# Deep Learning

Yann LeCun, Yoshua Bengio & Geoffrey Hinton
NYU UdeM UofT

Nature 2015

DMAIS@CAU
Yeongon Kim

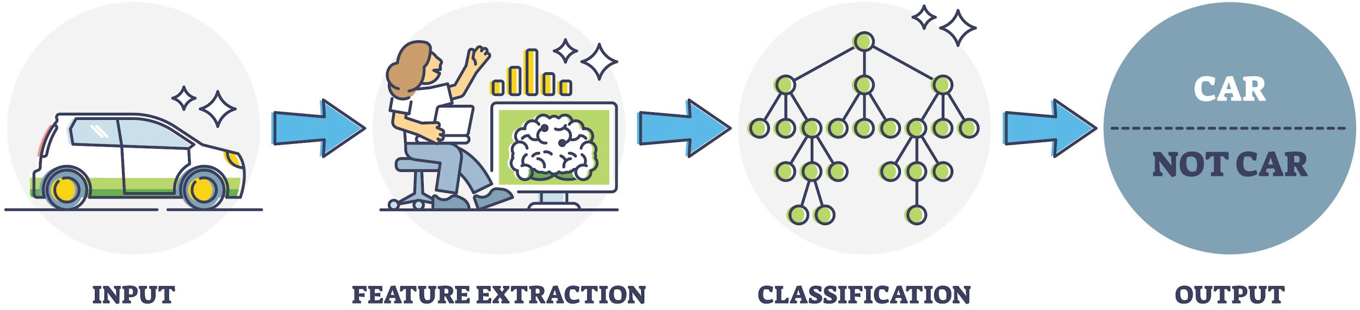
# INDEX

- **Machine Learning & Deep Learning**
- **General Procedure of Deep Learning**
- **Convolutional Neural Network**
- **Recurrent Neural Network**
- **Future of Deep Learning**

# Conventional Machine Learning

## ❑ What is Machine Learning?

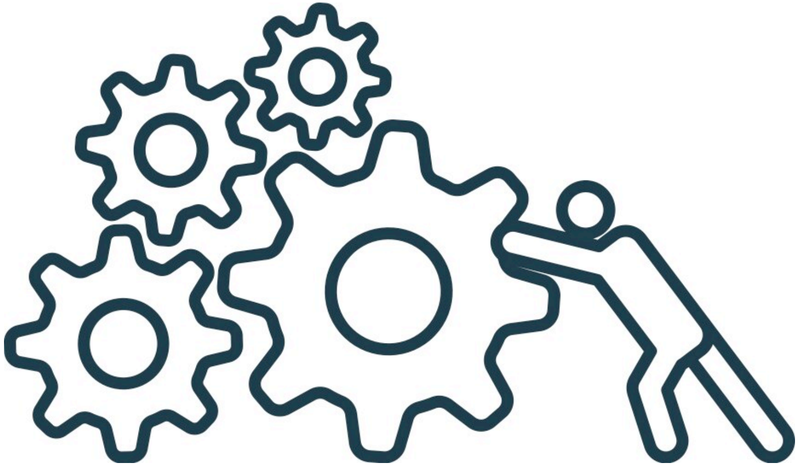
- Study of algorithms that learns from data and perform tasks without explicit instructions
- Hand-engineered feature extractor pulls meaningful features from raw data into a feature vector
- Classifier uses feature vectors to make the final prediction



# Limitations of Conventional Machine Learning

## ❑ Hand-Engineered Feature Extractor

- Model performance depends on the quality of the extracted features
- Careful engineering and considerable domain expertise is required



# Limitations of Conventional Machine Learning

## ❑ Selectivity–Invariance Dilemma

- **Selectivity**: Responding to features that are important for distinguishing between classes
- **Invariance**: Rremaining insensitive to features irrelevant for discrimination, such as an animal’s pose

➡ **How can we address this dilemma?**

# How Humans Recognize Objects

## ❑ Identify Objects by Core Concepts

- Capture the essential attributes of an object, not just form or texture
- Invariant to position and selective to relevant features

chair

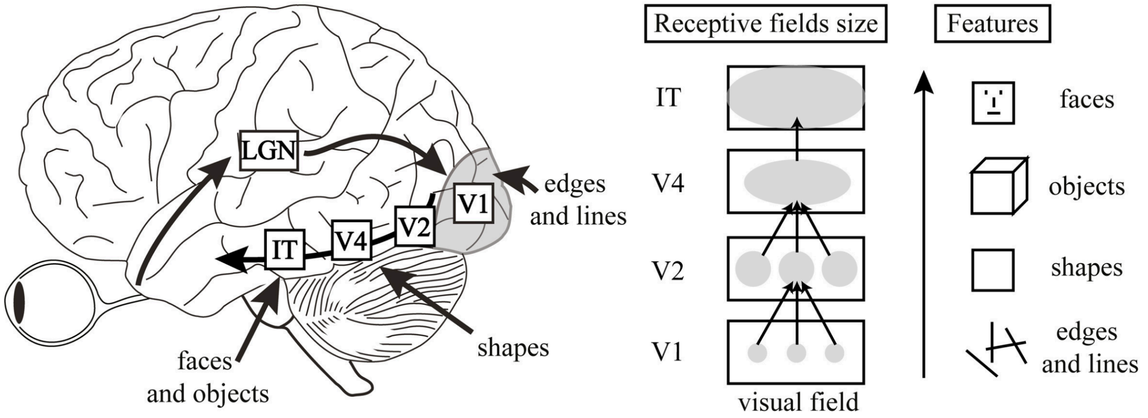
chair

# How Humans Recognize Objects

## ❑ Visual Cortex of Human Visual System

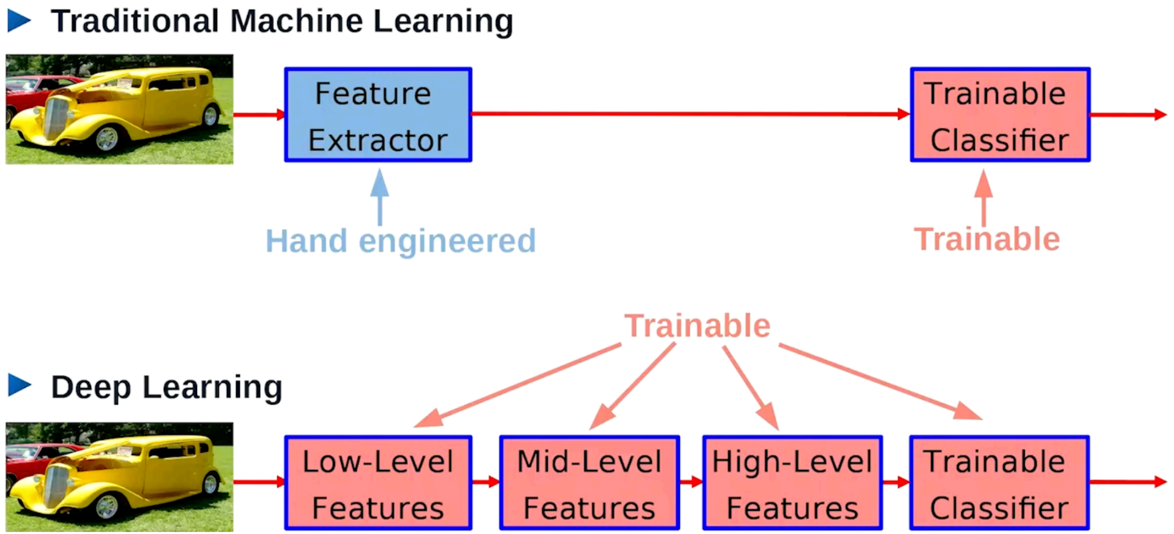
- Neurons are organized hierarchically and process visual information in stages
- Focusing on ‘**what**’ something is, rather than ‘**where**’ or ‘**how**’ it appears

