Deep Learning: The Basic Concepts and how the Networks are Organized and Trained

Course: 4800 Senior Seminar

Author: Mikayla Webber

Abstract

Deep learning, which is a subdivision of machine learning and subsequently a subdivision of artificial intelligence, was first introduced by Rina Dechter in 1986 although it was noted to have first been theorized in the early 1980’s. Recently, there has been a great deal of progress made within the deep learning field. This is partially due to the recent advancements in machine learning and information processing, the increase in chip processing abilities, and the increase in the size of data used for training.

To understand deep learning, it is first important to gain background knowledge on artificial intelligence and machine learning, which deep learning was derived from. Artificial intelligence is a specialized area of computer science with the goal to create intelligence machines. Research in artificial intelligence is dependent on two main parts: knowledge engineering and machine learning.

Machine learning is a branch of artificial intelligence that uses statistical techniques to help give computer systems the ability to learn with data but without being explicitly programmed. Machine learning tasks can be classified into three main categories: supervised learning, semi-supervised learning, or unsupervised learning.

Feature engineering, a core concept of deep learning, contributes to making the algorithms used in machine learning work and is not a single thing but an encompassment of many things. Some of which include feature importance, feature selection, feature extraction, feature construction, and feature learning.

Deep learning uses a model of computing that was inspired by the structure of the brain called artificial neural networks. Artificial neural networks, in general, contain many nodes that are connected and each one of these nodes perform a simple mathematical operation. Artificial neural networks depend on neurons. Neurons have a set of inputs which are given a weight and using some function they perform a calculation on those inputs. A neuron can be either linear or sigmoidal. If it is a sigmoidal neuron then it uses a specific logistic function which will return a value between zero and one. After performing the calculation, the neuron will then transmit the value to other neurons. Neurons are organized into layers, the first being the input layer, the last being the output layer, and the layers in between are known as hidden layers. When the information travels in only a single direction it is known as a feedforward neural network. However, with recurrent neural networks, the information can travel in both directions.

Regression analysis, or more specifically logistic regression, is important when it comes to neural networks. Regression analysis can estimate a relationship between input variables to predict the outcome variable. Logistic regression is like this neural network; however, it does not have any hidden layers, so it is easier to interpret and considered more reliable.

Deep learning, which is a subdivision of machine learning and subsequently a subdivision of artificial intelligence, was first introduced by Rina Dechter in 1986. Over the past several years deep learning has drastically increased in popularity. This is partially due to the recent advancements in machine learning and information processing, the increase in chip processing abilities, and the increase in the size of data used for training. Deep learning is based on learning representations of data instead of using task-specific algorithms and was inspired by the structure of the brain called artificial neural networks. Deep learning, although it does embody machine learning, has made many advances over machine learning. The main difference between machine learning and deep learning is within the performance. With the era of Big Data, performance with such a large amount of data is crucial and deep learning has excelled expectations. Some of the core concepts of deep learning are machine learning, feature engineering, and feature learning. These core concepts will then lead to some of the fundamental concepts which include artificial neural networks (also referred to as just neural networks), feedforward neural networks, logistic regression, and recurrent neural networks.

To understand deep learning, it is first important to gain background knowledge on artificial intelligence and machine learning, which deep learning was derived from. Artificial intelligence (AI) is a specialized area of computer science. The term artificial intelligence was coined in 1956 by a Stanford researcher, John McCarthy. That same year, 1956, at a conference in Dartmouth the key mission of the artificial intelligence field was defined. The core mission or goal is to create intelligent machines, or in other words, machines that can think. Based on how the researchers try to accomplish this goal they can belong to one of two distinct categories: strong AI or weak AI. For some researchers the aim was to build a system that would simulate human reasoning, thereby, thinking the same way that people do. Once achieved, the outcome can be used to not just build systems that think but to also describe how humans think as well. This is known as strong AI. However, with weak AI, the aim is to build an intelligence machine but the resulting system will not tell us how humans think. Recently, there has been another category emerging, an in-between of strong AI and weak AI, that deep learning is categorized with. This category is inspired by human reasoning and uses it as a guide. However, the researchers are not motivated by the goal to model it perfectly as with strong AI. In addition to strong AI and weak AI there is another distinction to be made: narrow AI and general AI. Narrow AI is a system that is designed for a precise task while general AI is a system that is designed for the ability to reason in general. This distinction is important when interpreting specific results because a system that can make recommendations based on past behavior will be different from a system that will learn to recognize images from examples.

