Fundamentals of Artificial Intelligence

Deep Learning

Shyamanta M Hazarika
Mechanical Engineering
Indian Institute of Technology Guwahati
s.m.hazarika@iitg.ac.in

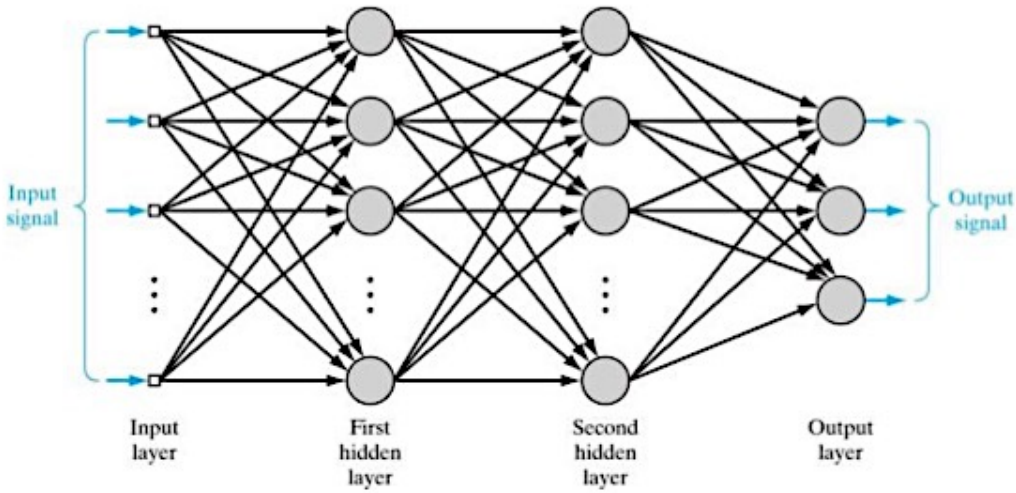
http://www.iitg.ac.in/s.m.hazarika/

Artificial Neural Network

- An artificial neural network comprises many interconnected, simple functional units, or neurons that act in concert as parallel information-processors, to solve classification or regression problems.
- Artificial neural networks (ANNs), now one of the most widely-used approaches to computational intelligence, started as an attempt to mimic adaptive biological nervous systems
- ANNs have been studied for more than 70 years; during which time they have waxed and waned in the attention of researchers; have made a strong resurgence again.

Artificial Neural Networks

1. Interact with the surrounding environment by using one layer of neurons to receive information (Units are known to be part of the input layers of the ANN)
2. Pass information back-and-forth between layers within the black-box for processing by invoking certain design goals and learning rules (Units are known to be part of the hidden layers of the ANN)
3. Relay processed information out to the surrounding environment via some of its atomic units (these are known to be part of the output layers of the neural network).



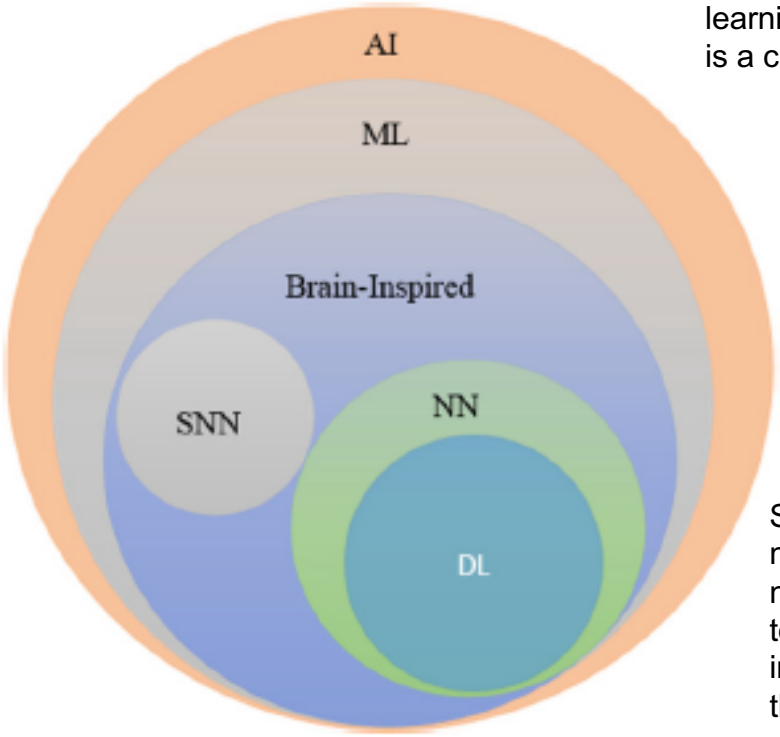
How do these Networks Learn?