➡ **Deep learning is inspired by the hierarchical structure of neural networks**



# Deep Learning

- ❑ **Representations of data are learned via multiple layers of abstraction**
  - Better selectivity and invariance
- ❑ **End-to-end learning using a general learning procedure**
  - No intermediate step such as feature extraction

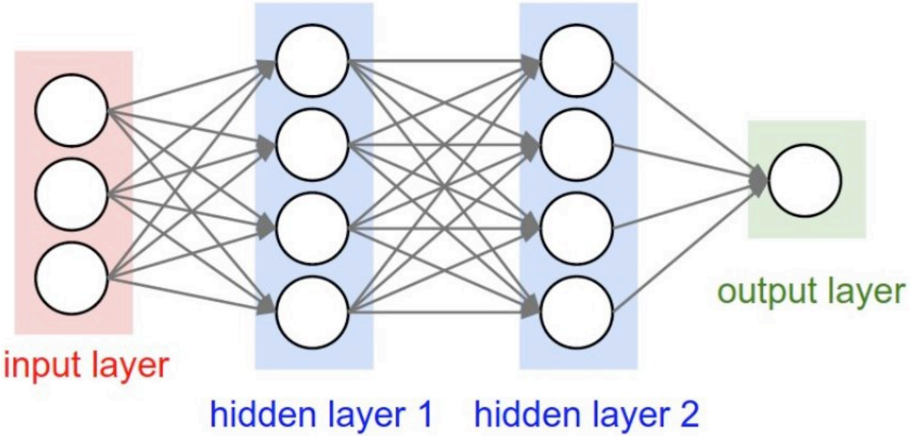


# INDEX

- Machine Learning & Deep Learning
- General Procedure of Deep Learning
- Convolutional Neural Network
- Recurrent Neural Network
- Future of Deep Learning

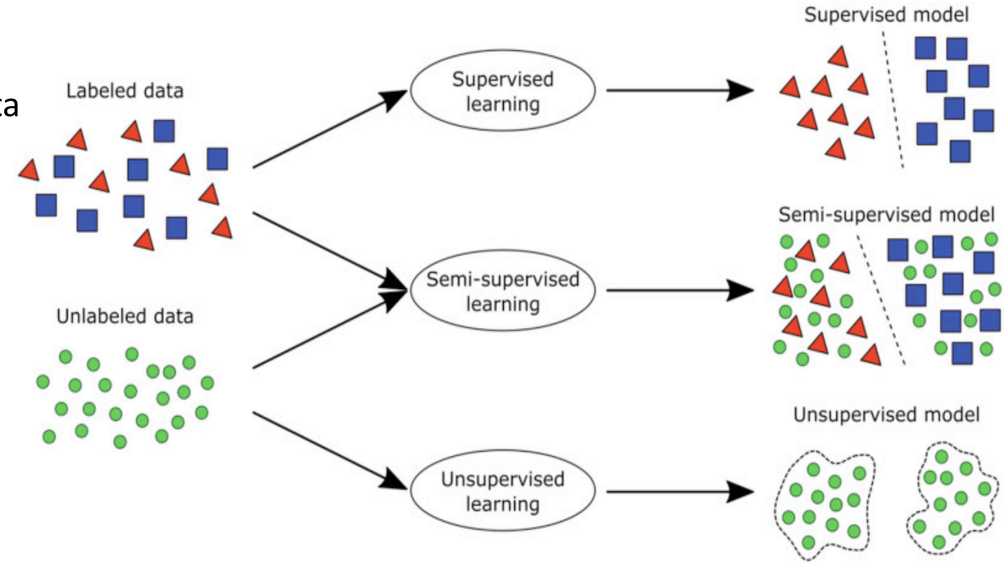
# Neural Network Architecture

- Hierarchical model inspired by the brain's neural structure
- Deep learning is implemented with deep neural networks
- Data is transformed through hidden layers and the final output is produced at the output layer



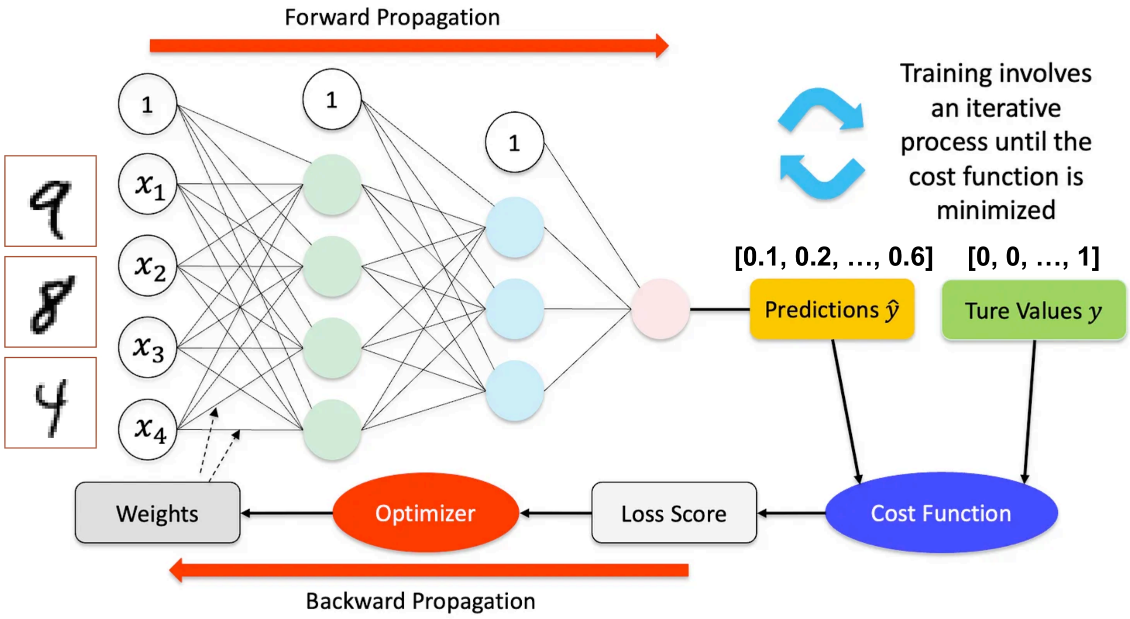
# Learning Paradigms

- **Supervised Learning**
  - Training a model using labeled data
- **Semi-supervised Learning**
  - Training with labeled data and unlabeled data
- **Unsupervised Learning**
  - Discovering patterns without labeled data



