Basic Information

Understanding Deep Neural Networks with Rectified Linear Units.

Author: Raman Arora, Amitabh Basu, Poorya Mianjy, Anirbit Mukherjee.

ICLR 2018, Cited by 75.

Summary

This paper is a theoretical work, mainly presents series of results regarding ReLU deep neural networks. It analyses the ReLU DNN through a special set "hard" functions that are not easily represented by shallow networks. First of all, it states any piece-wise linear (PWL) function can be represented by a ReLU-activated DNN. This leads to that any smooth function can be approximated by ReLU DNN. Secondly, the paper states that the expressiveness of a 3-layer DNN is stronger than a 2-layer DNN (with same level of hidden neurons). Then, it prove that exponential many piece-wise linear functions can be represented by ReLU DNN with polynomial number of neurons. This is the main contribution of this paper. At last, it proposed an algorithm to learn the global optimum of such DNN in exponential-time.

Strong points

1. Originality and contribution

It is important to have such theoretical analysis of DNN to push forward the development of machine learning. The main contribution of representation of PWL is interesting. The main contributions of this paper lies in:

- Tighten the previously bounds by Telgarsky et al. for ReLU DNNs on the depth.
- Constructed a family of hard functions, whose affine pieces scale exponentially with the dimensionality of the inputs. The hard functions has a better parameterization and the gap between 3-layer and 2-layer is proved bigger.
- Gave a procedure for searching for globally optimal solution of a 1-hidden layer ReLU DNN with linear output layer and convex loss.
- Presenting a constructive proof to show that ReLU-activated DNN can represent many linear pieces.

Overall, this is an incremental work in the direction of studying the representation power of neural networks.

2. **Writing** This paper is well written. From the introduction part, it tells the whole story in a smooth, detailed way and still give the reader the whole view of the research path of neural networks. It also provided enough background knowledge to readers to understand the paper in a high-level way. In the later parts, the authors use carefully chosen formulas to help understanding. They also explained the algorithm clearly and concretely. Though the paper is a little dense in places, but overall well organized and easy to follow. I really enjoy reading this paper.

Weak points

1. Originality and contribution

The main contribution of this paper is just extending the previous ideas. No new angle is introduced to look at the neural networks.

2. Clearness

It will help to understand the paper if the authors explain what is meant by "hard" functions (e.g. functions that are hard to represent, as opposed to step functions, etc.)

3. Technical details

- In deep neural networks, we usually use other activation function in the final layer other than a linear function. It will be interesting if the authors also considered the other non-linear final activation functions. That is because the ReLU or linear functions are usually not the final results. Wether softmax-like activation functions will affect the convergence or functions' approximation accuracy remain unknown in this paper.
- When the input dimension is n, the representability fails to show that a polynomial number of neurons is sufficient. Perhaps an exponential number of neurons is necessary in the worst case. But in practice, we care more about the average size, or size under certain conditions. It will be more interesting if the authors mentioned this.