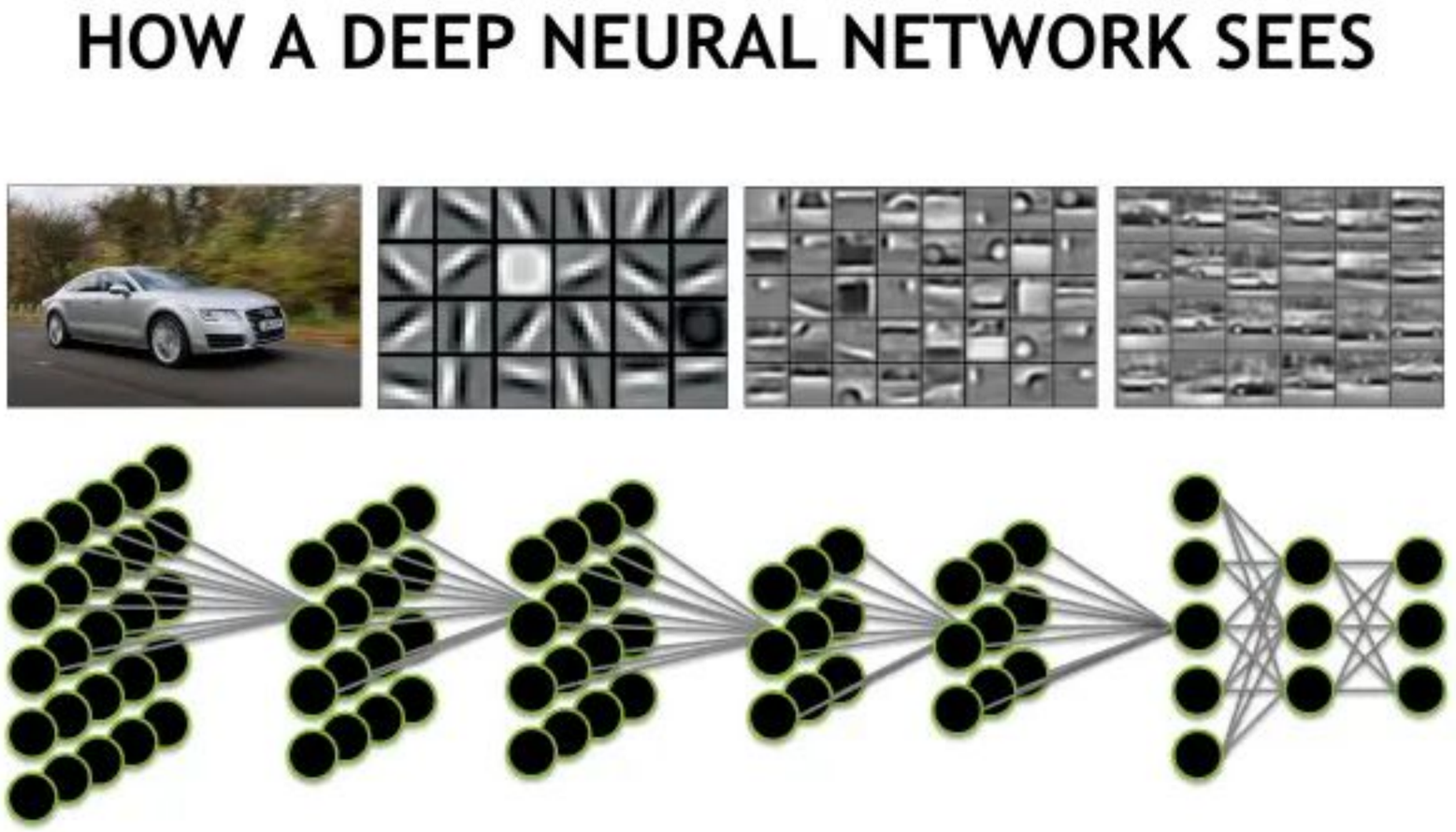
