Output-weighted optimal sampling for Bayesian regression and rare event statistics using few samples

Themis Sapsis, MIT

For many important problems the quantity of interest is an unknown function of the parameters, which is a random vector with known statistics. Since the dependence of the output on this random vector is unknown, the challenge is to identify its statistics, using the minimum number of function evaluations. This problem can be seen in the context of active learning or optimal experimental design. We employ Bayesian regression to represent the derived model uncertainty due to finite and small number of input–output pairs. In this context we evaluate existing methods for optimal sample selection, such as model error minimization and mutual information maximization. We show that for the case of known output variance, the commonly employed criteria in the literature do not take into account the output values of the existing input–output pairs, while for the case of unknown output variance this dependence can be very weak. We introduce a criterion that takes into account the values of the output for the existing samples and adaptively selects inputs from regions of the parameter space which have an important contribution to the output. The new method allows for application to high-dimensional inputs, paving the way for optimal experimental design in high dimensions