- Neural networks are capable of learning - By changing the distribution of weights, it is possible to approximate a function representative of the patterns in the input.
- The key idea is to re-stimulate the black-box using new excitation (data) until a sufficiently well-structured representation is achieved.
  - Each stimulation redistributes the neural weights - hopefully in the right direction.
  - Learning then, is the aggregation of a variable length of causal chains of neural computations seeking to approximate a certain pattern recognition task through linear/nonlinear modulation of the activation of the neurons across the architecture.

Deep Neural Networks

- Multi-layer neural networks have been around through the better part of the latter half of the previous century.
- The term ’deep’ in this context is a direct indicator of the space complexity of the aggregation chain across many hidden layers to learn sufficiently detailed representations.
- Deep Neural Networks has grown in light of its ability to scale with input data and its capacity to generalize across problems with similar underlying feature distributions.

The Taxonomy of AI



DL which uses either deep architectures of learning or hierarchical learning approaches), is a class of ML

Spiking neural networks are artificial neural networks that more closely mimic natural neural networks. In addition to neuronal and synaptic state, SNNs incorporate the concept of time into their operating model.

The taxonomy of AI. AI: Artificial Intelligence; ML: Machine Learning; NN: Neural Networks; DL: Deep Learning; SNN: Spiking Neural Networks

Feature Learning

- A key difference between traditional ML and DL is in how features are extracted.
- Traditional ML approaches use handcrafted engineering features by applying several feature extraction algorithms, and then apply the learning algorithms.
- Additionally, other boosting approaches are often used where several learning algorithms are applied to the features of a single task or dataset and a decision is made according to the multiple outcomes from the different algorithms.

Feature Learning

- On the other hand, in the case of DL, the **features are learned automatically** and are **represented hierarchically in multiple levels**. This is the strong point of DL against traditional ML approaches.

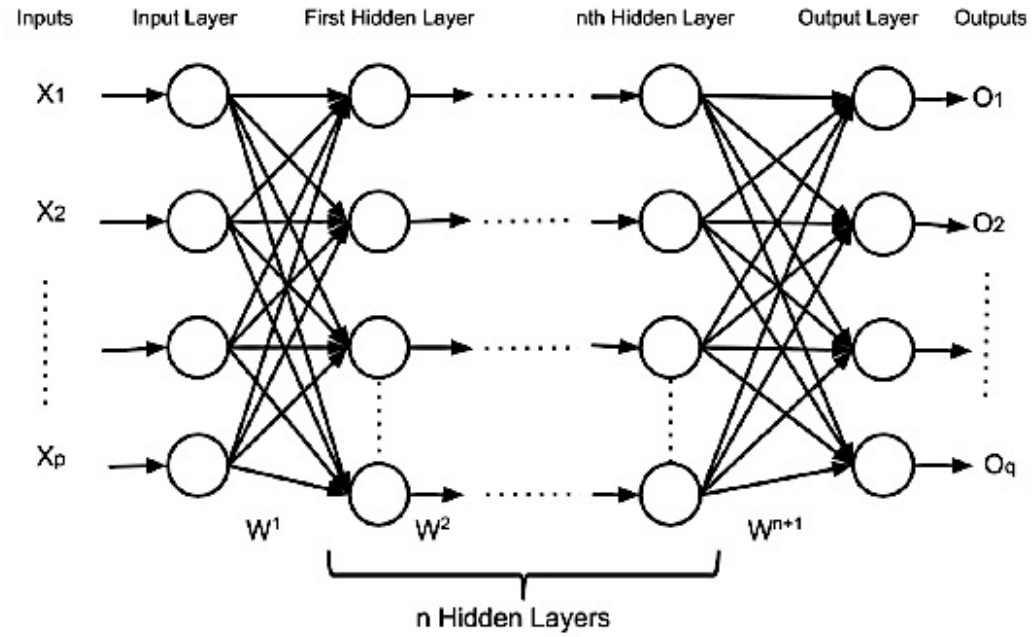
Approaches	Learning steps				
Rule-based	Input	Hand-design features	Output		
Traditional Machine Learning	Input	Hand-design features	Mapping from features	Output	
Representation Learning	Input	Features	Mapping from features	Output	
Deep Learning	Input	Simple features	Complex features	Mapping from features	Output

