CS 6600 Assignment 2
Brad Jensen A02214597

## Problem 1

*Learning Internal Representations by Error Propagation* by Rumelhart, Hinton and Williams showed that associative networks can handle much more challenging problems than previously thought. The authors point out that networks at the time struggled to solve problems when input and output patterns are significantly different from each other, as seen in the infamous XOR problem. The authors show that by observing the difference between the expected outcome of a system and the actual result, and then propagating that difference through the network, it can develop internal representations needed to solve complicated problems.

I enjoyed reading such a foundational work that laid the groundwork for the most widely used algorithm for neural network training, even today. The authors' excitement about their findings is contagious. When discussing the results of the symmetry problem, they wrote that "this problem was interesting to us because the learning system developed a much more elegant solution to the problem than we had previously considered." They must have been ecstatic to see such results in a time of pessimism for neural networks. Their mathematical explanation of the Generalized Delta Rule is straightforward and easy to follow. The authors provided many examples of problems that backpropagation was able to solve. These examples were easy to understand, and provided me with an approachable way to see how their method can solve problems that simpler models cannot.

Being the introduction of the Generalized Delta Rule and backpropagation, it is natural that not all the limitations were discussed or even known about at the time. Indeed, the authors do acknowledge many opportunities for future research. It is understandable that they were yet to discover issues such as gradient explosion. The paper deals exclusively with much simpler problems than we are used to seeing ANNs tackle today. It would have been interesting for them to at least discuss more complicated issues that could be addressed with back propagation. It is interesting to see them discuss parallel programming as a potential use case, but I would have liked to read more about what the authors felt about the possible implications of their work.

I found the authors' observations about the hidden unit structure fascinating. They discussed minimizing the number of nodes as much as possible, but they also noted that more elegant solutions are often found by giving the network more hidden units than it needs. They could then observe which nodes are being utilized. This has me thinking about ANN layer structure and the possible effects of different numbers of layers and node density per layer. I would like to explore existing research on this topic and experiment with different structures firsthand.