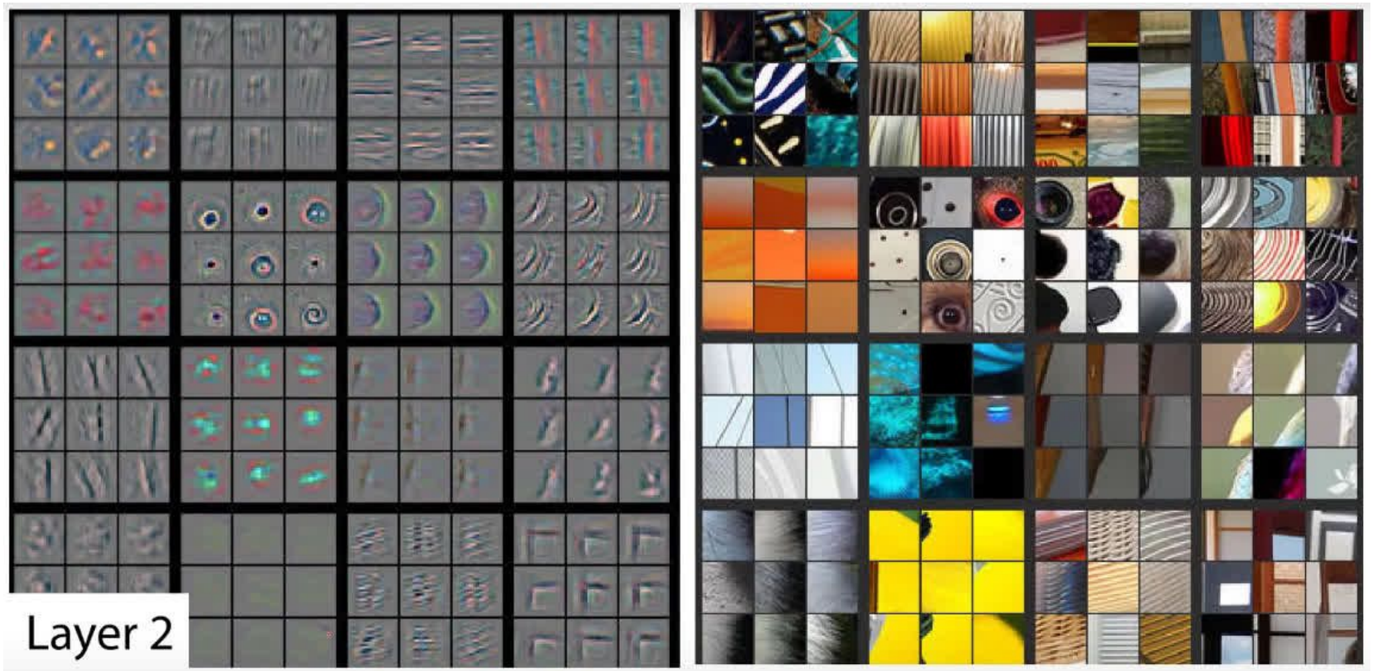


