Gaussian Processes for Active Sensor Management

Alexander N. Dolia, University of Southampton

This poster is based on

- A.N.Dolia, C.J.Harris, J.Shawe-Taylor, D.M.Titterington, **Kernel Ellipsoidal Trimming**, submitted to the Special Issue of the Journal Computational Statistics and Data Analysis on Machine Learning and Robust Data Mining. under review.

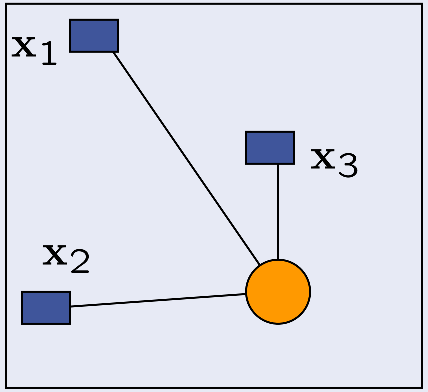
- A.N.Dolia, T.De Bie, C.J.Harris, J.Shawe-Taylor, D.M.Titterington. **Optimal experimental design for kernel ridge regression, and the minimum volume covering ellipsoid**, Workshop on Optimal Experimental Design, Southampton, 22-26 September, 2006

Joint work with: Dr. **Tijl De Bie**, Katholieke Universiteit Leuven
Prof. **John Shawe-Taylor**, University of Southampton
Prof. **Chris Harris**, University of Southampton
Prof. **Mike Titterington**, University of Glasgow

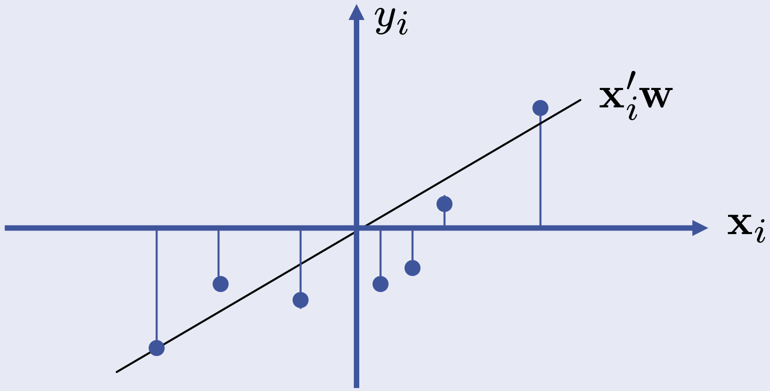
Problem Statement

Aim is to **estimate locations of the sensors and number of repetitions** given a set of possible sensors locations, cost of measurements and upper bound for the number of repetitions at given sensor locations in order to get **good prediction** $f(\mathbf{x})$

- Sensor network: N sensors measure signals at positions $\mathbf{x}_i$
- Sensors measure function $y_i = f(\mathbf{x}_i) = \mathbf{x}_i'\mathbf{w} + n_i$
- Weight vector $\mathbf{w}$ gives information about 'system'
- Position sensors optimally at $\mathbf{X}_D$
- Estimate $\mathbf{w}$ based on $\mathbf{X}_D$



Optimal experiment design?



Optimal experiment design (OED) idea:

- Given a set of n data points $\mathbf{X} = \{\mathbf{x}_i\}$
- Choose multiset $\mathbf{X}_D = \{\mathbf{x}_{D,i}\} \subseteq \mathbf{X}$ with N data points, N_i times $\mathbf{x}_i$
- Measure at $\mathbf{x}_{D,i} \to \mathbf{y}_D = \{y_{D,i}\}$ with $y_{D,i} = \mathbf{x}_{D,i}'\mathbf{w} + n_i$
- Estimate $\mathbf{w}$ based on $\{\mathbf{X}_D, \mathbf{y}_D\} \to \hat{\mathbf{w}}$

Optimal experiment design for RR

- Result is thus a non-convex optimization problem:

$$\begin{aligned} \min_\alpha \quad & -\log\det\left(\sum_i \alpha_i \mathbf{x}_i\mathbf{x}_i' + \gamma\mathbf{I} + \frac{1}{4}\gamma^2\left(\sum_i \alpha_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\right) \\ \text{s.t.} \quad & \alpha'\mathbf{e} = 1 \\ & \alpha \geq 0 \end{aligned}$$

- Minimize tight upper bound:

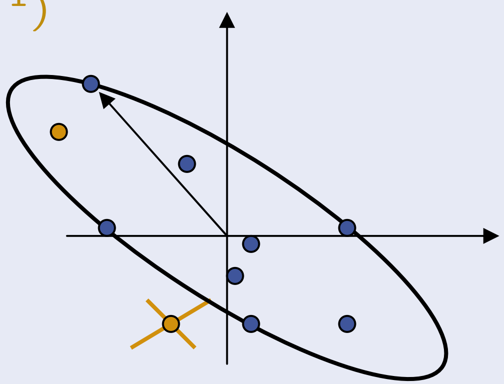
$$\begin{aligned} \alpha_\gamma^* = \quad \text{argmin}_\alpha \quad & -\log\det\left(\sum_i \alpha_i \mathbf{x}_i\mathbf{x}_i' + \gamma\mathbf{I}\right) \\ \text{s.t.} \quad & \alpha'\mathbf{e} = 1 \\ & \alpha \geq 0 \end{aligned}$$

- This is a convex optimization problem again

Regularized MVCE

- What about the dual of the regularized D-OED?

$$\begin{array}{ll} \min_{\mathbf{M},\mu} & \log\det(\mathbf{M}) + \mu + \gamma \text{trace}(\mathbf{M}^{-1}) \\ \text{s.t.} & \mathbf{x}_i'\mathbf{M}^{-1}\mathbf{x}_i <= \mu \end{array}$$



- The optimum is given by:

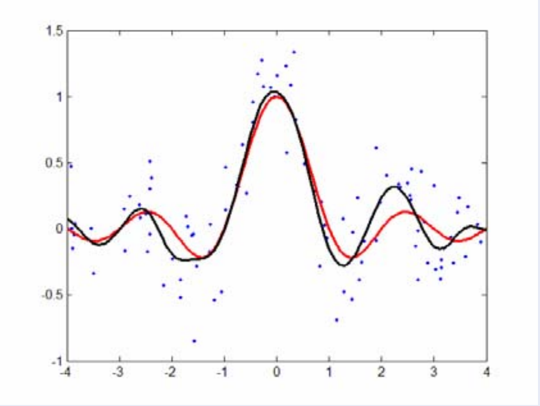
$$\mathbf{M}_\gamma^* = \sum_i \alpha_{\gamma,i}^* \mathbf{x}_i\mathbf{x}_i' + \gamma\mathbf{I}$$

where α_γ^* is the solution of the regularized D-OED problem

- Interpretation: $\text{trace}(\mathbf{M}^{-1}) = \sum_i \frac{1}{\lambda_i}$ → fit an ellipsoid, but make sure none of the eigenvalues of $\mathbf{M}_\gamma^*$ is too small...

Kernel ridge regression (KRR)

- Kernel ridge regression (KRR):



$$\mathbf{K}_D = \mathbf{X}_D \mathbf{X}'_D$$

$$\beta = (\mathbf{K}_D + \tilde{\gamma}\mathbf{I})^{-1}\mathbf{y}$$

$$\hat{\mathbf{w}}_{RR} = \mathbf{X}'_D \beta = \sum_i \beta_i \mathbf{x}_{D,i}$$

$$f(\mathbf{x}) = \mathbf{x}'\hat{\mathbf{w}}_{RR} = \sum_i \beta_i \mathbf{x}'\mathbf{x}_{D,i} = \sum_i \beta_i k(\mathbf{x}, \mathbf{x}_{D,i})$$

- Everything expressed in terms of $\mathbf{K}_D$ (i.e. in terms of inner products/kernels): 'kernel trick'

- If we want to do OED for KRR, we need to write it entirely in terms of kernel evaluations/inner products—can we?

Kernel MVCE

- Mahalanobis distances $\mathbf{x}'(\sum_i \alpha^*_{\gamma,i}\mathbf{x}_i\mathbf{x}_i' + \gamma\mathbf{I})^{-1}\mathbf{x}$ in terms of inner products/kernel evaluations?