Research in artificial intelligence is dependent on two main parts: knowledge engineering and machine learning. Knowledge engineering is an intricate part of artificial intelligence and it refers to all the technical, scientific, and social aspects involved in the making of a knowledge-based system. Artificial intelligence must have access to objects, properties, categories, and the relations between all of them to implement knowledge engineering (“What is Artificial Intelligence (AI)?”). Another intricate part of artificial intelligence is machine learning because learning with suitable supervision involves classification and numerical regressions, whereas learning with no supervision requires the ability to identify patterns within streams of input.

Machine learning is a branch of artificial intelligence. Machine learning uses statistical techniques to help give computer systems the ability to learn with data but without being explicitly programmed. In other words, computer systems will be able to learn things and be able to do thing with their own experience, rather than simply following the instructions set by humans. Machine learning originally progressed from the studies of pattern recognition and the computational learning theory within artificial intelligence. The term machine learning was coined in 1959 by Arthur Samuel, a pioneer in the artificial intelligence and computer gaming fields. A more formal definition of machine learning, given by Tom M. Mitchell a Carnegie Mellon University professor, of the algorithms that are studied in the machine learning field is “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*” (Brownlee, “What is Machine Learning?”). This statement offers an operational definition of the machine learning tasks.

Machine learning tasks can be classified into three main categories: supervised learning, semi-supervised learning, or unsupervised learning. Supervised (also referred to as classification or regression), which is used most frequently, is when there is an example data set for the algorithm with the examples labeled. This helps direct the algorithm in the right direction. With unsupervised learning (also referred to as pattern analysis) it is the opposite; the examples are not labeled so there is no additional guidance for the algorithm. Semi-supervised learning, which is considered a subcategory or class of supervised learning, is when there is more than one algorithm and they all start with a small number of labeled examples. Through communication with each other they share what they think about some sizable amount of unlabeled data and learn from it.

Feature engineering contributes to making the algorithms used in machine learning work and is important in the understanding of deep learning. Feature engineering is a fundamental process and is considered vitally important in applied machine learning although it is considered more of an art rather than a science. Feature engineering is the extraction of patterns from the data that will help make it easier for the machine learning models to differentiate between classes. For example, if there is a machine learning model that wanted to distinguish between the grass and sky in some picture it would take the number of blue pixels versus the number of green pixels as an indicator. This feature would help it limit the number of classes that would need to be considered for a decent classification. The reason feature engineering is considered more of an art rather than a science is because the same features that can be used for one data set might not be able to be used on others.

Feature engineering is not a single thing but an encompassment of many things. To begin, a feature is measurable characteristic. For example, in computer vision a potential feature could be an edge or an object although there are many other possibilities. Features can be ranked by an allocated score that they receive to determine their importance and usefulness, this is known as feature importance. This could be used as a pre-cursor so only the most useful features are chosen (the features with the highest score). A feature may be deemed higher in importance if it is highly associated with the object that is being predicted, or in other words, the dependent variable.

With feature importance, every feature is given a score and then ranked based on the score which will provide details on their importance and usefulness. Feature selection helps further narrow down the list of potentially useful features by removing redundant features and by selecting a subset of the most useful ones based on the problem, all automatically. While some feature selection algorithms use the feature importance method, more advance methods look for the subsets of features by a trial and error system. This more advance method creates and evaluates models to find the most predictive and valuable subset of features automatically (Brownlee, “Discover Feature Engineering”).

In some cases, when there is a large complex data set that can easily include millions of attributes and potential features, which would require a great deal of memory and computational power, an additional process is needed. This process, feature extraction, is needed to reduce the dimensionality and redundancy thereby offering a smaller, more manageable, data set that can be modelled. The key to this process is that the methods can solve the problem of high dimensional data and can do so automatically. Several researchers believe that enhancing feature extraction is crucial to effective model building.