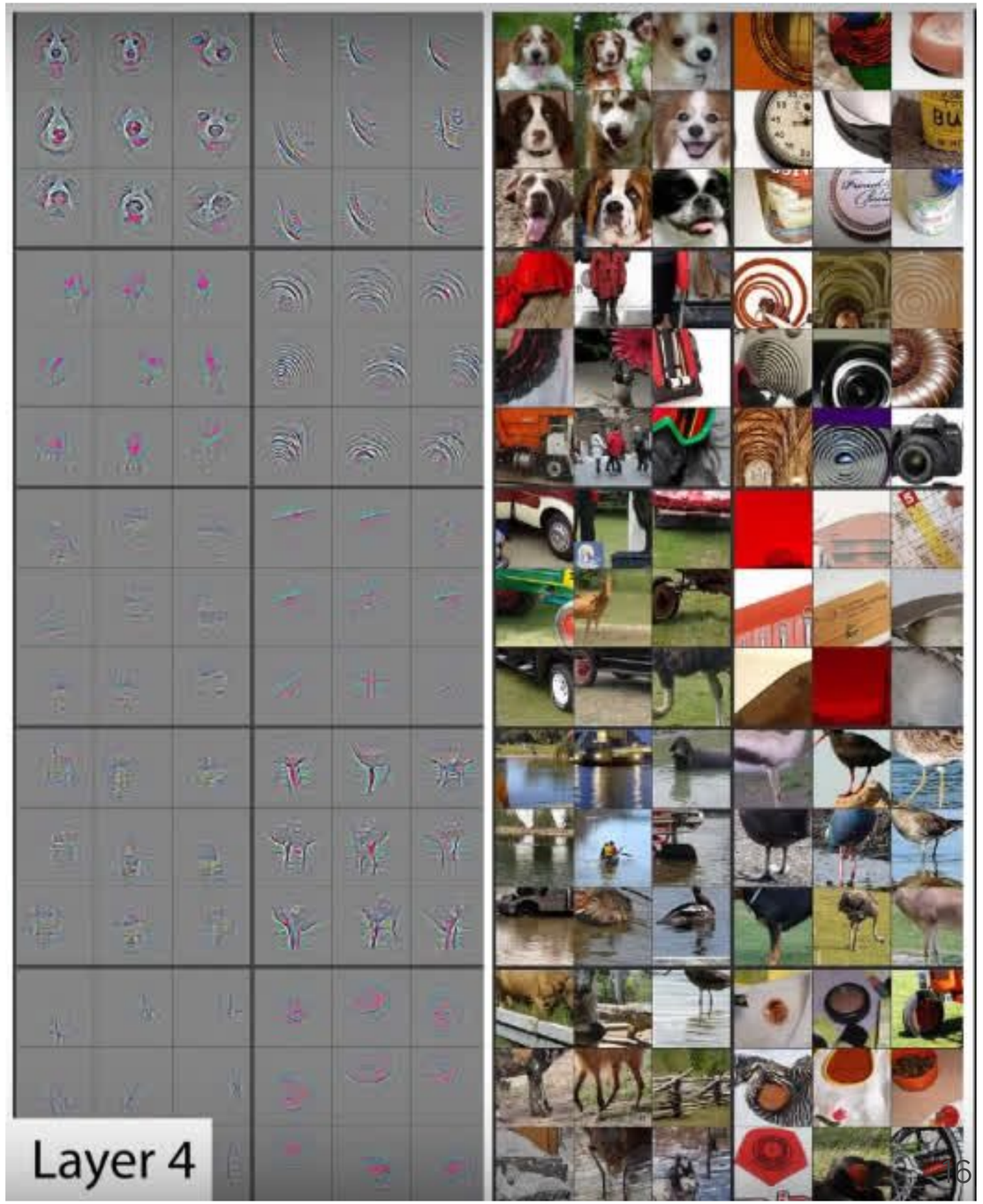
Image credit: NVIDIA

Visualizing Features



CNNs handle invariance!

(e.g., position/translation, rotation, scale)