- Let $\mathbf{AKA} = \mathbf{V\Lambda V}'$ (eigenvalue decomposition), then (derivation not shown...):

$$\mathbf{x}'(\sum_i \alpha^*_{\gamma,i}\mathbf{x}_i\mathbf{x}_i' + \gamma\mathbf{I})^{-1}\mathbf{x} = \frac{1}{\gamma}\left(\mathbf{x}'\mathbf{x} - \mathbf{x}'\mathbf{X}'\mathbf{AV\Lambda}(\mathbf{\Lambda} + \gamma\mathbf{I})^{-1}\mathbf{V}'\mathbf{AXx}\right)$$

- Express in terms of $k(\mathbf{x},\mathbf{x}) = \mathbf{x}'\mathbf{x}$ and $\mathbf{k} = \mathbf{Xx}$, then:

$$\mathbf{x}'(\sum_i \alpha^*_{\gamma,i}\mathbf{x}_i\mathbf{x}_i' + \gamma\mathbf{I})^{-1}\mathbf{x} = \frac{1}{\gamma}\left(k(\mathbf{x},\mathbf{x}) - \mathbf{k}'\mathbf{AV\Lambda}(\mathbf{\Lambda} + \gamma\mathbf{I})^{-1}\mathbf{V}'\mathbf{Ak}\right)$$

completely expressed in terms of kernels

Novelty detection
MVCE and duality
Regularized MVCE
Kernel MVCE

OED: summary

	D-OED	MVCE
standard	$\min_{\alpha} \; -\log\det\left(\sum_i \alpha_i \mathbf{x}_i \mathbf{x}_i'\right)$ s.t. $\alpha'\mathbf{1} = 1$ $\alpha \geq 0$	$\min_{\mathbf{M},\mu} \; \log\det(\mathbf{M}) + \mu$ s.t. $\mathbf{x}_i'\mathbf{M}^{-1}\mathbf{x}_i <= \mu$
regularized	$\min_{\alpha} \; -\log\det\left(\sum_i \alpha_i \mathbf{x}_i \mathbf{x}_i' + \gamma\mathbf{I}\right)$ s.t. $\alpha'\mathbf{1} = 1$ $\alpha \geq 0$	$\min_{\mathbf{M},\mu} \; \log\det(\mathbf{M}) + \mu + \gamma \text{trace}(\mathbf{M}^{-1})$ s.t. $\mathbf{x}_i'\mathbf{M}^{-1}\mathbf{x}_i <= \mu$
kernel	$\min_{\mathbf{a}} \; -\log\det(\mathbf{AKA} + \gamma\mathbf{I})$ s.t. $\mathbf{a}'\mathbf{a} <= 1$ $\mathbf{a} \geq 0$	

Experiment

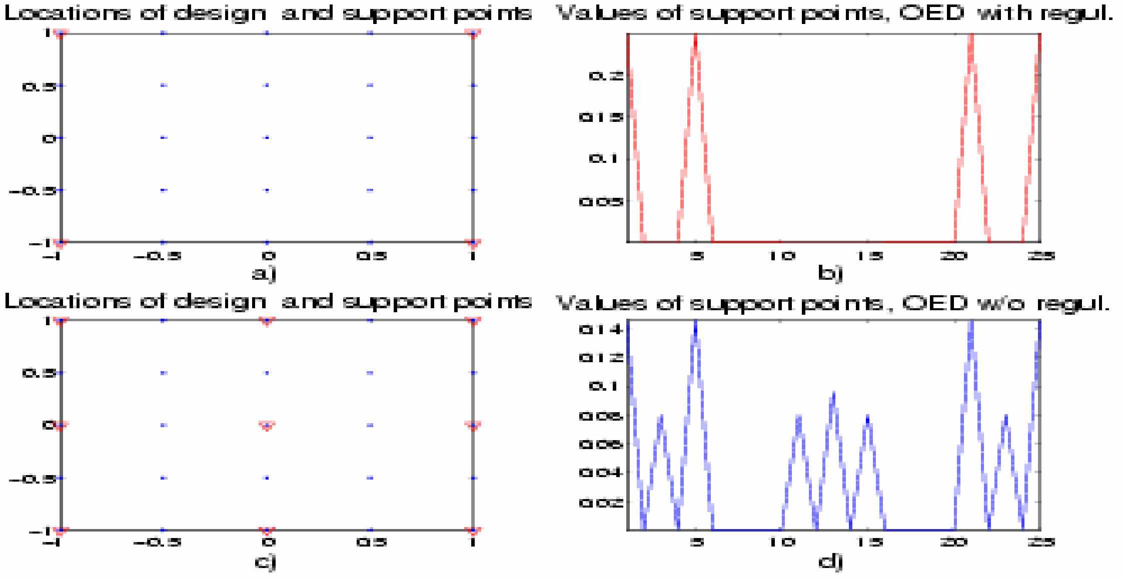


FIGURE 4.1: Support points of regularized and non regularized experimental design, $\bar{\gamma} = 20$ and $\gamma = 2\bar{\gamma}/N$, $N = 100$

