MACHINE LEARNING FROM EXAMPLES: A NON-INDUCTIVIST ANALYSIS

Edoardo Datteri,° Hykel Hosni,* Guglielmo Tamburrini°
°Department. of Philosophy, Università di Pisa, Italy
*School of Mathematics, University of Manchester, UK
e-mail: edoardo@mail-arts.sssup.it; hykel@maths.man.ac.uk; gugt@inwind.it

It has been suggested that AI investigations of mechanical learning undermine sweeping anti-inductivist views in the theory of knowledge and the philosophy of science. In particular, it is claimed that some mechanical learning systems perform epistemically justified inductive generalization and prediction. Contrary to this view, it is argued that no trace of such epistemic justification is to be found within a rather representative class of mechanical learning agents. An alternative deductive analysis of mechanical learning from examples is outlined.

First drawing on behaviour-based robotics, we examine relatively simple autonomous robots that learn from experience, insofar as they acquire new sensorimotor capabilities and generalize from observation. Analysis of a representative behaviour-based architecture reveals that even these rudimentary learning mechanisms embody crucial assumptions about their environment. In this context, the epistemic problem of induction is not "solved". Rather it is reformulated as the problem of assessing whether projections based on such background assumptions are reasonable to believe.

This epistemic problem is more informatively addressed by reference to the symbolically richer, ID3-style learning algorithms. Pervasive overfitting of training data jeopardizes the idea that epistemically justified induction is at work there: overfitting reminds one that good approximation to the target concept or rule on training data is not, in itself, diagnostic of good approximation over the whole instance space of that concept or rule. Once again, both ID3 learning and successful post-pruning of overfitting trees rely on the conjectural representativeness of concept instance collections.

Having found no trace of epistemically justified induction, an alternative deductivist account is outlined, drawing on some families of non-monotonic consequence relations, and emphasizing deductive trial and error-elimination processes in autonomous learning mechanisms. These processes interleave default-based introduction of projective hypotheses from observed samples, retraction of falsified hypotheses, and heuristic selection of new background assumptions for more effective learning.