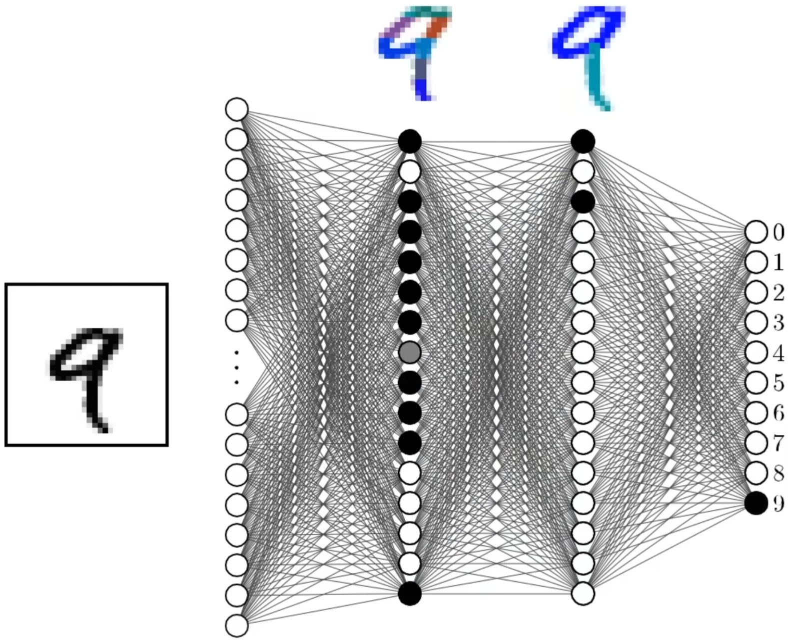
# Supervised Learning

- Most common form of deep learning and machine learning
- Labeled data is used to compare model predictions with actual values
- Weights are updated in the direction that minimize loss



# Forward Propagation

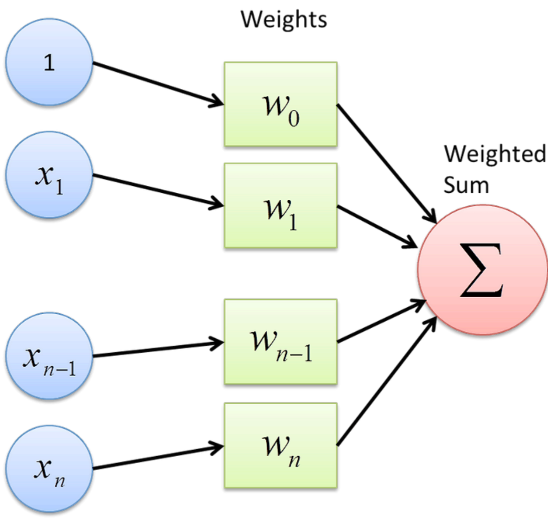
- Transforming input into output through the model
- Each neuron's activation feeds into neurons in the next layer, propagating through the network



# Forward Propagation

## ❑ Weighted Sum

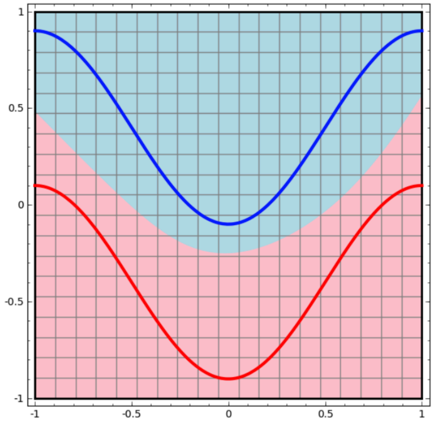
- Operation by which each neuron receives its input signals
- Neuron's activation level is represented by the weighted sum of its inputs
- Bias is added to a neuron’s weighted sum to adjust its activation threshold

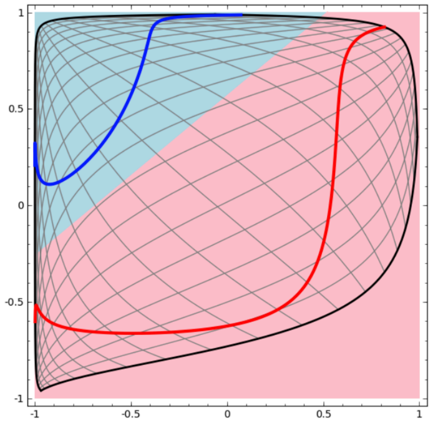


# Forward Propagation

## ❑ Activation Function

- Pass values through the activation function before forwarding them to the next layer
- Introduces **non-linearity** to neural networks, enabling them to learn more complex patterns



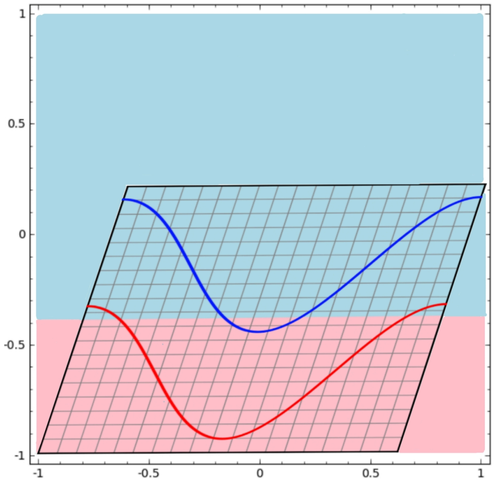



# Forward Propagation

- **Why are weighted sums alone insufficient for learning complex patterns?**
  - No matter how many times linear transformation is applied, they remain linear
  - Difficult to learn complex patterns without non-linear activation

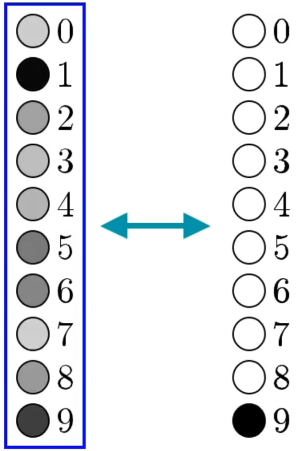
$$\mathbf{y} = W_4^T \left( W_3^T \left( W_2^T \left( W_1^T \mathbf{x} + \mathbf{b}_1 \right) + \mathbf{b}_2 \right) + \mathbf{b}_3 \right) + \mathbf{b}_4$$

$$\mathbf{y} = W^T \left( \mathbf{x} + \mathbf{b} \right)$$



# Objective Function

- Compute the error, which indicates how much the predictions differ from the true labels
- Goal of the model is to minimize the error