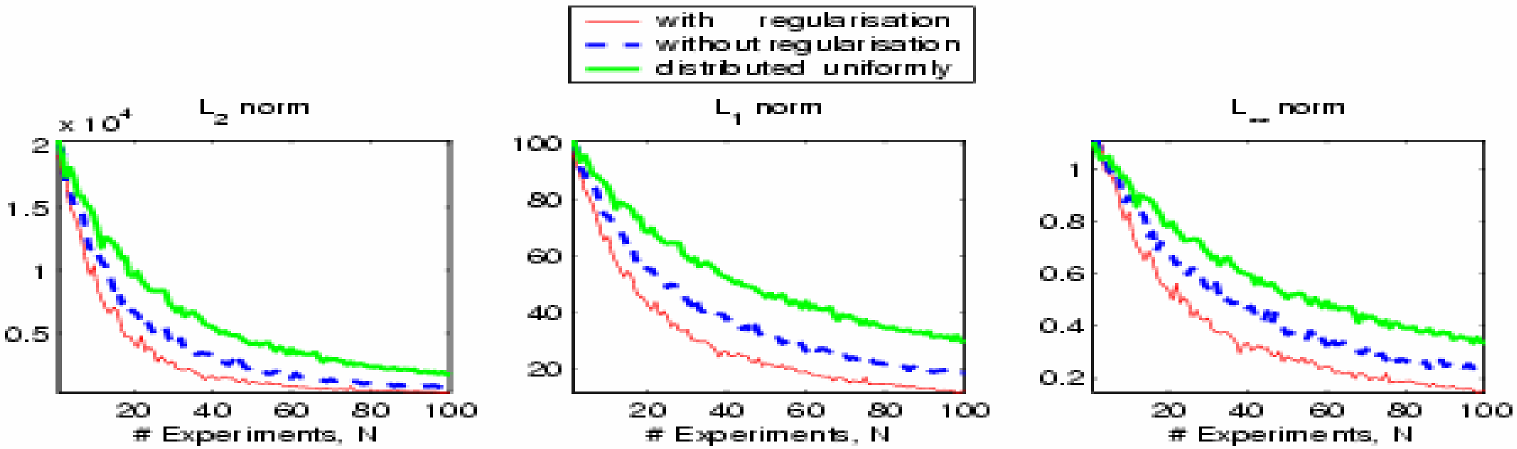


FIGURE 4.2: nonkernel OED performance, $\bar{\gamma} = 20$ and $\gamma = 2\bar{\gamma}/N$

Generalised D-optimal Experimental Design

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:

$$\begin{aligned} \boldsymbol{\alpha}^*_\gamma = \quad & \mathrm{argmin}_{\boldsymbol{\alpha}} \quad -\sum_{s=1}^{S} w_s \log\det\left(\mathbf{A}\mathbf{\Lambda}\mathbf{K}_s\mathbf{\Lambda}\mathbf{A} + \gamma\mathbf{I}\right), \\ & \text{s.t.} \qquad \mathbf{a}'\mathbf{a} \le 1, \\ & \qquad\qquad 0 \le \mathbf{a} \le \sqrt{\frac{1}{\nu\ell}}\mathbf{e}, \\ & \qquad\qquad \mathbf{w}'\mathbf{e} = 1, \quad \mathbf{w} \ge 0. \end{aligned}$$

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), $\mathbf{K}_s = \{k_s(\mathbf{x}_i, \mathbf{x}_j)\}|_{(i=1,\ldots,\ell, \quad j=1,\ldots,\ell)}$, the given matrix $\mathbf{\Lambda} = diag\{\lambda_1, \ldots, \lambda_\ell\}$ 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 $\mathbf{\Lambda}$.

Conclusions

- Two seemingly very different algorithms within one optimization framework
- A way to perform optimal experimental design in high dimensional spaces, such as kernel induced feature spaces
- A way to perform minimum volume covering ellipsoid estimation in high dimensional spaces to perform novelty detection
- Nice features: Convex optimisation and sparse solution