

6.867 Term Project: An Exploration of Deep Learning and Convolutional Neural Nets for Image Classification

Kathryn Evans, Andres Hasfura and Remy Mock

November 22, 2015

1 Introduction

Neural networks are a useful machine learning framework that's primary benefit is that instead of specifying the basis functions relating input to output, more or less learns the connectivity, since the optimal basis functions could be complicated and non-intuitive, such as in the case of images. Such an approach is Artificial Neural Networks (ANN). The idea of a neural network replaces the known basis functions with features, which are parametric functions of activations, learned from the input data in the first layer of learning, and then utilize a second layer to learn the relationship of the output data from those selected features.

Through this course, a simple singular layer artificial neural network was presented. However, much like the neural networks in the human brain which provided the inspiration for artificial neural networks, the architecture used for machine learning neural networks can be more complicated and complex than a single hidden layer. The idea of utilizing many layers is known as deep learning. Increasing layers, while drastically more parameters and computation, allows for more complex input/output relationship and

an ability to classify based on both information from low and high level features.

Deep learning, although not a recent idea, has recently exploded in popularity due to rise in labeled data and general purpose GPU programming and is revolutionizing very important subfields within artificial intelligence. Machine learning, machine vision, and natural language processing are examples of area in which the use of deep learning has produced large jumps in performance on difficult test sets. Deep nets are now being used anywhere from pedestrian detection for autonomous vehicles [1], to facial expression recognition [2] to classifying whether or not a selfie is good [3].

Not only do deep nets contain more hidden layers but a multitude of different types of layers. Each layer type has different connectivity and objectives allowing for a greater richness of information. For use with images convolutional layers are especially beneficial for examining only the relevance of spatially nearby pixels in the context of determining features. Arrangements of convolutional layers as well as other types of layers leads to Convolutional Neural Networks (CNN).

In this project, we want to explore the benefits of deep nets as well as convolutional neu-

ral networks for image classification. Our goal is to generalize the benefit on image classification performance of firstly, additional fully connected hidden layers and secondly, more complicated convolutional layers and architectures. Not only do we want to gain familiarity with concept and tools used in deep learning but we also hope to document generalities by applying these approaches to multiple image datasets. The generalities we hope to explore include the following: How does the number of layers effect the amount of training data needed? How to choose architectures?

2 Approach and Methodology

In this project, we seek to implement a version of multilayer neural nets for supervised learning on specifically images. We plan on trying our implementations on datasets of varying difficulty, from MNIST to ImageNet so that we do not have to worry about data generation and labeling and provide generalization about the neural net choices. Because we are specifically interested in applying our implementation to images we hope to see benefit of making the network out of convolution layers, where not all nodes are connected and we will compare this to a implementation with fully connected layers.

We will then compare our rudimentary approach to a professional library for deep learning, specifically TensorFlow. Obviously the professional library will allow for better results and more interesting conclusions but building and implementing a simplified version will give a better understanding into how the neural network works.

2.1 Neural Network Basic

2.2 Multi-layer Neural Network

2.3 General Architecture of a Convolutional Neural Network

The idea of the CNN is to no longer have fully connected layers. Instead not every node is connected to every other node. Instead these relationships are described by a kernel. Of particular interest in the convolutional kernel. But instead of hard coding the kernel, the kernels used are learned by training.

2.3.1 Convolutional Layer

By using a convolutional kernel we can transform the 2-D images into a 3-D space, this is done by convolving the images by several kernels. For a $M * N$ size image, using k kernels that are $m * n$, after convolution, there are $k * (M - m + 1) * (N - n + 1)$ sub images. Using this method, we decrease the size of each image, but at the same time we improved the feature by increase the influence of nearby pixels.

2.3.2 Re-LU Layer

2.3.3 Pool Layer

2.4 Professional Libraries

2.4.1 TensorFlow

2.5 Datasets Considered

3 Results

3.1 Benchmarking with MNIST

MNIST [4] Already done and thoroughly explored by many people included Yann LeCun [5] however it provides a good system to check our results by since it is so well documented. Without exploring our results for different and being able to concretely compare them to results for different layers and types of layers to published and documented results, we can make sure our implementation is working before taking on a more complicated dataset.

Only considered datasets with no distortions

3.2 Other dataset

References

- [1] M. Szarvas, U. Sakai, and J Ogata. Real-time pedestrian detection using lidar and convolutional neural networks. *Intelligent Vehicles Symposium, 2006 IEEE*, pages 213–218, 2006.
- [2] Wei Li, Min Li, Zhong Su, and Zhigang Zhu. A deep-learning approach to facial expression recognition with candid images. *Machine Vision Applications (MVA), 2015 14th IAPR International Conference on*, pages 279–282, May 2015.
- [3] Andrej Karpathy. What a deep neural network thinks about your selfie. <http://karpathy.github.io/2015/10/25/selfie/>, 2015.
- [4] C.J.C. Burges Y. LeCunn, C. Cortes. The mnist database of handwritten digits. <http://yann.lecun.com/exdb/mnist/index.html>.
- [5] Y. Bengio Y. LeCun, L. Bottou and P. Haffne. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, November 1998.

ANN Type	LeCun Test Error	Our Test Error
2-layer NN, 300 hidden units, mean square error	4.7	
2-layer NN, 1000 hidden units	4.5	
3-layer NN, 300+100 hidden units	3.05	
3-layer NN, 500+150 hidden units	2.95	
Convolutional net LeNet-1	1.7	
Convolutional net LeNet-4	1.1	
Convolutional net LeNet-5	.95	

Table 1: Comparison of results for multilayer ANN and CNN with published results [5].