Gaussian Processes for Active Sensor Management

Alexander N. Dolia, Chris Harris, John Shawe-Taylor {AD,CJH,JST}@ECS.SOTON.AC.UK
School of Electronics and Computer Science, University of Southampton, Southampton, UK

Tijl De Bie TIJL.DEBIE@GMAIL.COM
OKP Research group, Katholieke Universiteit Leuven, Belgium

Editor: Neil D. Lawrence, Joaquin Quin onero Candela, Anton Schwaighofer

Abstract

In this paper we study the active sensor management problem using continuous optimal experimental design (OED) framework. This task comprises the determination of allocation for a limited number of sensors over the spatial domain and the number of repetitive measurements in these locations in order to improve the overall system performance. We present a principled approach to active sensor management with repetitive measurements for Gaussian Processes (GPs) using a generalised D-optimality criteria and soft margin constrains. The resulting optimum of the convex optimization of the optimal experimental design for GP is generally sparse, in the sense that measurements should be taken at only a limited set of possible sensor locations. We demonstrate the use of our method on artificial dataset.

Keywords: Optimal experimental design, Sensor Management, Gaussian Processes.

1. Introduction

It is intriguing that two such important problems in their own right such as the minimum volume covering ellipsoid (MVCE) problem and D-optimal experimental design (OED), are Lagrangian duals of each other. Many applications in control, system identification, visual/audio tracking, data mining, and of course experimental design, robust statistics and novelty/outlier detection can be solved by (or reformulated as) the MVCE/OEP optimization problem (Sun and Freund, 2004; Titterington, 1975). Active sensor management is related to machine learning problems such as active learning (MacKay, 2003).

But many OED methods (Titterington, 1975; Guestrin et al., 2005) do not take into account the following problems. Firstly, it is necessary to determine the solution of optimal experimental design problem that is optimal for a given set of different models and prior information about the probability of model occurrences. Secondly, the upper bound for the number of repetitive measurements and cost of taking measurements in particular location can vary over spatial domain. Finally, it is required that OED method uses the convex optimisation problem.

The importance of OED and sensor management cannot be overestimated. Measurements are often very expensive or dangerous, especially in military or industrial applications, and a priori maximization of the information revealed by them can save large amounts of money and/or reduce the risk being taken. However, thus far there has been no OED formulation which handle aforementioned problems and is optimal if *Gaussian Processes* is used in the subsequent regression step. An important contribution of this paper is to fill this gap. In this paper we propose to use the generalised D-optimal experimental design with soft margin constrains.

2. Optimal experimental design

Let the data matrix be $\mathbf{X} = (\mathbf{x}_1' \ \mathbf{x}_2' \ \cdots \ \mathbf{x}_\ell)'$, let $\mathbf{y} = (y_1 \ y_2 \ \cdots \ y_\ell)$ denotes the observations, and let $\mathbf{n} = (n_1 \ n_2 \ \cdots \ n_\ell)$ denote the experimental noise. We assume the following model for the data:

$$\begin{aligned} \mathbf{y}_s = f_s(\mathbf{X}) + \mathbf{n}, \text{where} \quad E\{\mathbf{n}\} = 0, \quad E\{\mathbf{n}\mathbf{n}'\} = \sigma^2\mathbf{I}, \\ f_s \sim GP(m_s, \mathbf{K}_s), \quad f_s \sim GP(m_s, \mathbf{K}_s + \sigma^2\mathbf{I}), \quad s = 1, ..., S, \end{aligned} \tag{1}$$

where the function f_s is distributed as a GP with mean function m_s and covariance (kernel) function $\mathbf{K}_s$, S is the number of considered models. We aim to estimate the regression function f_s. For convenience and without loss of generality, we will assume σ^2 to be equal to 1 in this paper.

OED is concerned with the optimal selection of the data points $\mathbf{x}_i$ on the grid (out of a set of such data points) at which the model should be sampled, in order to optimize some measure of performance of the estimator of the weight vector $\beta_{i,s}$ of the s-th model $f_s(x) = \sum_i \beta_{i,s} k_s(\mathbf{x}_i, \mathbf{x})$.

