Benjamin Katz

Due February 20th Week 5

**Perceptron**

Rosenblatt attempts to understand how it is that learning takes place. He is interested in both how he theorizes this occurs in humans, but more importantly, how the concept exists apart from any one particular physical system. If it is possible to explain how learning works in a very abstract mathematical way, the theory could be instantiated in multiple physical systems like a human brain or a 1950’s computer.

Rosenblatt's theory was unique from all other cognitive theories at the time because it had the capacity of being disproved. All other theories did not claim that their model would be able to predict a response given a set of inputs. Rosenblatt did claim as much. His theory includes S-points, A-units and R-cells. S-points perceive the stimuli in the environment and send signals to the A-units, which then sends signals to the R-cells, either probabilistically or deterministically to trigger some response. At this point no learning has occurred. This is simply how Rosenblatt pictures the decision making process. The learning occurs in the feedback sent from the R-cells to the A cells which can adjust the values, path, or probability of the signals. Rosenblatt calls this reinforcement learning.

Rosenblatt is able to create a system where learning can take place given an initial starting point of random connections between A-units and R-cells. The reinforcement learning allows the system to recognize similar stimuli and thus trigger similar responses. This model completely demystifies the human thought process and lays the groundwork for building other thinking systems in the future.

**Backprop**

Rumelhart explores the added benefit of hidden layers when building a learning system. Rumelhart thinks that the classic perception of the perceptron having the input layer directly connected to the output layer is intuitive for learning a response. However, it is not sufficiently powerful for learning when inputs can be similar in different ways.

Backprop proposes assigning every node a weight and a bias. The goal of this system is to get the output layer to look like our desired output layer for as many inputs as possible. By adjusting the weights and biases of each node, that goal can be achieved. Backprop proposes a system for changing those weights and biases such that the error between the actual and the desired is minimized. Backprop does this through gradient descent. It finds the error, and the partial derivative of the error with respect to the weights and could adjust the weights such that it reduces the error proportional to the learning rate.

This research is extremely important because it does not rely on the trial and error or simple reward and punishment method proposed by Rosenblatt in his bivalent system. This method is a much more efficient system for calculating which weights should be adjusted and by how much.

**ImageNet**

ImageNet trained a classifier on a much larger dataset than has ever been used previously. With so many possible classes in which to classify the dataset, overfitting becomes a major concern. This was alleviated in part due to the massive dataset. Because this was such a massive dataset it, in turn, was a massive project. Because of this, ImageNet was forced to implement many optimizing techniques which is a major contribution to continued ML development. They trained using a ReLu activation as opposed to a more typical sigmoid or tan. Because of this, their results converged much faster which is very important when dealing with large datasets. Other optimizations include multiple GPU’s to help with the matrix multiplication, along with other optimizations that made small improvements.

ImageNet also created new techniques to combat overfitting. When dealing with so many classifications overfitting can be a real problem. To combat this ImageNet augments their image data. They train their model on multiple smaller patches of the same image. By taking the average of the predictions of all the softmaxes, the model prevents substantial overfitting. While there were no ground breaking new theories proposed in the ImageNet paper, it non the less shows strategies for building large neural networks. Much of this was found by trial and error and simply cleverly implementing ideas that were already mathematically proven to aid in the building of these models.

ImageNet proved that with enough data, compute, and clever optimization strategies, it is possible to build a highly accurate supervised learning image classifier.

Part 2:

<https://colab.research.google.com/drive/1jhC7dxi4vRLw7HLvZ0KYwGw-KV82BM_8?usp=sharing>