Reflection - Sparse Gaussian Process - Snelson, Ghahramani

Nicholas Hoernle **April 25, 2018**

The authors attempt to solve the $O(N^3)$ computational overhead of the GP inference (dominated by the matrix inversion operation of the Kernel matrix). They rather suggest the selection of $M << N$ pseudo-datapoints to use as training data, thus requiring a smaller and therefore more computationally tractable Kernel matrix. Alternative approaches select a subset of real datapoints but experience problems with hyperparameter optimization due to local optima that result from the subset of data that are selected. These authors propose to rather introduce pseudo-datapoints, the positions of which can be optimized jointly to the hyperparameter optimization.

In the last paragraph, the authors give a heuristic argument for the largest computational overhead of the SPGP and alternative approaches. They argue that while the SPGP is more computationally expensive than the alternatives, the gradient ascent algorithm dominates the computation cost and therefore the SPGP specific solution is favorable due to the better accuracy that is achieved. The plots in Figure 3 suggest that the SPGP also converges more quickly (**by iteration**) than the alternatives and therefore the algorithm might end up being more computationally efficient in practice. There lacks a conclusive investigation into the wall time computation of the alternatives that are presented. I am also concerned by the mention of the high computation cost of the gradient ascent which, in extreme cases, could defeat the entire aim of reducing the N^3 computation of the standard GP.