COMP-SCI-5567-0001
Deep Learning

Home Work - 4

By
Name: Sai Prudhvi Charan Pothumsetty,
Student ID: 16343752,
Professor: Jesse Lowe,
University of Missouri Kansas City.

Video Link

https://d2y36twrtb17ty.cloudfront.net/sessions/cb8524ae-0a14-4a79-ab09-b14d012a845e/b802c64d-4887-4673-8690-b14d012a8467-70553a39-6f8b-4ee8-b475-b14d018902d3.mp4?invocationId=3bacc3a7-cdf6-ee11-8291-12c206d2fd2b

Github link:

https://github.com/Prudhvicharan/deeplearning_HW4

1. **Low-Dimensional Manifold Exploration**:

- What is the central finding of the paper regarding the training process of deep networks?
  Based on this paper, the authors argue that deep networks training explores a low-dimensional manifold in prediction space. In simpler terms, despite the complexity of deep networks, their training process follows a low-dimensional path.

- How does it relate to the concept of a "low-dimensional manifold"?
  The concept of a low-dimensional manifold is key to understanding the paper's findings. A low-dimensional manifold refers to a lower-dimensional space that can be embedded in a higher-dimensional space. In this context, the paper suggests that the training process of deep networks takes place in a lower-dimensional space, even though the networks themselves exist in a higher-dimensional space.

2. **Information-Geometric Techniques**:

- How do the authors use geometric information techniques to analyze the trajectories of deep network predictions during training?
  The authors leverage information geometry, a mathematical tool, to analyze the trajectories of deep network predictions during training. This approach allows them to gain insights into the behavior of networks with different configurations.

- What insights do believe these techniques provide?
  Information geometry provides valuable insights according to the paper. It helps us understand how different deep networks, despite their variations in configuration, tend to follow similar paths during training, shedding light on the convergence and similarity of trajectories in the prediction space.

3. **Common Manifold for Different Architectures**:

- According to the paper, how do networks with varying architectures, sizes, optimization methods, regularization techniques, and weight initializations behave during training?
  The paper interestingly highlights that networks with varying configurations follow similar paths in the same projected manifold during training. This suggests that the underlying probabilistic models of these networks reside on the same manifold.

- Is it ***surprising*** *to you* that these diverse methods can be shown to follow similar paths in the same projected manifold in the prediction space?
  Based on the paper, it is surprising that networks with varying configurations can follow similar paths during training. This finding suggests that the training process is guided by underlying factors that are independent of the specific network configuration.

4. **Trajectories and Network Initialization**:

- Discuss the behavior of networks **initialized** at different points in the prediction space.
  The paper reveals that networks initialized at different points in the prediction space converge to the truth along a similar manifold. This means that regardless of where a network starts, its training process follows a similar path.

- What factors do the authors claim correlate most strongly with convergence?
  The authors claim that the structure of the dataset, the initialization near a specific point, and the behavior of the optimization method correlate most strongly with convergence. This suggests that these factors play a more significant role in convergence than the specific network configuration.

5. **Implications for Deep Learning**:

- What implications does the existence of an effectively ***low-dimensional manifold*** during training have for the field of deep learning?
  The existence of a low-dimensional manifold during training has significant implications, according to the paper. It provides insights into the dynamics of deep network training and sheds light on the feasibility of training complex networks.

- How might this knowledge impact network design and optimization strategies?
  This knowledge can help us develop more efficient training procedures and improve our understanding of generalization and model selection in deep learning, as argued by the paper. It can also guide the development of more effective training algorithms and regularization techniques.