Prediction Label

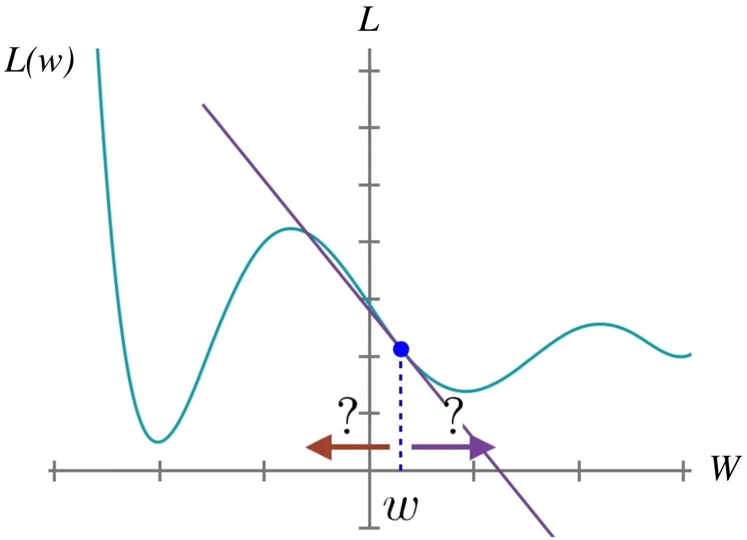
$$
\left\{
\begin{aligned}
&(0.22-0.00)^2+\\
&(0.99-0.00)^2+\\
&(0.37-0.00)^2+\\
&(0.26-0.00)^2+\\
&(0.30-0.00)^2+\\
&(0.53-0.00)^2+\\
&(0.48-0.00)^2+\\
&(0.19-0.00)^2+\\
&(0.40-0.00)^2+\\
&(0.77-\boxed{1.00})^2
\end{aligned}
\right.
$$

Mean Squared Error

# Gradient Descent

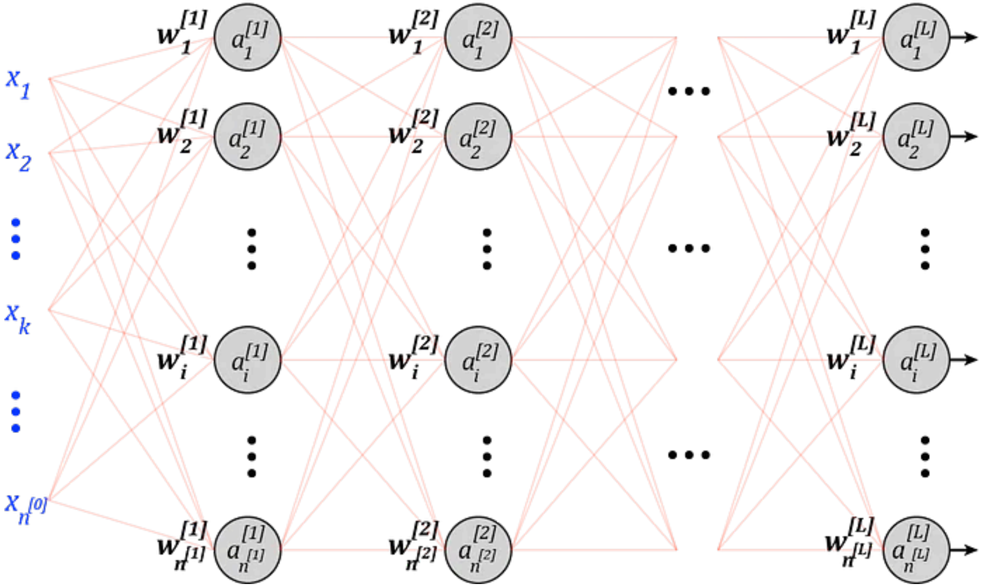
- Update the parameters opposite the loss gradient to minimize loss

$$w_{t+1} = w_t - \alpha \frac{\partial L}{\partial w_t}$$



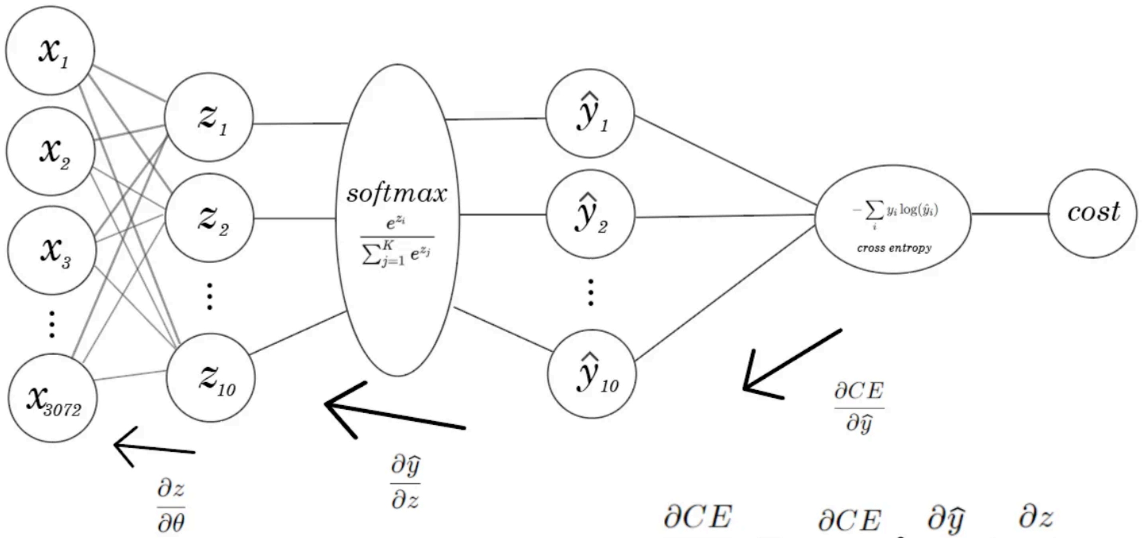
# Gradient Descent

- How can we compute and apply updates to every parameter across all layers of the network?



