Inductive assumptions in ML (Abstract for KHK publication)

Inductive assumptions in machine learning - always required, often ignored

Every machine learning system encodes inductive assumptions. Inductive assumptions are unsafe in that they invariably introduce inductive errors. As long as these errors can be controlled this might not be bad, but as I argue, in practice things often look different. Before delving into practical details, I review some popular theoretical approaches to ML (PAC-learning, algorithmic information theory) which promise tight error control and in turn make inductive assumptions very explicit. Already at this level a distinction of such assumptions can be made - ones that allow for quantitative error estimation and ones that don't. Assumptions that don't allow for quantitative error estimations are tightly related to the problem of induction, the others on explicit human decisions.
For feasibility reasons most of the theoretical bounds on error are inapplicable in practice. Approximations to inductive assumptions have to be made which explicitly break them. I specifically discuss the general principle of cross-entropy minimization as a way to approximate the deviation from an unknown distribution. In ML this is better known as training and test set cross validation - a case in point for an implicit inductive assumption in practice. Approximative practice motivates a typology of inductive errors along the dimensions of control and explicitness. I conclude by discussing what these dimensions mean for the epistemology of error.