

## CSE 3521 DEEP RESIDUAL NETS

BRENDAN WHITAKER

I attended the seminar titled *Representation, Optimization, and Generalization in Deep Learning* by **Peter Bartlett** on October 19th, 2017. The problem being addressed was in general any application of deep learning, but most of the examples used by the speaker were centered around computer vision. The deep learning method ties into our course as a part of machine learning which focuses on optimal data representation. For example, the speaker discussed the CIFAR10 datasets used for computer vision testing. Datasets of this type consist of a number of relatively poor-quality images of several categories (i.e. trucks, cats, horses, ships,...). The problem is to learn how to sort these images into the correct categories below a specified error rate in as few training episodes as possible.

The first half of the talk consisted of an introduction to deep residual networks, which are characterized by large numbers of layers of nonlinear function compositions

$$h = h_m \circ h_{m-1} \circ \dots \circ h_1$$

used to approximate an empirical risk minimizer  $h^*$ . In this method, all layers compute near-identity functions:

$$\|h_i - \text{Id}\| = O\left(\frac{\log m}{m}\right),$$

and we see that as  $m$ , the number of layers, tends to infinity, the layer functions become arbitrarily close to the identity. This is responsible for the improvement in error rate as the number of layers increases during training of deep residual nets. The speaker detailed how as the number of layers increased from 20 to 110 in training, the performance increased as opposed to the case with deep plain nets, in which the error rate actually increased from 20 to 56 layers.

This method is entirely dissimilar to that of Watson playing Jeopardy, since the approach of deep learning was in its infancy when Watson was first being developed, and wasn't implemented in Watson when it competed on the show. I learned a lot about what direction researchers are pushing to break new ground on current AI methods, and I discovered how heavily the fields of linear algebra, statistical theory, and machine learning are intertwined.