**Deep Learning using Big Data**

Research Question:

An exploration of pruning on different types of Neural Networks (ANNs, CNNs, RNNs).

**Overview**

Contrary to popular belief, and media accounts emphasizing the similarity of DL to the brain, rough guidelines are drawn from neuroscience to the fundamentals of DL. The basic idea of having many computational units that become intelligent via interaction with each other is inspired by the brain. While neuroscience is an important source of inspiration, it need not be taken as a rigid guide. We know actual neurons compute very different functions than modern rectified linear units, but greater neural realism has not yet led to an improvement in machine learning performance. Also, while neuroscience has successfully inspired several neural network architectures, we do not yet know enough about biological learning for neuroscience to offer much guidance for the learning algorithms we use to train these architectures (**SUMMARIZE**).

The aim of this paper is to conduct research into whether pruning methods for neural networks hamper model performance with a specific application to financial market sentiment.

This paper takes a primary focus on the application of pruning methods on different types of Neural Networks. The motivation behind this research focus was to assess the tradeoff between model size reduction, and hence run time of neural networks, versus model performance. The three types of Neural Networks explored in this research paper are Artificial Neural Networks (ANNs), Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs).

The purpose of this literature review is to summarize and comment on existing literature in the areas of; applications of pruning methods related to different neural network architectures, deep learning/pruning methods for sentiment analysis, and methods and tools for big data storage. Following a critical evaluation of sources, the aim is to outline a primary motivation for research, assess deep learning model architectures (pruning methods) related to a specific application field, analyze the data storage requirements of this project and to identify any potential research gaps in these fields.

Sentiment is associated with different attributes and, as such, the literature points towards no universal definition. Schleifer and Summers (1990) define sentiment as an investors overall attitude toward financial markets, DeLong et al. (1990) defines it as investors’ formation of beliefs about future cash flows and investment risks that are not justified by existing facts, and Brown and Cliff (2004) are of the opinion it represents the ‘expectations of market participants relative to a norm: a bullish (bearish) investor expects returns to be above (below) average, whatever average may be’. A myriad of research papers comment on the efficient market hypothesis (EMH), the rational behavior of investors, questioning the ‘random walk of future asset prices’ and whether or not other factors can act as future indicators for stock price movements. Behavioral finance argue that sentiment proxies may be candidates for explaining subtle pricing anomalies. Lee et al. (1991) argues that sentiment may be a candidate explanation for different asset valuations across investors. Trading strategies such as the one by Zhang and Skiena (2010) have been developed utilizing natural language processors (NLPs) on blog/news data to generate consistent returns for investors. Katayama and Tsuda (2020) investigated whether a trading strategy based on a sentiment analyzer using deep learning can be profitable using Nikkei Telecom from the Nikkei Shimbun morning edition.

Deep learning (DL) is the subfield of artificial intelligence that focuses on creating large neural network models that are capable of making accurate data-driven decisions, particularly suited to contexts where the data available is complex and large (Kelleher, DL). There have been three waves of developments in DL: cybernetics in the 1940s-1960s, connectionism in the 1980s-1990s and the current resurgence beginning in 2006 (Goodfellow, DL). DL has attracted much attention from the academic community due to its state-of-the-art performance in many research domains such as speech recognition (G. E. Dahl, D. Yu, L. Deng and A. Acero, "Context-dependent pretrained deep neural networks for large-vocabulary speech recognition", IEEE Trans. Audio Speech Lang. Process., vol. 20, no. 1, pp. 30-41, Jan. 2012), (G. Hinton et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups", IEEE Signal Process. Mag., vol. 29, no. 6, pp. 82-97, Nov. 2012.), collaborative fultering (R. Salakhutdinov, A. Mnih and G. Hinton, "Restricted Boltzmann machines for collaborative filtering", Proc. 24th Int. Conf. Mach. Learn., pp. 791-798, 2007.), and computer vision (D. Cireşan, U. Meler, L. Cambardella and J. Schmidhuber, "Deep big simple neural nets for handwritten digit recognition", Neural Comput., vol. 22, no. 12, pp. 3207-3220, 2010.),(M. Zeiler, G. Taylor and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning", Proc. IEEE Int. Conf. Comput. Vis., pp. 2018-2025, Nov. 2011.). Artificial Neural Networks (ANNs) consist of an input layer, an output layer and a specific number of hidden layers. Neurons are contained in the input and output layers related to the number of input and output parameters. It is then an objective to determine the number of hidden layers and the number of neurons in each hidden layer. Activation functions are applied by each neuron in the hidden and output layers to the weighted sum of inputs.