Images credit: Convolutional Neural Networks (CNNs) Explained

Pre-Training and Transfer Learning

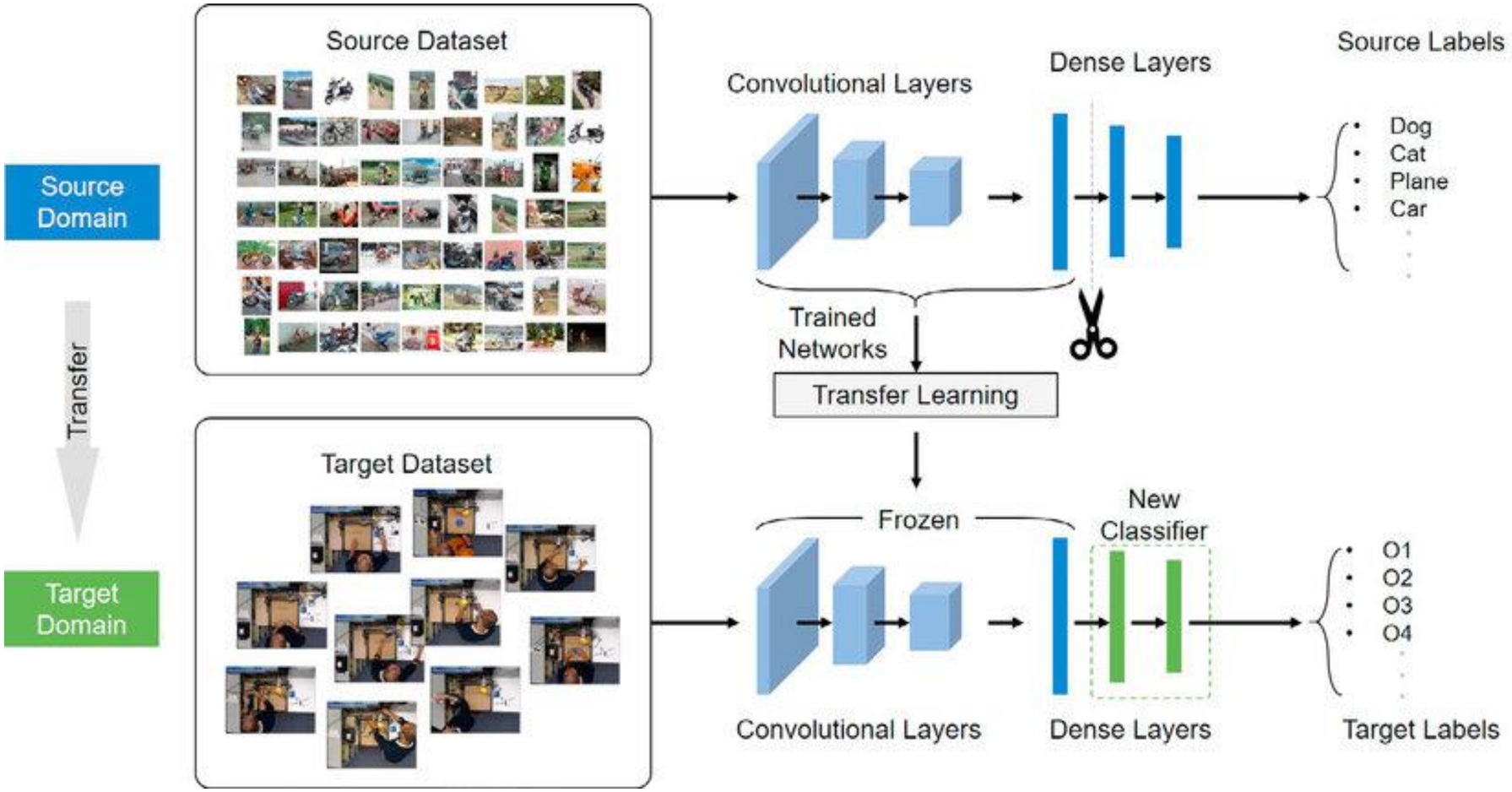
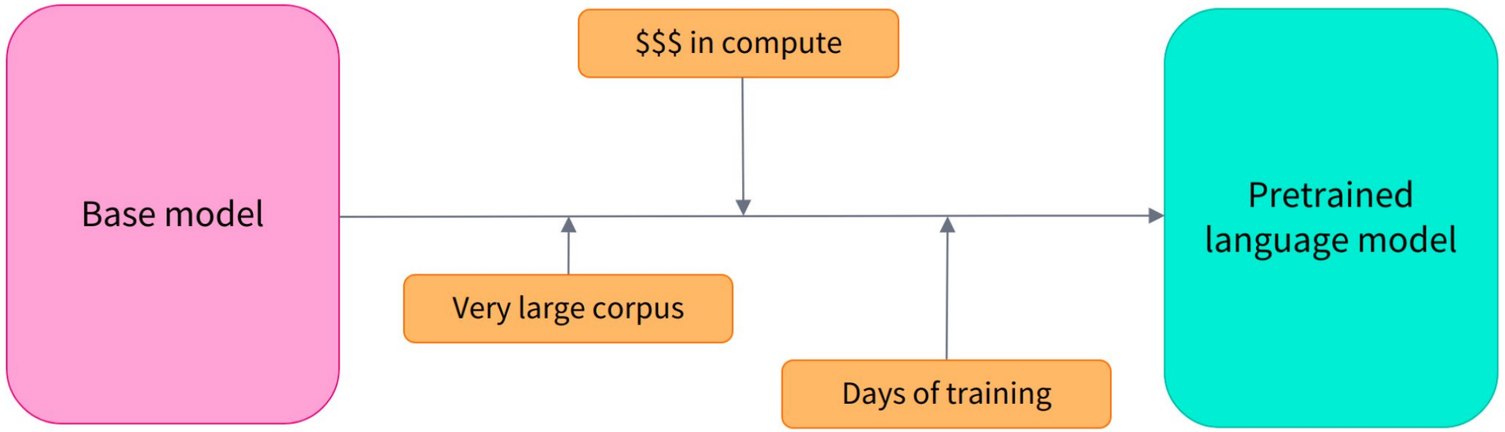