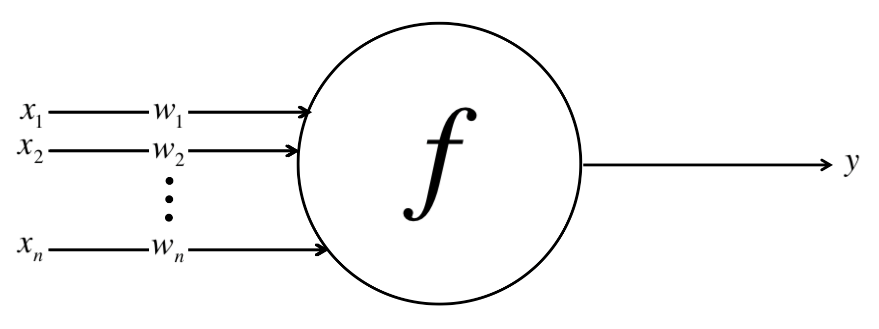
Feature construction, which is a part of feature engineering, is the most talked about as being an artform. Feature construction must be done manually and is a slower process. However, it can make a large impact on the results. While feature importance and feature selection can rank features thereby describing their usefulness, those features must be manually created. This involves determining the structures within the data set and determining what way would be best to reveal them to the predictive modeling algorithms. For example, depending on the data set this could either mean combining, decomposing, or separating features to create new features (Brownlee, “Discover Feature Engineering”). Nevertheless, there are many other options when it comes to feature construction.

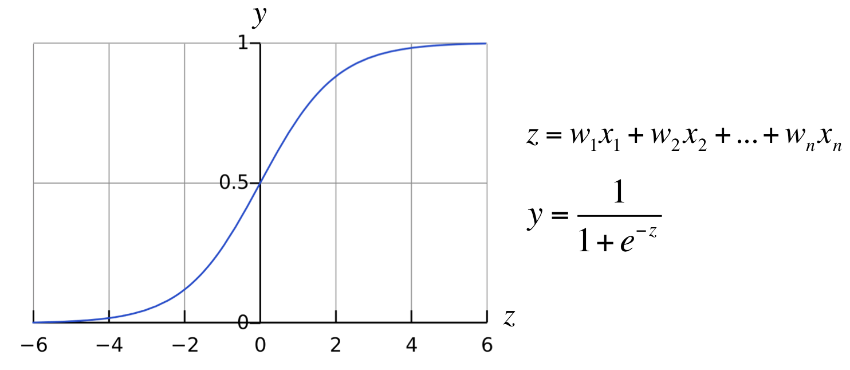
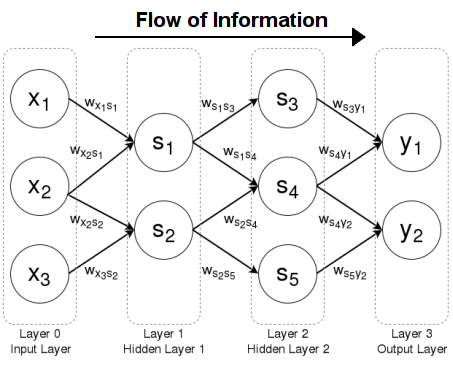
Feature learning is currently theorized but some deep learning methods, such as restricted Boltzmann machines and autoencoders, have begun to make some success. Feature learning hopes to make the task of feature construction and feature extraction less daunting and hopefully automatic. Potentially an easier way to think about feature learning is as feature engineering done by algorithms automatically.

Now with the core concepts of machine learning, feature engineering, and feature learning covered it is onto deep learning. Deep learning, which is also referred to as hierarchical learning, is a subdivision of machine learning and subsequently a subdivision of artificial intelligence. Deep learning originated from artificial neural research and was noted to have first been theorized in the early 1980’s (Deng & Yu 2014). However, it was first officially introduced by Rina Dechter in 1986. Deep learning is based on learning multiple layers of representations which correspond to a hierarchy of concepts, factors, or features where the higher-level concepts are defined from the lower-level concepts, and the same lower-level concepts can help to define many of the higher-level concepts (Deng & Yu 2014). Basically, deep learning is based on learning multiple layers of representations that can help understand data such as images, sounds, or text. Recently, deep learning has increased in popularity due to the advancements in machine learning and information processing, the increase in chip processing abilities, and the increase in the size of data used for training (Deng & Yu 2014). These progressions have helped enabled the effective use of labeled and unlabeled data and learn feature representations (both distributed and hierarchical).

