

Is it irrational for a Bayesian prediction modeler to set data aside for model evaluation?

A seminar by Giovanni Parmigiani

Harvard University

Thursday 23 Oct 2025 | 14:30-15:30

Room BENVENUTI

Department of Statistical Sciences

Hold-out data sets are data sets used in model evaluation and not in model training, even though they are available at the time of training. Using hold-out data is common practice when developing prediction models, it is often a recommended step before the implementation of machine learning and artificial intelligence in practice, and can be a regulatory requirement in some cases. On the other hand, from a Bayesian standpoint, there is no rational justification that I know of to not look at all the data available when training a model. Can it be rational to hold out data from training to do an out-of-sample evaluation? In this paper, I formally define a game that I hope may serve as the normative foundation for a Bayesian model developer's decisions about model validation. In a simple setting, I present a conjecture that may provide an early view into the conditions for rationality of hold-out sets.



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