# Back-Propagation

- Procedure to compute the gradient of a loss function with respect to the weights using chain rule
- Intermediate activations are cached during the forward pass and reuse them in the backward pass




$$\frac{\partial CE}{\partial \theta_{ij}} = \frac{\partial CE}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial \theta_{ij}}$$

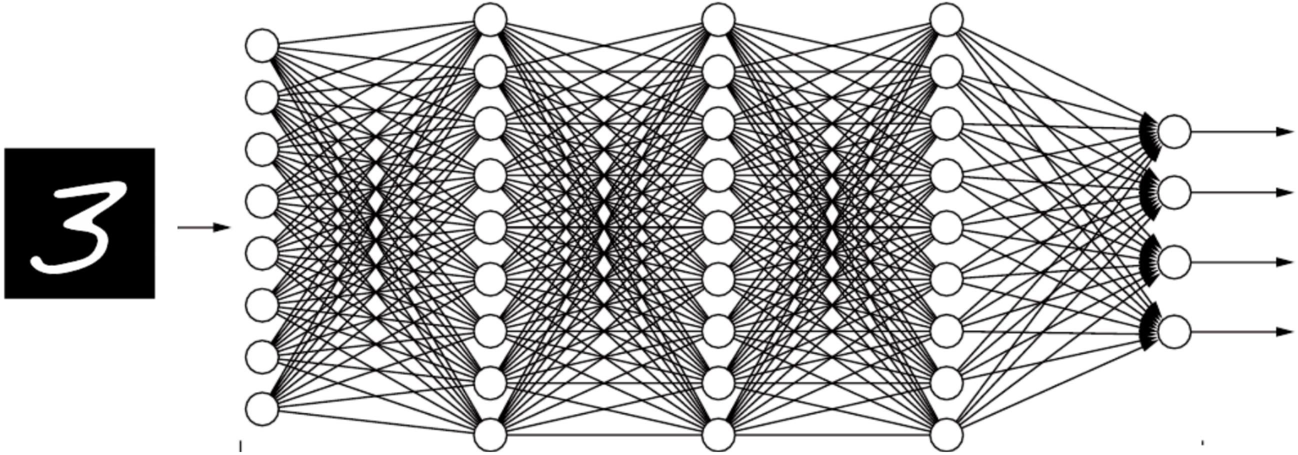
# INDEX

- Machine Learning & Deep Learning
- General Procedure of Deep Learning
- Convolutional Neural Network
- Recurrent Neural Network
- Future of Deep Learning

# Needs For Specialized Models: Images

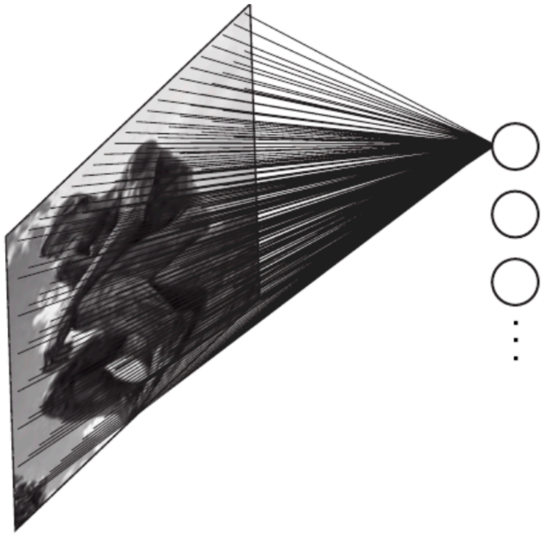
## Multi-Layer Perceptron

- Feedforward neural network composed of fully connected layers



# Needs For Specialized Models: Images

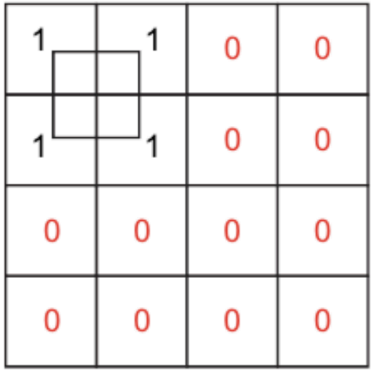
- **Requires a lot of weight parameters**
  - Small image size of 128x128 needs 16384 weights for each unit



# Needs For Specialized Models: Images

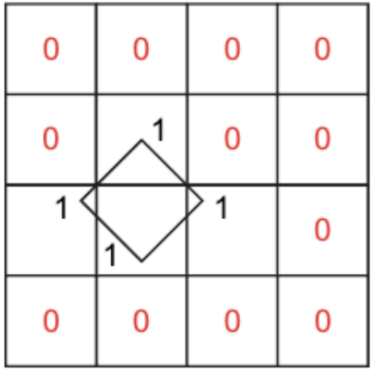
## ❑ Loss of spatial information

- Learns individual pixels, small shifts are treated as entirely different inputs
- Low invariance

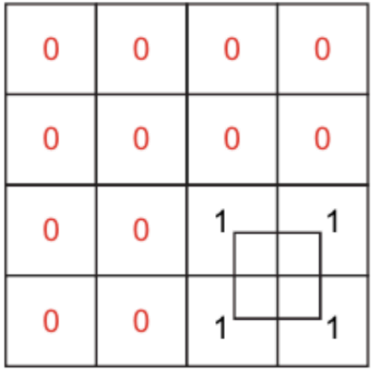



[1,1,0,0,1,1,0,0,0,0,0,0,0,0,0,0]

**Rectangle found**




[0,0,0,0,0,1,0,0,1,1,1,0,0,0,0,0]

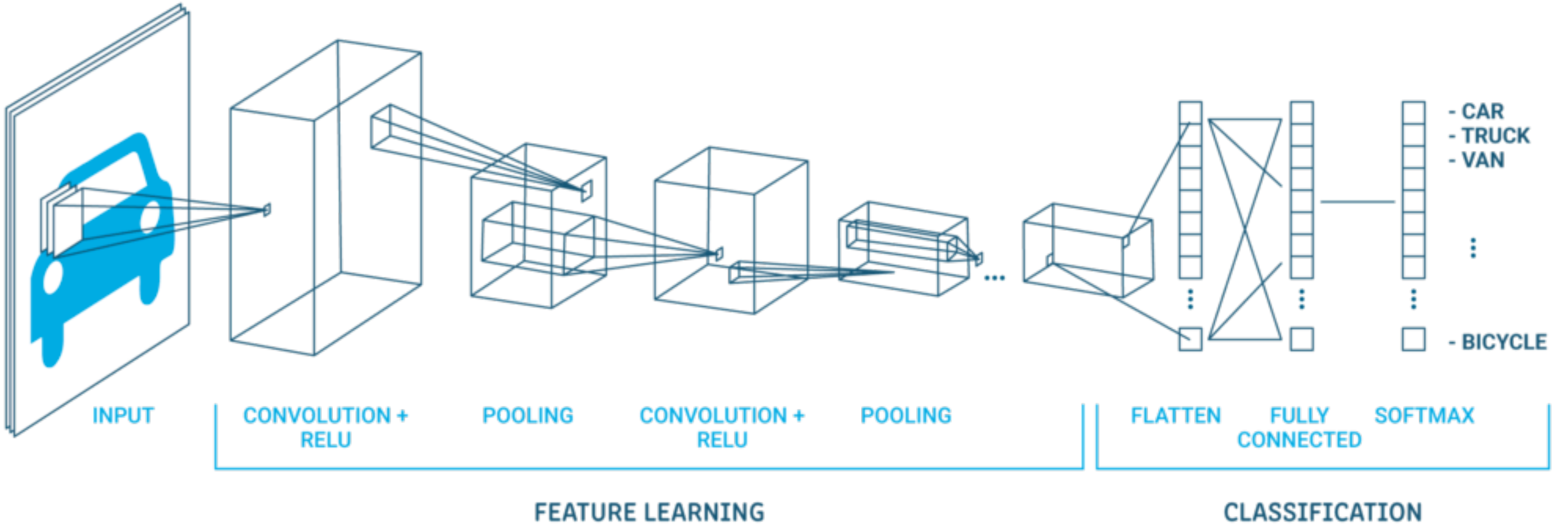



[0,0,0,0,0,0,0,0,0,0,1,1,0,0,1,1]