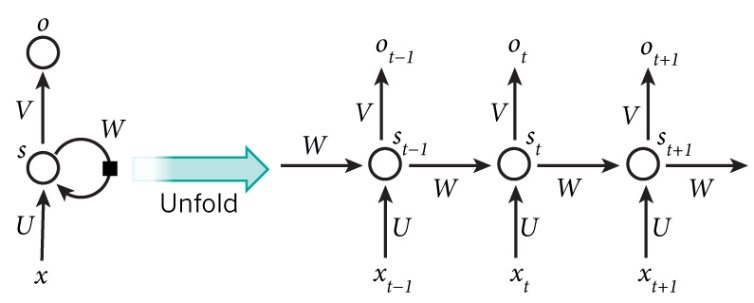
Deep learning is considered a technique that implements machine learning. There are many ways in which deep learning and machine learning contrast. While machine learning uses statistical techniques to help give computer systems the capability to learn with data but without being explicitly programmed, deep learning interprets data features and the relationships by using neural networks that pass relevant information through stages of data processing. With machine learning, there are usually various algorithms that are directed to examine the different variables in the data sets. However, with deep learning once the algorithms are implemented they are typically self-directed for the data analysis. Another difference is that machine learning usually has a few thousand data points that are used for the analysis. Whereas, with deep learning, there are usually a few million data points (Badkar). The main difference, however, is in the performance. While the data set is small, deep learning does not perform anymore optimally than other older learning algorithms but when the data set is large the deep learning performance surges.

Deep learning uses a model of computing that was inspired by the structure of the brain called artificial neural networks, which are also referred to as just neural networks or neural net. Artificial neural networks, in general, contain many nodes that are connected and each one of these nodes perform a simple mathematical operation. The output of the node is dependent upon the mathematical operation and the set of parameters that are specific to that node. By connecting these nodes, complex functions can be calculated and learned.

A neural network is based on the single foundational concept of a neuron which is illustrated on the left (Buduma). Essentially, each of the neurons in the neural network have a set of inputs which are given a weight, in the illustration above this would be *x1...n* and *w1...n*. The neuron, using some function, then performs a calculation on these inputs, which would be *f* in the illustration. After the calculation is performed, the neuron will then take a combination of the weighted inputs, which would be *y* in the illustration above. However, this will only occur if it is a linear neuron.

The other option besides a linear neuron is a sigmoidal neuron which is considered more complicated. With a sigmoidal neuron, instead of using some function to perform the calculation on the inputs, it uses a logistic function (illustrated below) and it uses a weighted sum of the inputs as it’s input. With this function, a value between zero and one will be returned. Therefore, when the returned value is closer to zero that means the weight sum was negative. Since sigmoidal neurons offer more versatility when it comes to using learning algorithms than linear neurons they are usually the preferred choice. In the illustration above, the *y =* where is known as the logistic sigmoid function and is one of the ways a sigmoidal neuron can be distinguished from a linear neuron. However, whether it is a linear neuron or a sigmoidal neuron both will transmit the value it computes to other neurons as its output. Similarly, to neurons in the brain, the neurons in the neural network are organized in layers. The neurons in the bottommost layer will receive signals from the inputs that get transmitted to other neurons and when then reach the topmost layer the neurons there will have their outlets connected to the output layer. The layers of neurons which are between the input layer and the output layer are commonly referred to as hidden layers. Usually, there are no connections between neurons within the same layer. The neural network just described is what is known as a feedforward neural network and is mostly used for supervised learning (illustrated on the left). This type of neural network was the first type of artificial neural network invented and is named feedforward because the information only travels in one direction, forward (McGonagle, “Feedforward Neural Networks”). The illustration above might help clarify what has been described. In this illustration, *x1*, *x2*, and *x3* are the inputs which are then transmitted to the first hidden layer, *s1* and *s2*, when are then transmitted to the second hidden layer, *s3*, *s4*, and *s5*, which then get transmitted to the output layer, *y1* and *y*2.

Regression analysis, more specifically logistic regression, is important when it comes to neural networks. Regression analysis can estimate a relationship between input variables to predict the outcome variable. Logistic regression, which applies to the logistic sigmoid function discussed above, uses the given input variables to predict an outcome variable that can take on one of a limited set of class values. Logistic regression relates to the logistic sigmoid function because it applies this function to the weighted input values and then generates a prediction of which class the input data belongs with. If it is generating a prediction to determine which of two classes the input data belongs to then it is called logistic regression (Dettmers). However, if there are multiple classes then it is called multinomial logistic regression. Logistic regression is like this neural network; however, it does not have any hidden layers, so it is easier to interpret and considered more reliable. Logistic regression is considered more reliable because if some of the properties for the input variables hold true then one can generate a reliable model with little input data. This can be beneficial when there is little data available. However, since it is fast and simple it can be also be used for large data sets.