Image credit: Tao, Wenjin & Al-Amin, Md & Chen, Haodong & Leu, Ming & Yin, Zhaozheng & Qin, Ruwen. (2020). Real-Time Assembly Operation Recognition with Fog Computing and Transfer Learning for Human-Centered Intelligent Manufacturing. Procedia Manufacturing. 48. 926-931. 10.1016/j.promfg.2020.05.131.

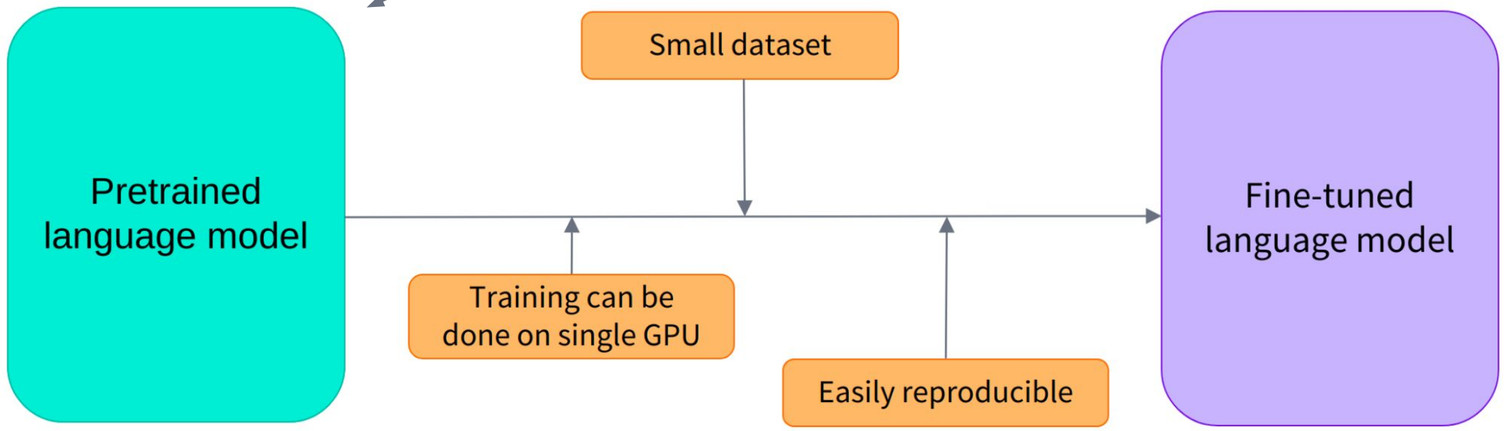
Pre-Training

› Requires a huge corpus of data and can take several weeks



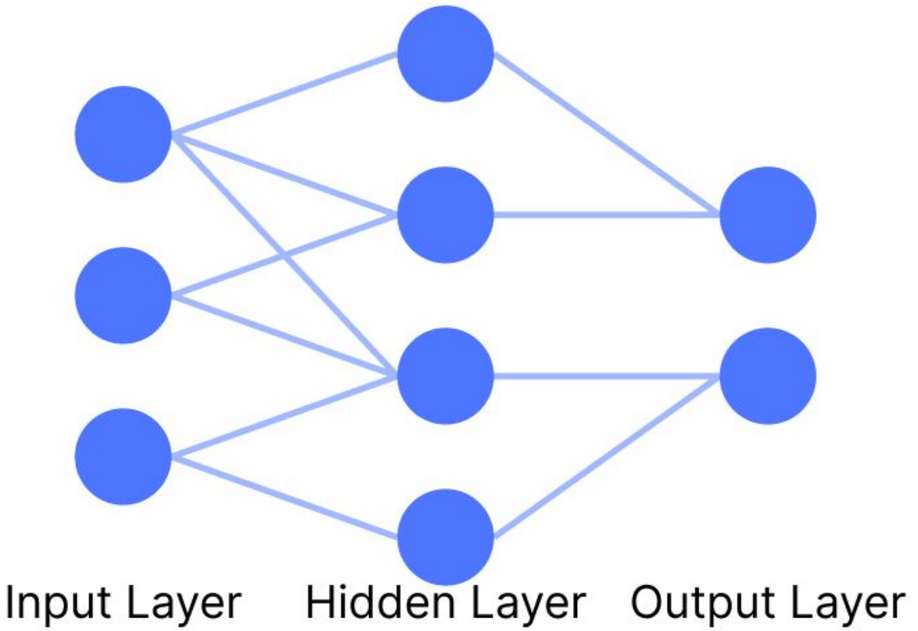
Transfer Learning / Fine-Tuning

Ex: arXiv corpus resulting in a science/research-based model



Images credit: Hugging Face

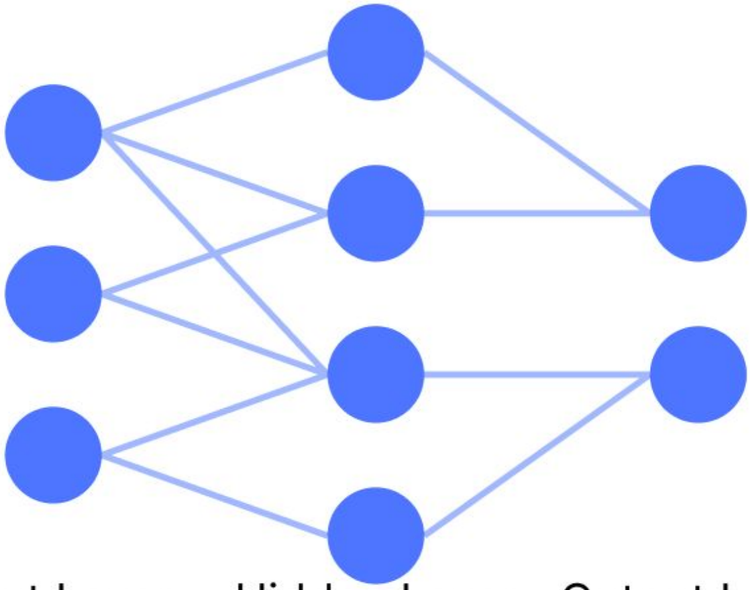
Connectionist AI



- Learns from data
- Pattern recognition
- Black box decisions

Connectionist vs Symbolic AI

Connectionist AI



- Learns from data
- Pattern recognition
- Black box decisions

Symbolic AI

IF temperature > 30°C

THEN status = "hot"

IF status = "hot"

THEN action = "let cool"

Rule-based Knowledge Base

- Explicit rules
- Logical reasoning
- Transparent decisions

Activity

Draw the Trust Line

Goal: Decide what must be **Symbolic**, what can be **Connectionist**, and what should be **Hybrid**

Rule: If a wrong answer is costly → the system must be able to **refuse / escalate**

***** Don’t use any devices for assistance**
***** Follow guidance on activity steps**

Activity

Scenario: Budget vs Actual Review

Each month, leaders ask: “**Why were we off plan?**”

The system proposes explanations and actions:

- **Compute** the difference (Budget vs Actual)
- **Suggest** the most likely reason (“what changed?”)
- **Recommend** an action (“what should we do?”)
- **Write** a 2–3 sentence summary for leadership

For each: choose: **Connectionist / Symbolic / Hybrid**

***** Don’t use any devices for assistance**
***** Follow guidance on activity steps**

Activity

Discussion:

- Where is a “confidently wrong” answer unacceptable?
- What should the system do when it doesn’t know?
- What evidence should be required before it can speak?
- What does a “safe hybrid” look like in one sentence?

***** Don’t use any devices for assistance**
***** Follow guidance on activity steps**