

Evidence of successively generated models

Let us investigate an algorithm of regression model construction. The constructed model will be used to solve problems of the Financial Sector: it might be a scoring model, an energy consumption forecast model or European option volatility smile model. We suppose that given historical data are not sufficient to discover hidden dependencies in an investigated problem. So we propose the following approach to the model construction. Together with historical data we use expert-given set of primitive functions. It is recommended to collect functions, which already widely used to model the investigated problem. Then we assign a generating function, which will be used to generate the set of the competitive models. We estimate evidence of the models using coherent Bayesian inference and select a model of the best structure. Since generating functions make a countable set of models, we organize an iterative generation-selection procedure.

Each cycle of the procedure include the following steps. First, we modify competitive models so that the structural distance between an original and a derivative model will as minimal as possible. Second, we estimate parameters and hyperparameters of the derivative model to cut-off some model modifications at the following steps and reduce the algorithm complexity. Third, we analyze the evidence of the derivative model to find the probability to become it a model of the optimal structure. Also, we analyze some restrictions applied to the model structure and robustness of the model.

As the result we obtain a model, interpretable from the expert's point-of view; if fits historical data well and robust. Some additional tests are applied to verify the result model: cross-validation and retrospective forecasting to ensure quality of the further use.

Keywords: model generation, model selection, coherent Bayesian inference