Deep Architectures

- There are numerous deep architectures available in the literature. Comparison of architectures is difficult as different architectures have different advantages based on the application and the characteristics of the data involved.
  - In vision, Convolutional Neural Networks, and for sequences and time series modelling Recurrent neural networks is preferred.
- Deep learning is a fast evolving field. Every year various architectures with various learning algorithms are developed to create human-like efficient machines in different domains of application

Deep Feed-forward Networks

Multilayer neural network containing multiple numbers of hidden layers, we call it deep neural network



Deep Feed-forward Neural Network with n Hidden layers, p input units and q output units with weights W.

Deep Feed-forward Networks

- Multiple hidden layers help in modelling complex nonlinear relation more efficiently compared to the shallow architecture.
- A complex function can be modelled with less number of computational units
  - compared to a similarly performing shallow network due to the hierarchical learning possible with the multiple levels of nonlinearity.
- Backpropagation using gradient descent is the most common learning algorithm used to train this model.

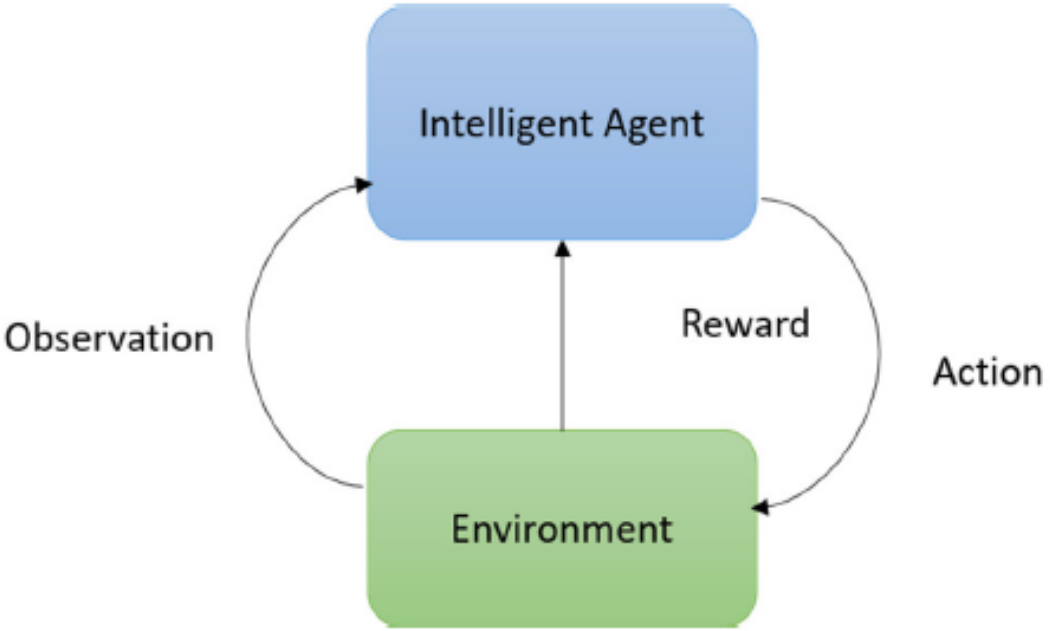
Deep Reinforcement Learning

Q-learning is defined as a model-free reinforcement learning approach which is used to find an optimal action-selection policy for any given (finite) Markov Decision Process (MDP).

