Information Theoretic Analysis of Neural Networks

Seyed Mehdi Ayyoubzadeh

February 22, 2022

1 Overview and Literature Search

In this project, we aim to review the information-theoretic analysis of deep neural network architectures. Although Deep Neural Networks (DNNs) outperformed in many real-world tasks, there are still a few works that analyze these Deep architectures from an information theory perspective. Understanding the statistical properties of DNNs can provide more intuition about these architectures that can lead to improving them. For supervised learning problems, the goal of any Deep architectures is to find an efficient representation of the inputs that can provide the necessary information about the outputs. This is closely related to the finding of minimal sufficient statistics. The Deep architectures try to find a mapping that can optimally compress the input while preserving the information about output. This is the goal of a problem in information theory called the Information Bottleneck problem. This problem arises in many areas. We can inspect this problem in DNNs by analyzing and looking at the Information Plane in such networks. In [1], Goldfeld and Polyanskiy stated the information bottleneck in machine learning problems. They have given different examples in different areas of communication and machine learning that the Information Bottleneck problem arises. They have also formulated this problem for DNNs. In [2], Tishby and Zaslavsky specifically examined this problem for DNNs and tried to analyze the behavior of the DNNs by inspecting the mutual information and Information plane between layers. In [3], Schwartz-Ziv and Tishby follow up the previous work and extend the study to discover more features about DNNs. They studied the effect of Stochastic Gradient Descent (SGD) in the Information Plane for DNNs. They have also explained why and how adding more layers can benefit DNNs. Besides, they studied how much the hidden layers can form optimal Information Bottleneck representation for the data.

2 Proposal Details

First, we will talk about the Information Bottleneck problem in information theory and its applications in different areas of machine learning and communications. Then, we develop the optimization formulation and see the special cases of this problem [1]. We will see how the Information Bottleneck theory helps to understand DNNs. We try to study the information plane of DNNs review the dynamics of them during the training of DNNs [3]. The experimental results should support training DNNs has two main steps, fitting the data and then compressing the representation of the input data while preserving the useful information for predicting the output. And the second step takes the major part of the training of DCNNs.

References

[1] Z. Goldfeld and Y. Polyanskiy, "The information bottleneck problem and its applications in machine learning," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, pp. 19–38, May 2020.

[2] N. Tishby and N. Zaslavsky, "Deep learning and the information bottleneck principle," 2015.

[3] R. Shwartz-Ziv and N. Tishby, "Opening the black box of deep neural networks via information," 2017.