**Rectangle not found**

# Convolutional Neural Networks

- Structure designed to process data that come in the form of multiple arrays like images
- CNN consists of two parts: feature learning, decision making(classification)



# Convolution Layer

- Slides small kernels over the input to detect local patterns and generate feature maps

| | | | | | | |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 60 | 113 | 56 | 139 | 85 | 0 |
| 0 | 73 | 121 | 54 | 84 | 128 | 0 |
| 0 | 131 | 99 | 70 | 129 | 127 | 0 |
| 0 | 80 | 57 | 115 | 69 | 134 | 0 |
| 0 | 104 | 126 | 123 | 95 | 130 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Kernel

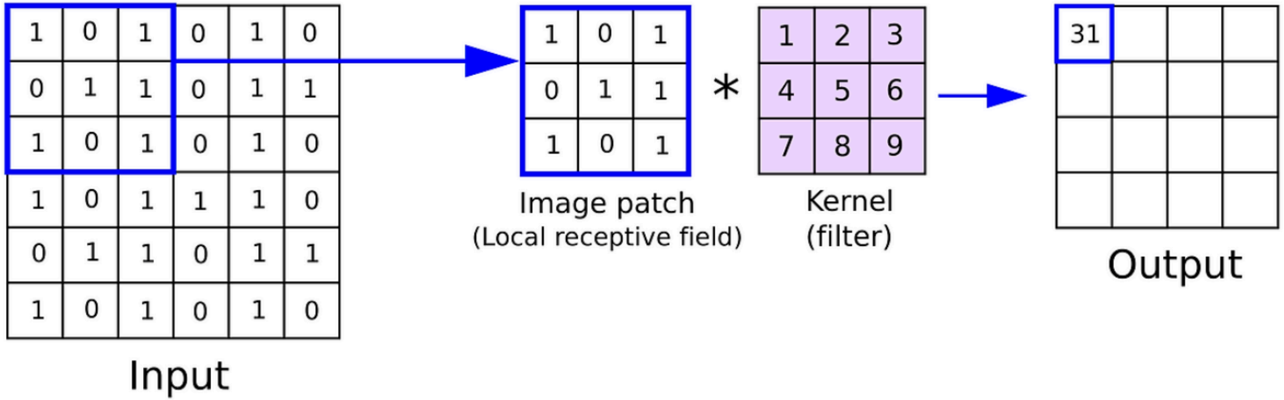
| | | |
|---|---|---|
| 0 | -1 | 0 |
| -1 | 5 | -1 |
| 0 | -1 | 0 |

| | | | | |
|---|---|---|---|---|
| 114 | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

# Convolution Layer

## ❑ Local Patches

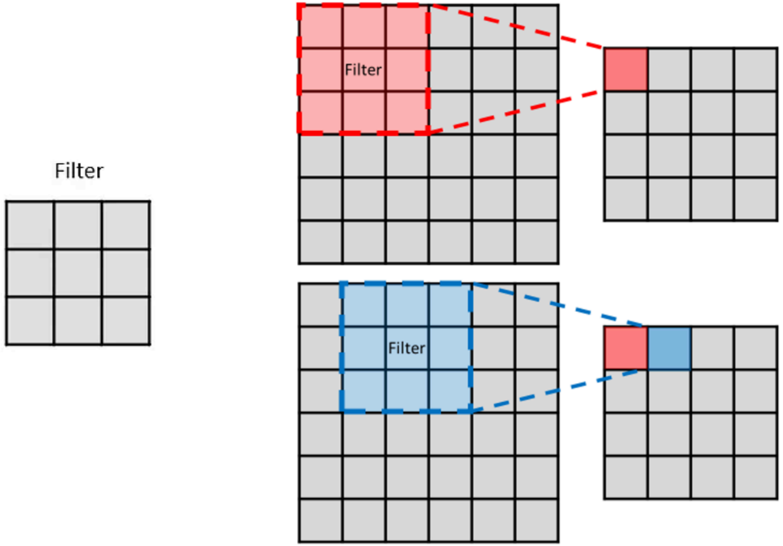
- Based on the insight that neighboring pixels are highly correlated
- Reduce the number of parameters by limiting the number of connections between units



# Convolution Layer

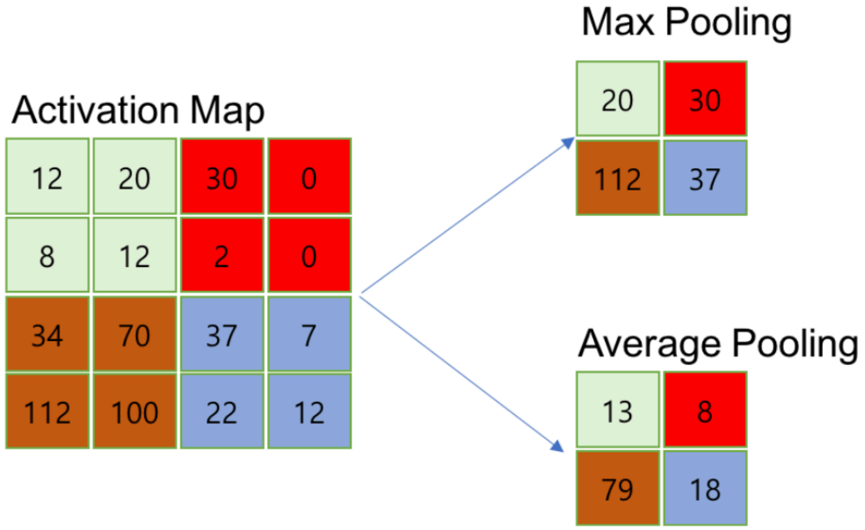
## ❑ Shared Weights

- Detect features regardless of their position within the image, providing translation invariance
- If a motif can appear in one part of the image, it could appear anywhere



# Pooling Layer

- Take the activated feature map as input and perform a pooling operation
- Merge semantically similar features into one
- Provides little translation invariance to the model
  - By aggregating units, it checks for presence within a region rather than exact positions



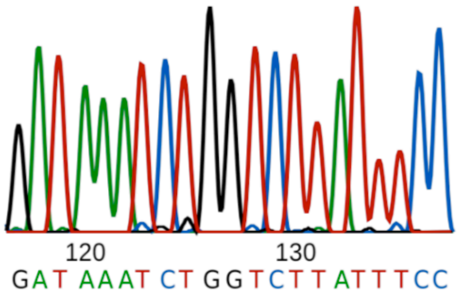
# INDEX

- Machine Learning & Deep Learning
- General Procedure of Deep Learning
- Convolutional Neural Network
- **Recurrent Neural Network**
- **Future of Deep Learning**