Use a neural network (particularly DNN) as an approximation instead of the state table.

The inputs of DNN are the state and action and the outputs are numbers between 0 and 1 that represent the utility encoding the states and actions properly.

That is the place where the deep learning approaches contribute to making better decisions with respect to the state information.



DRL is a learning approach which learns to act with general sense from the unknown real environment

Deep Reinforcement Learning

- According to the learning strategy, the RL technique is learned through observation.
  - For observing the environment, the promising DL techniques include CNN, RNN depending upon the observation space.
- As DL techniques encode data efficiently, therefore, the following step of action is performed more accurately.
- According to the action, the agent receives an appropriate reward respectively. As a result, the entire RL approach becomes more efficient to learn and interact in the environment with better performance.

Why Deep Learning?

- Universal Learning Approach
  - The DL approach is sometimes called universal learning because it can be applied to almost any application domain.
- Robust
  - Deep learning approaches do not require the precisely designed feature. Instead, optimal features are automatically learned for the task at hand. Robustness to natural variations of the input data is achieved.
- Generalization
  - The same DL approach can be used in different applications or with different data types - transfer learning. In addition, approach is helpful when do not have sufficient available data

Why Deep Learning?

- Scalability
  - The DL approach is highly scalable.
  - Microsoft invented a deep network known as ResNet. This network contains 1202 layers and is often implemented at a supercomputing scale.
  - There is a big initiative at Lawrence Livermore National Laboratory (LLNL) in developing frameworks for networks like this, which can implement thousands of nodes.

In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance.

Applications

Deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation,

1. Automatic speech recognition
2. Image recognition; Image restoration.
3. Natural language processing
4. Drug discovery and toxicology
5. Recommendation systems
6. Bioinformatics; Medical Image Analysis
7. Forecasting.

Applications

Deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation,

1. Automatic speech recognition
2. Image recognition; Image restoration.
3. Natural language processing
4. Drug discovery and toxicology
5. Recommendation systems
6. Bioinformatics; Medical Image Analysis
7. Forecasting.

Automatic machine translation has been around for a long time, but deep learning is achieving top results in two specific areas:
1. Automatic Translation of Text.
2. Automatic Translation of Images.
Text translation can be performed without any preprocessing of the sequence, allowing the algorithm to learn the dependencies between words and their mapping to a new language.

Applications

Deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation,

1. Automatic speech recognition
2. Image recognition; Image restoration.
3. Natural language processing
4. Drug discovery and toxicology
5. Recommendation systems
6. Bioinformatics; Medical Image Analysis
7. Forecasting.

Deep learning can be used to use the objects and their context within the photograph to color the image, much like a human operator might approach the problem.

Challenges with Scarcity of Data

- With growing availability of data as well as powerful and distributed processing units Deep Learning architectures can be successfully applied to major industrial problems.
  - However, deep learning is traditionally big data driven and lacks efficiency to learn abstractions through clear verbal definitions if not trained with billions of training samples.
- To make DL work with smaller available data sets, some of the approaches in use are
  - data augmentation, transfer learning, recursive classification techniques as well as synthetic data generation.
  - One shot learning is also bringing new avenues to learn from very few training examples

Adopting Unsupervised Approaches

- A major thrust is towards combining deep learning with unsupervised learning methods.
- Systems developed to set their own goals and develop problem-solving approaches are the future research directions surpassing supervised approaches requiring lots of data apriori.
- Thrust of AI research including Deep Learning is towards Meta Learning, i.e., learning to learn
  - Involves automated model designing and decision making capabilities of the algorithms.
  - It optimizes the ability to learn various tasks from fewer training data.

Influence of Cognitive Neuroscience

- Inspiration drawn from cognitive neuroscience, developmental psychology to decipher human behavioral pattern are able to bring major breakthrough in applications such as
  - enabling artificial agents learn about spatial navigation on their own which comes naturally to most living beings
- To build super-intelligent machines, we must gain a deeper understanding of the human brain. Equally, exploring AI can help us gain a better understanding on what's going on in our own heads.