3. Proposed optimisation problem

For the given set of kernels $\mathbf{K}_s$, the matrix $\mathbf{\Lambda}$ and the weights w_i and the parameter γ, we propose the following formulation:

$$\boldsymbol{\alpha}_\gamma^* = \quad \operatorname{argmin}_{\boldsymbol{\alpha}} \quad -\sum_{s=1}^{S} w_s \log\det\left(\mathbf{A}\mathbf{\Lambda}\mathbf{K}_s\mathbf{\Lambda}\mathbf{A} + \gamma\mathbf{I}\right), \tag{2}$$

$$\begin{aligned} \text{s.t.} \quad & \mathbf{a}'\mathbf{a} \le 1, \\ & 0 \le \mathbf{a} \le \sqrt{\frac{1}{\nu\ell}}\mathbf{e}, \\ & \mathbf{w}'\mathbf{e} = 1, \quad \mathbf{w} \ge 0. \end{aligned} \tag{3}$$

where the diagonal matrix $\mathbf{A}$, with $\mathbf{A}_{ii} = a_i \triangleq \sqrt{\alpha_i} \ge 0$, such that (with $\mathbf{a} = (a_1 \ a_2 \ \cdots \ a_n)'$) from $\mathbf{e}'\boldsymbol{\alpha} = 1$ we have that $\mathbf{a}'\mathbf{a} = 1$, $\boldsymbol{\alpha}$ define the solution of the continuous OED problem (Titterington, 1975), $\mathbf{K}_s = \{k_s(\mathbf{x}_i, \mathbf{x}_j)\}|_{(i=1,\ldots,\ell,\ j=1,\ldots,\ell)}$, the given matrix $\mathbf{\Lambda} = diag\{\lambda_1, ..., \lambda_\ell\}$ defines cost of taking measurements in different locations, ν is the upper bound for the number of repetitive measurements (ν can be different for the different sensor locations). If the cost of taking measurements in different locations is the same we can omit the matrix $\mathbf{\Lambda}$.

4. Experiments

We have to point out one technicality of our method we have not discussed so far. Because of the relaxation step we used in our derivation, the design is not discrete, meaning that the weights α_i are not exactly equal to integer values N_i (the number of repetitive measurements in ith sensor location) divided by N (the total number of measurements or experiments). In order to derive a discrete design $\{N_i\}$ from a continuous design vector $\boldsymbol{\alpha}$, we propose to regard $\boldsymbol{\alpha}$ as a probability distribution over the different data points $\mathbf{x}_i$ (note that $\boldsymbol{\alpha}'\mathbf{e} = 1$), and randomly sample N times from this distribution. This yields an empirical (and discrete) approximation $\frac{\hat{N}_i}{N}$ for α_i, if $\hat{N}_i$ samples are at the ith position. In our experiments, we computed 100 such random discrete approximations, and kept the discrete design achieving the lowest cost according to (2) with $\alpha_i = \frac{\hat{N}_i}{N}$. In our experiment $S = 1$, $w_1 = 1$ and $\boldsymbol{\Lambda} = diag\{1, .., 1\}$.

To make an objective evaluation possible, we considered the sinc function (on the interval $[-4, 4]$, discretized with data points 0.1 apart from each other), and performed kernel ridge regression based on the data $\mathbf{x}_{D,i}$ and $y_{D,i} = \text{sinc}(\mathbf{x}_{D,i}) + n_i$ with $n_{D,i}$ Gaussianly distributed random noise with standard deviation 0.2. A Gaussian kernel with kernel width $\rho = 0.5$ is used. In this way objective performance measures according to how well the regression function approximates the target sinc function can be computed. We carried out experiments comparing the design $\mathbf{X}_D$ found by our method, with a design where the design points are uniformly spaced, and with a design where the design points are uniformly randomly sampled from the domain. The experiments are summarized in Figure 1. We can conclude that our approach results in a considerable improvement as compared to the random and the uniform designs, especially for larger sample sizes N (which may be due to the fact that for larger sample sizes the discrete design is a better approximation of the continuous one).

5. Conclusions

We have investigated the sensor management problem based on D-optimal experimental design. We derived the regularized version based on the subsequent use of kernel ridge regression or GPs as an estimation technique for the regression weight vector. We believe that the presented nonparametric approach is likely to prove a very useful alternative to parametric approaches.

References

C. Guestrin, A. Krause, and A. Singh. Near-optimal sensor placements in gaussian processes. In *Proceedings of International Conference on Machine Learning*, 2005.