Recurrent neural networks (RNNs) are artificial neural networks that have directed cycles within the computation graph (an illustration of a simple recurrent neural network and the unfolding of the calculations involved is displayed on the next page). It is called a recurrent neural network because they can perform the same calculation for every element of a sequence, however, the output is dependent upon the previous calculations. This is unlike feedforward neural networks that only have information being transmitted in one direction. With recurrent neural networks the information can travel in loops from each layer, therefore, the state of the model is determined by the previous state (McGonagle, “Recurrent Neural Network”). Another contrast between feedforward and recurrent neural networks is that recurrent neural networks have a memory. This is because the hidden layers in a recurrent neural network can have connections to themselves and this allows the model to store information about the past calculations and thereby, model sequences of input and output pairs.

Deep learning, a subdivision of machine learning, is based on learning numerous layers of representations that can help understand data such as images, sounds, or text. Deep learning, although it does embody machine learning, has made many advancements over machine learning. The main difference between machine learning and deep learning is within the performance. With the era of Big Data, performance with such a large amount of data is crucial and deep learning excelled expectations. Deep learning, as well as machine learning tasks, can be classified as supervised learning (being the most common), semi-supervised learning, and unsupervised learning. Some of the core concepts of deep learning are machine learning, feature engineering, and feature learning. Feature engineering, which contributes to making the algorithms used in machine learning work, is the extraction of patterns from the data that will help make it easier for the machine learning models to differentiate between classes. Feature learning, although it is currently theoretical, can be thought of as feature engineering done by algorithms automatically. These core concepts then led us into the fundamental concepts of artificial neural networks and how a neuron can receive information from an input layer, perform some calculation, then transmit that information to other hidden layers until it reaches the output layer. When all of the information goes in a single direction, this is known as a feedforward neural network. However, if the information can go in both direction then it is known as a recurrent neural network. Logistic regression is like this neural network; however, it does not have any hidden layers, so it is easier to interpret and considered more reliable. This is but a brief insight into an extremely vast field where every day there are more networks and architectures being created. Deep learning has great potential and is being applied more and more to everyday life. It can be found in search engines (both text and image search), photo tagging, speech and facial recognition applications, natural language processing, automated email marketing, online advertising, and with the increase in popularity deep learning will be found in many more places.

Sources

Badkar, Akshay. “Difference Between Machine Learning and Deep Learning.” *Iamwire*,

21 Nov. 2017.

Brownlee, Jason. “Discover Feature Engineering, How to Engineer Features and How to Get

Good at It.” *Machine Learning Mastery,* 26 Sept. 2014.

Brownlee, Jason. “What is Machine Learning?” *Machine Learning Mastery*, 17 Nov. 2013.

Buduma, Nikhil. “KDnuggets.” *KDnuggets Analytics Big Data Data Mining and Data Science*. Chandra, B., and Rajesh K. Sharma. “Fast learning in Deep Neural Networks.” *Elsevier*,

vol. 171, 2 Dec. 2014, pp. 1205–1215.

Deng, Li, and Dong Yu. *Deep Learning: Methods and Applications*. Now, 2014.

Dettmers, Tim. “Deep Learning in a Nutshell: Core Concepts.” *NVIDIA Developer Blog*,

3 Nov. 2015.

Hinton, Geoffrey E., et al. “A Fast Learning Algorithm for Deep Belief Nets.”

*Neural Computation*, vol. 18, no. 7, July 2006, pp. 1527–1554.

McGonagle, John, and Christopher Williams. “Recurrent Neural Network.” *Brilliant Math &*

*Science Wiki*.

McGonagle, John. “Feedforward Neural Networks.” *Brilliant Math & Science Wiki*.

Schmidhuber, Jürgen. “Deep learning in neural networks: An overview.” *Neural Networks*,

vol. 61, Jan. 2015, pp. 85–117.

Schmidhuber, Jürgen, et al. “Deep Learning.” *Scholarpedia*, vol. 10, no. 11, 2015.

“What Is Artificial Intelligence (AI)? – Definition from Techopedia.” *Techopedia.com*.