Sources of Uncertainty in Supervised Machine Learning - A Statisticians' View

Cornelia Gruber | LMU Munich | Germany

Hide details

Authors:

Cornelia Gruber | LMU Munich | Germany Patrick Oliver Schenk | LMU Munich | Germany Dr. Malte Schierholz | LMU Munich Prof. Dr. Frauke Kreuter | LMU Munich Prof. Dr. Göran Kauermann | LMU Munich

Supervised machine learning and predictive models have achieved an impressive standard today, enabling us to answer questions that were inconceivable a few years ago. Besides these successes, it becomes clear, that beyond pure prediction, which is the primary strength of most supervised machine learning algorithms, the quantification of uncertainty is relevant and necessary as well.

However, before quantification is possible, types and sources of uncertainty need to be defined precisely.

While first concepts and ideas in this direction have emerged in recent years, this paper adopts a conceptual, basic science perspective and examines possible sources of uncertainty. By adopting the viewpoint of a statistician, we discuss the concepts of aleatoric and epistemic uncertainty, which are more commonly associated with machine learning. The paper aims to formalize the two types of uncertainty and demonstrates that sources of uncertainty are miscellaneous and can not always be decomposed into aleatoric and epistemic. Drawing parallels between statistical concepts and uncertainty in machine learning, we emphasise the role of data and their influence on uncertainty.