Analysis of "Learning Internal Representations by Error Propagation"

Connor Osborne: A01880782

September 2021

1 Discussion:

In this paper Rumelhart, Hinton, and Williams attempt to prove that a learning theorem for a multilayered machine of perceptrons does exist and that the worry over reaching local minima and thus the system not being able to learn can be mitigated. They do so by demonstrating a generalization of the perceptron learning procedure using a rule called the generalized delta rule. The generalized delta rule is a way of implementing a gradient decent method for finding weights that minimize the sum squared error of the system. This occurs by comparing the output of the system with the desired output and changing the weights of the system based on the result. They then prove that the issue of local minima can be mitigated for a wide variety of learning problems through the use of forward and backward propagation using a set of weights where the weights of the system are not all equal to each other.

2 Likes:

The parts that liked most about this paper were the tables, figures, and explanations of functions and theorems. The tables and figures were all well labeled and had great explanations. Each one was explained in detail while still being clear and concise such that the reader can easily understand the information being conveyed. The article itself also had clear and concise explanations with more details than the headings of the figures and tables which helped build upon information already conveyed. The explanations of the functions and theorems were also very useful as they were clear and concise while also providing a step-by-step transition from one formula to the next. This made the thought process of the writers easy to understand, and it made the knowledge more useful while continuing through the rest of the article.

3 Dislikes:

There was only one thing about this article that I didn't like, but it happens several times throughout this work. The writers had the habit of referencing other works and chapters without explaining what the reference was and sometimes they didn't even describe how it applied to the situation. For example, on page 4 of the article, the writers state that their example generates the standard delta rule as explained in chapters 2 and 11, but we have no reference to those chapters.

4 Inspiration:

The section I found most fascinating was when, in passing, the writers touched on the fact that not all weights need to be variable. The system could still work well if certain weights remained static. This caused me to think it over for a bit before moving on and after testing I came to the conclusion that it was correct but only if there were still some weights that could change. Since the learning process relied on the the propagation and change of weights, without that no learning could take place. My thoughts were proven correct as later in the article the writers demonstrated that the system had to be "symmetry breaking", meaning that without weights that change, nothing is propagated and the system can make no learning progress.

Another topic that the writers discussed in passing I found interesting was when they talked about NNs being run via parallel computing. I took the parallel computing course last year and since we started talking about NNs made of perceptrons I recognized a parallelization problem in these systems. However, as the writers explained and I learned first hand, parallel programming is difficult and time consuming and they demonstrated that the ANNs they discuss and we have been using in class do the same job with a lot less work and coding involved.