# Needs For Specialized Models: Sequences

## ❑ Sequence Data

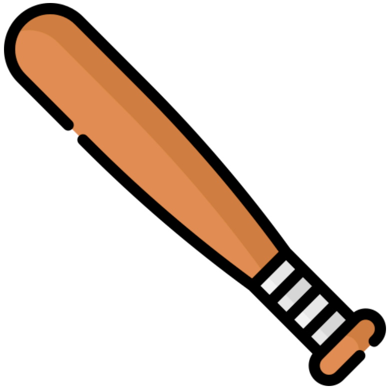
- Series of elements arranged in a specific order
- Order of elements in a sequence carries the essential information
- Essential information carried by order is called **context**




# Needs For Specialized Models: Sequences

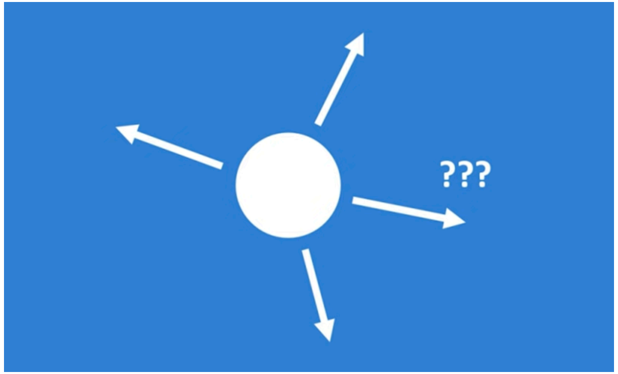
- **Why retaining the context is important**
  - Meaning of a word changes depending on the preceding context

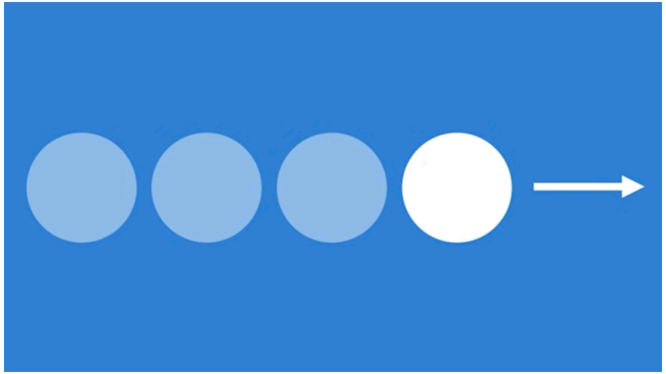
"… He grabbed the bat."



# Needs For Specialized Models: Sequences

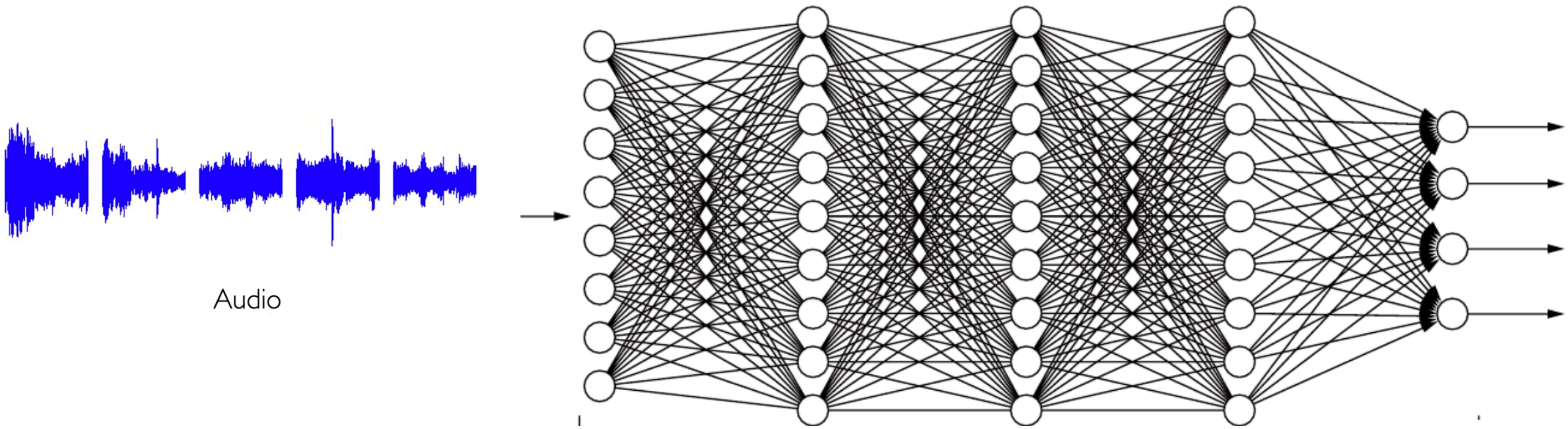
- **Why retaining the context is important**
  - In sequential modeling, information from previous time steps is essential for predicting the next step



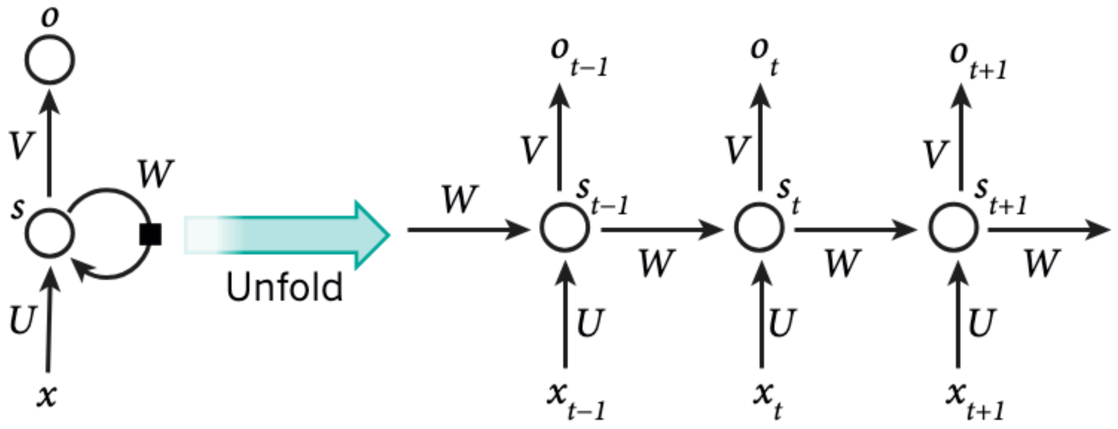
# Needs For Specialized Models: Sequences

- **MLPs cannot retain store contextual information**
  - Since sequences have variable lengths, must split to fit fixed-length input layers
  - Each input is treated as independent information