D. MacKay. *Information theory, inference, and learning algorithms*. Cambridge, 2003.

P. Sun and R.M. Freund. Computation of minimum-volume covering ellipsoids. *Operations Research*, 52(5):690–706, 2004.

D.M. Titterington. Optimal design: Some geometrical aspect of d-optimality. *Biometrika*, 62(3):313–320, 1975.

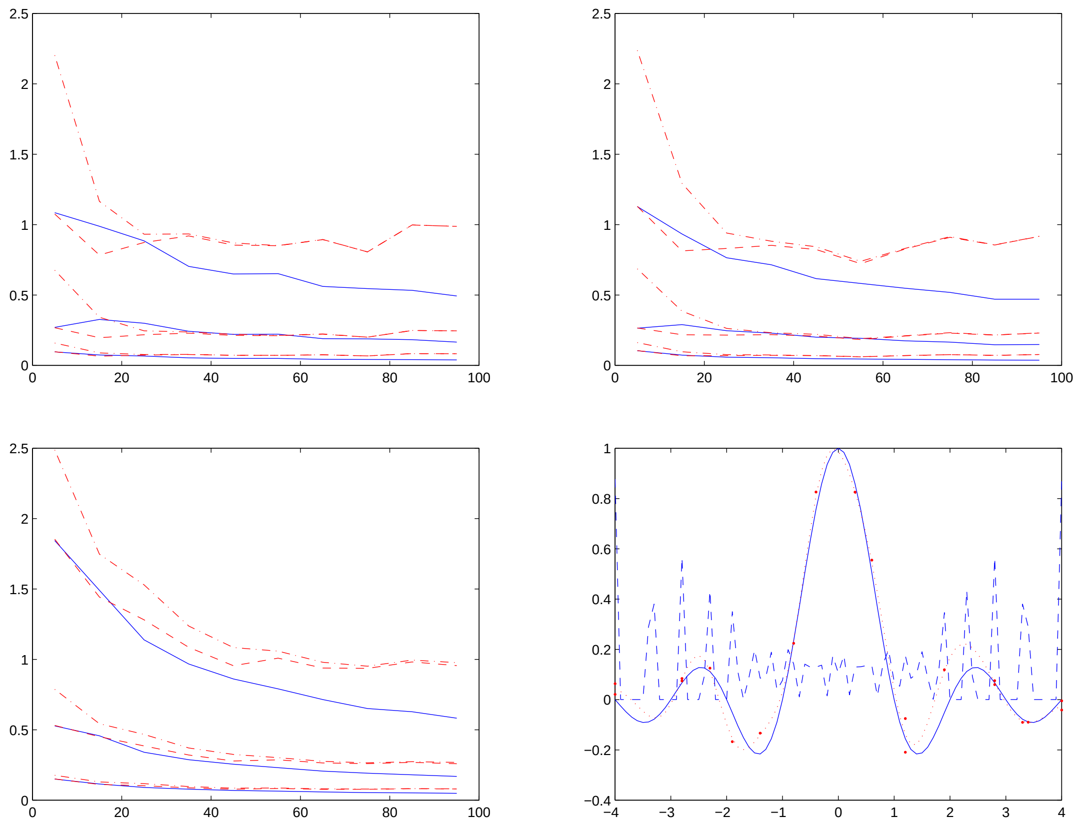

Figure 1: Top left ($\gamma = 0.01$), top right ($\gamma = 0.1$) and bottom left ($\gamma = 1$) pictures: the average 2-norm (three highest curves), ∞-norm (three middle curves) and 1-norm errors (three lowest curves) over the interval $[-4, 4]$, as a function of the number of data points in the design, and this for random (dash-dotted lines), uniform (dashed lines), and optimal designs as computed by our method (full lines). For each of the sample sizes, the average error over 100 random noise sequences $\mathbf{n}_D$ is shown, and with 100 different random designs for the random design error curves (the uniform and optimal designs are fixed). Clearly, all error measures decrease as the sample size increases. Mainly for large sample sizes, the optimal design outperforms the random and uniform designs. The bottom right figure shows the sinc function on $[-4, 4]$, the relaxed design vector α_i as a function of $\mathbf{x}_i$ (dashed line), the values $y_{D,i} = \text{sinc}(\mathbf{x}_{D,i}) + n_{D,i}$ according to a discrete design $\{N_i\}$ derived from this relaxed design (dots), and the resulting regression function (dotted line). Here we used $\gamma = 0.1$.