Directly maximizing mutual information to tackle the memorization problem

CS330 Final Project Proposal

Bryce Long
molohov@stanford.edu

I-Sheng Yang
ishengy@stanford.edu

1 Objective

For our final project, we propose exploring different ways to solve the meta-learning memorization problem by directly maximizing mutual information between task training data and output predictions, through newly proposed techniques and/or our own algorithmic experimentation, time permitting. Tackling the memorization problem is vital for meta-learning scenarios where tasks cannot be made mutually exclusive.

2 Related Works

The problem is introduced in [1] and proposes minimizing a new meta regularization objective in order to restrict information flow between the task test data and shared parameters to the output prediction. The hope is that this will encourage the model to use task training data to make predictions, and not memorize inputs from task test data, so the network can generalize. Despite the impressive results, this method is indirect because the restriction is only encouraged - not enforced - and there are some cases presented in the paper where the method fails to prevent the memorization problem.

3 Technical Outline

Our goal is to find an network-level approach that directly learns to use task training data to make predictions. Professor Finn has suggested a technique in [2] that uses another regularization technique on cGANs to encourages their generated outputs to be diverse by penalizing generated outputs that do not depend on the conditioning variable. Analogies between generated outputs and meta-test predictions can be drawn, as well as between the conditioning variable and the meta task training data. Thus, employing this cGANs along with the regularization technique as a meta-learning network could cause its predictions to be aware of the differences of its task training data, and could perform well on tasks that are not mutually exclusive.

In addition to the suggested approach, we could find and brainstorm other techniques try to implement them also, time permitting.

To evaluate the performance of our implementation(s), we will first reproduce a baseline of one of the examples in [1], such as pose prediction or unshuffled N-way image classification on Omniglot. We will then apply our method(s) and compare test accuracy against the baseline meta regularization technique. As the meta regularization technique formally defines a metric of mutual information between two distributions, we would like to see if we can derive something similar for our directed approaches, or adapt the existing metric on our network.

References

[1] Mingzhang Yin, George Tucker, Mingyuan Zhou, Sergey Levine, and Chelsea Finn. Meta-learning without memorization, 2020.

[2] Dingdong Yang, Seunghoon Hong, Yunseok Jang, Tianchen Zhao, and Honglak Lee. Diversity-sensitive conditional generative adversarial networks, 2019.