# Recurrent Neural Networks

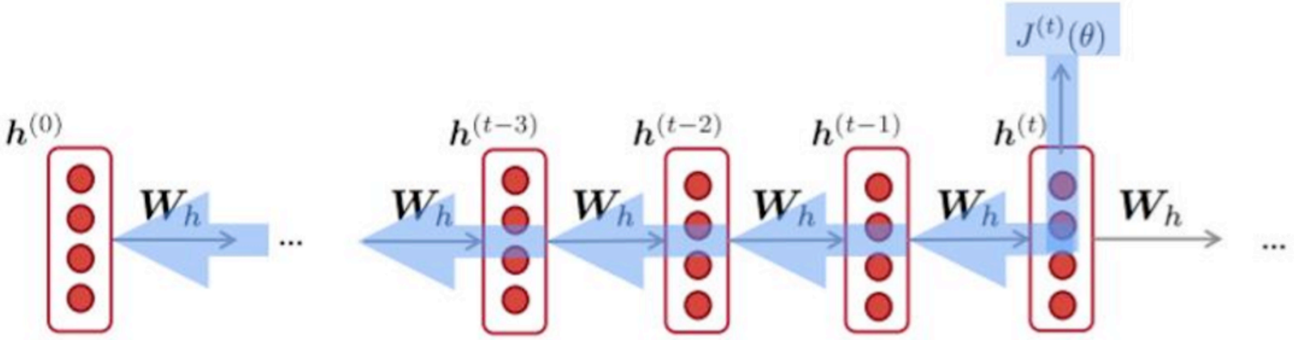
- Models for handling sequential inputs, such as speech and language
- Hidden units called state vector
  - Contains information about the history of all the past elements
- Same weights are used at each time step
  - Provides time invariance



# Limitation of RNNs

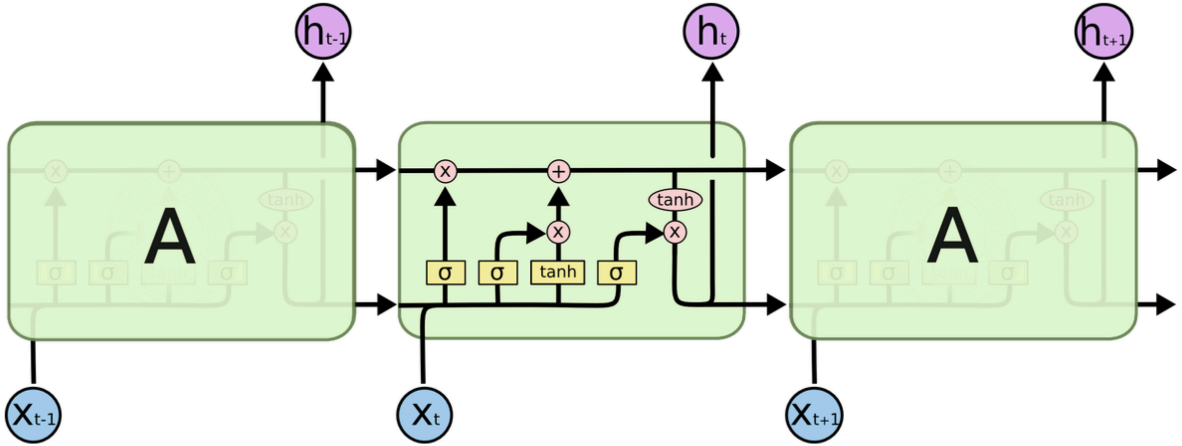
## ❑ Long-Term Dependencies

- Difficult to learn to store information for very long
- In back-propagation, the weight matrix W is multiplied repeatedly
- |W|<1, vanishing gradients occur, |W|>1, exploding gradients occur



# Long Short-Term Memory

- Designed to learn long-term dependencies by using gates and memory cell
- Memory cell is a gated accumulator that preserves information across many time steps

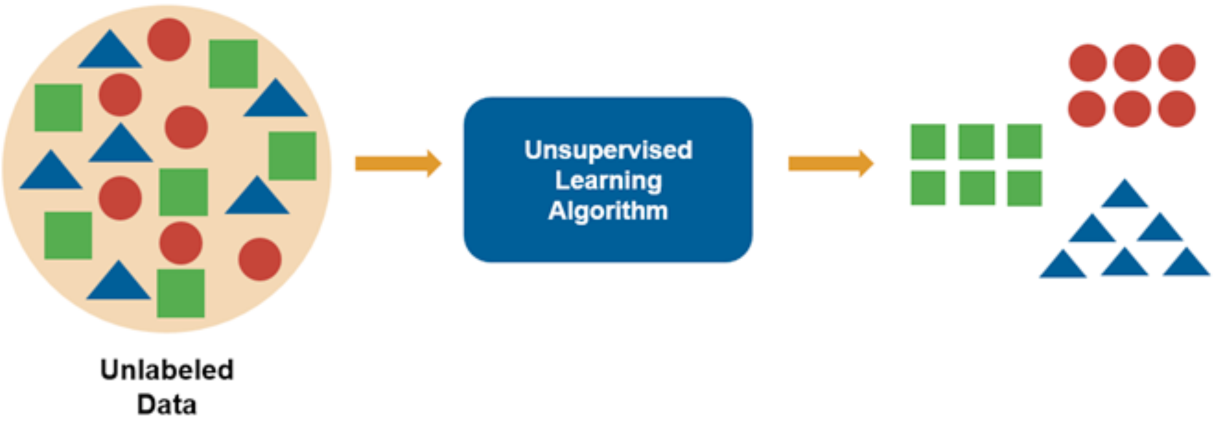


# INDEX

# Future of Deep Learning

## ❑ Unsupervised learning

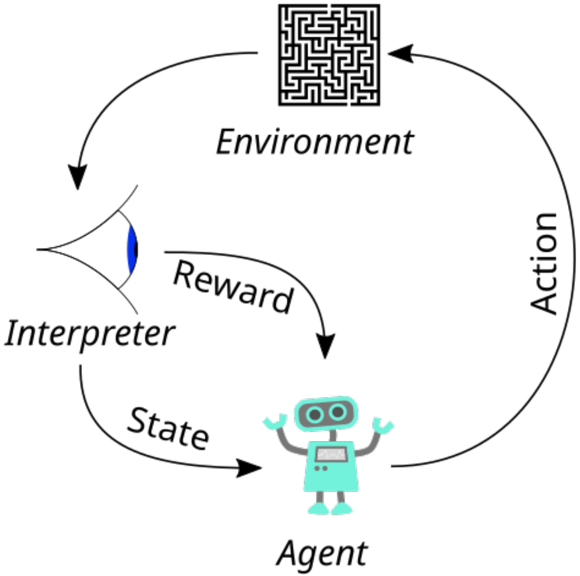
- Human and animal learning is largely unsupervised
- Humans discover the structure of the world by observing it, not by being told



# Future of Deep Learning

## ❑ Reinforcement Learning

- An agent interacts with its environment, takes actions, and learns from rewards to find the optimal policy
- E.g., an active vision system learns where to look to maximize information



# Thank you!