Project Proposal Report

Tentative project title: On the Robustness of Activation Functions
Student name: Zachary Harrison
Project summary:

Throughout my data science career, I have been both fascinated and appalled at the apparent and flagrant lack of objective science. There are countless times that I've heard "use your intuition" or "you'll get a feel for it" in reference to choosing good learning rates, number of epochs, number of nodes in layers, the number of hidden layers, etc… Everything about data science is so cutting edge that it's hard to believe that so much of the hyperparameter tuning is done by feel alone.

So, you can imagine my discomfort when I realized that Sigmoid and Tanh, the only activation functions for which I have genuine intuition as to their usefulness, are widely considered the worst activation functions. During this research project, I would love to give myself and others a more in-depth understanding of different activations functions' usefulness (or lack thereof). And if I just so happen to end up showing Sigmoid and Tanh the love they rightfully deserve, then I'm all for it.

The goal of this project is to help future data scientists use activation functions appropriately. That is, to find types of datasets for which activation functions thrive. The way I'm going to do this is by comparing the performance of different activation functions (Sigmoid, Tanh, ReLU, LReLU, PReLU, ELU, MPELU, etc…) on different types of datasets (like Spatio-Temporal, Graph, CNN, RNN, LSTM, etc). I will then do some cross-analysis for different datasets to hopefully find the best activation function.

Finally, I have one last goal of this project: proposing a new activation function. I have not developed it yet, but Dr. Hamdi suggested it as a natural conclusion of this research and I couldn't agree more.

References (list at least 3 papers from the conference or journals of the project description page):

1. Activation functions and their characteristics in deep neural networks
2. The Power of Approximating: a Comparison of Activation Functions
3. How to Choose an Activation Function
4. Small nonlinearities in activation functions create bad local minima in neural networks
5. Neural Networks, Manifolds, and Topology

Note: I read through most of these references and I am PUMPED!!! I don't talk about a lot of them in my powerpoint or this report because they are **extremely** math heavy. As someone who is double-majoring in mathematics and computer science, I'm in love with the idea of using mathematics to find solutions to problems as abstract as data science (despite the rest of the world's obvious distaste for math). And using topology in the context of decision boundaries is like a wet dream for me. I'm super excited to get working on this project!

Availability of dataset/code:
See my github for the dataset/code. Since I haven't actually started the project, there won't be a whole lot of content yet.