



Generalised MAP inference

Concept. To largely generalise the procedure of statistical inference for any model using an algorithm which builds from techniques we developed in the previous chapter. When we say ‘statistical inference’ here; we specifically mean computing the maximum a posteriori (MAP) estimate for any arbitrary stochastic model which has been defined in the stochadex simulator. In order for our algorithm to evaluate the MAP of a model, we show that the user must specify the prior distribution over model parameters, and the model must itself be defined within the stochadex. In order to describe how our algorithm works, in this chapter, we will discuss some concepts which are commonplace within the field of Bayesian inference and provide a few simple examples of how our algorithm might work in various instances. For the mathematically-inclined, this chapter will give a very brief exposition for Bayesian statistical inference methodology — in particular, how it relates to the evaluation of MAP estimates. For the programmers, the software described in this chapter lives in the public Git repository: <https://github.com/umbralcalc/learnadex>.

5.1 Methodology

First there’s the problem of knowing what the likelihood of the data is. The measurements y, y', y'', \dots have either a known distribution which the user specifies or the filtering algorithm runs on this to get the data likelihood.

Once you have a data likelihood, then inference should proceed on the sampling domain side of things where the filtering algorithm is used again! This time with a tunable conditional probability that is a gaussian with mean and variance estimated directly from the history (don’t optimise it like in Bayesian optimisation!) with a tunable exponential timestep kernel (timestep being the optimiser step here) and the resulting function can be optimised (or draw monte-carlo samples from) in an EM algorithm approach. The resulting Gaussian function could also exploit gradients for SGD.

Reference the contrast to ABC methods here, which involve approximating the data likelihood with a simple proximity function with a tolerance ϵ . Also talk about the BOLFI method which does indeed use the full Bayesian optimisation as it goes.

Bibliography