Learning Inductive Biases with Neural Networks

Reuben Feinman (reuben.feinman@nyu.edu)
Center for Neural Science
New York University

Brenden M. Lake (brenden@nyu.edu)
Department of Psychology and Center for Data Science
New York University

Abstract

People use rich prior knowledge about the world in order to efficiently learn new concepts. These priors–commonly referred to as ”inductive biases”–pertain to the space of internal models considered by a learner, and they help maximize the amount of information that is extracted from limited data. Recently, it was shown that performance-optimized deep neural networks (DNNs) develop inductive biases similar to those possessed by human children. However, these models use unrealistic training data, and it remains unclear whether they develop their biases in the same way as humans. We investigate the development of inductive biases in DNNs and perform novel regional parametric analyses of these biases. Our findings suggest...

Keywords: learning-to-learn; neural networks; inductive biases

Introduction

TODO: intro.

Experiments

This is where experiment information will go.

Acknowledgements

This research was supported by a Henry M. McCracken fellowship and NYU start-up faculty funding (?). We thank Subhankar Ghosh for useful code and preliminary experiment ideas.

References

Ritter, S., Barrett, D. G. T., Santoro, A., & Botvinick, M. M. (2017). Cognitive psychology for deep neural networks: a shape bias case study. In *Proceedings of the 34th international conference on machine learning* (pp. 2940–2949).

The source code repository for this paper can be found at `http://github.com/rfeinman/learning-to-learn`