EPISTEMOLOGY AND THEORY OF MACHINE LEARNING

TERM PAPER: SOME POSSIBLE TOPICS

NOTE. These are just suggestions. Feel free to choose different questions related to the mentioned literature, or to choose different topics altogether – as long as the topic is related to material we discussed in the seminar.

- Harman & Kulkarni (2007, sect. 3.6) write that the use of simplicity in structural risk minimization (SRM) is not to assume that the world is simple. Yet Bargagli Stoffi et al. (2022) argue that SRM works better in simpler worlds and worse in complex worlds, suggesting that it does operate under such an assumption. How to resolve this apparent tension?
- Buchholz (in process) argues that the "standard analysis" of the curve-fitting problem can still be made to fit benign overfitting, by broadening the concept of simplicity. Is the argument successful? What exactly is the trade-off now at play?
- Räz (2022) brings Salmon's statistical relevance model to bear on explaining deep learning. What exactly would be the relevant explanation? Does it help us to better understand benign overfitting, or to interpret trained machine learning models?
- Shalev-Shwartz & Ben-David (2014, ch. 7) suggest that the MDL paradigm gives "a formal justification" to Occam's razor. Do they give an argument, and if so, is it successful? How does this sit with the more skeptical views on simplicity of Harman & Kulkarni (2007) and/or Strevens (2009)?
- Lin (in process) sets out a "unified inductive logic," encompassing statistical learning theory, that is based on notions of convergence to the truth. Do Strevens' (2009) critical observations about "philosophical learning theory" apply to Lin's account, or even spell trouble for it?
- Forster & Sober (1994) draw a number of philosophical conclusions based on the AIC method in model selection. Assess one or more of these claims, but now on the basis of the method of SRM. Do similar claims follow, and why (not)?
- Dotan (2021) argues that the no-free-lunch theorems show the necessity of non-epistemic values in theory choice. Is the argument successful? Must inductive bias be value-laden? Is Rushing's (2022) critique of Dotan's argument successful?
- Schurz (2008) proposes a Reichenbachian justification of induction, based on the machine learning theory of competitive online prediction (also known as "prediction with expert advice"). Is the argument successful? What role does inductive bias play here?