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

Course: 4800 Senior Seminar

Author: Mikayla Webber

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.

References

Chandra, B., and Rajesh K. Sharma. “Fast learning in Deep Neural Networks.” *Elsevier*,

vol. 171, 2 Dec. 2014, pp. 1205–1215., doi:10.1016/j.neucom.2015.07.093.

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

*Neural Computation*, vol. 18, no. 7, July 2006, pp. 1527–1554., doi:10.1162/neco.2006.18.7.1527.

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

vol. 61, Jan. 2015, pp. 85–117., doi: 10.1016/j.neunet.2014.09.003.

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

doi:10.4249/scholarpedia.32832.

Sutskever, Ilya, et al. “On the Importance of Initialization and Momentum in Deep Learning.”

*Proceedings of the 30th International Conference on Machine Learning*, Edited by

Sanjoy Dasgupta and David McAllester, vol. 28, no. 3, 17 June 2013, pp. 1139–1147. *Proceedings of Machine Learning Research*, proceedings.mlr.press/v28/sutskever13.html.