Dual Regularized Total Least Squares in Learning Theory

Sergiy Pereverzyev Jr.[1], Shuai Lu[2], Sivananthan Sampath[2]

[1]Industrial Mathematics Institute, Johannes Kepler University of Linz, Austria
[2]Radon Institute for Computational and Applied Mathematics, Linz, Austria

IP-TA 2010, Warsaw, February 9-12

problem of learning

5th IP:M&S, Antalya, Turkey

A. Hasanov:
"focus on the problem of learning as an inverse problem,
using some **new regularization techniques**,
in order to open *new perspectives* for solving the considered problems
more fundamentally"

problem of learning 2

[T.Poggio, S.Smale, 2003]

$$\text{wish: } y = f(x)$$

problem of learning 2

[T.Poggio, S.Smale, 2003]

$$\text{wish: } y = f(x)$$

$$p(x, y) \text{ — probability distribution on } \mathbb{R}^2$$

problem of learning 2

[T.Poggio, S.Smale, 2003]

$$\text{wish: } y = f(x)$$

$$p(x,y) \text{ — probability distribution on } \mathbb{R}^2$$

$$f(x)\text{: risk } I[f] = \int_{\mathbb{R}^2} (y - f(x))^2 p(x,y)\,dxdy \;\rightarrow\; \min$$

problem of learning 2

[T.Poggio, S.Smale, 2003]

$$\text{wish: } y = f(x)$$

$$p(x, y) \text{ — probability distribution on } \mathbb{R}^2$$

$$f(x)\text{: risk } I[f] = \int_{\mathbb{R}^2} (y - f(x))^2 p(x, y)\,dxdy \rightarrow \min$$

$$\text{minimizer (target function): } f_0(x) = \int_{\mathbb{R}} y p(y|x)\,dy$$

problem of learning 2

[T.Poggio, S.Smale, 2003]

$$\text{wish: } y = f(x)$$

$$p(x,y) \text{ — probability distribution on } \mathbb{R}^2$$

$$f(x)\text{: risk } I[f] = \int_{\mathbb{R}^2} (y - f(x))^2 p(x,y)dxdy \to \min$$

$$\text{minimizer (target function): } f_0(x) = \int_{\mathbb{R}} yp(y|x)dy$$

$$\min_f I[f] = \min_f \|f - f_0\|^2_{L^2((0,l),p(x)dx)} = \min_f \int_{\mathbb{R}} (f(x) - f_0(x))^2 p(x)dx$$

problem of learning 2

[T.Poggio, S.Smale, 2003]

$$\text{wish: } y = f(x)$$

$$p(x, y) \text{ — probability distribution on } \mathbb{R}^2$$

$$f(x)\text{: risk } I[f] = \int_{\mathbb{R}^2} (y - f(x))^2 p(x, y) dxdy \rightarrow \min$$

$$\text{minimizer (target function): } f_0(x) = \int_{\mathbb{R}} yp(y|x)dy$$

$$\min_f I[f] = \min_f \|f - f_0\|^2_{L^2((0,l),p(x)dx)} = \min_f \int_{\mathbb{R}} (f(x) - f_0(x))^2 p(x)dx$$

$$f \in \mathcal{H}_K \text{ — RKHS defined by a kernel } K(x, y)$$

RKHS defined by $K(x, y)$

$$K(x, y) = (xy)^{\alpha} + \beta \exp(-\gamma |x - y|^2)$$

symmetric positive definite function

RKHS defined by $K(x, y)$

$$K(x,y) = (xy)^\alpha + \beta \exp(-\gamma |x-y|^2)$$

symmetric positive definite function

$$f(x) = \sum_{i=1}^{n} a_i K(x, t_i)$$

$$g(x) = \sum_{j=1}^{m} b_j K(x, \tau_j)$$

$$(f, g)_{\mathcal{H}_K} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j K(t_i, \tau_j)$$

original goal

Find $f \in \mathcal{H}_K$ such that $\|\mathcal{T}_K f - f_0\|^2_{L_2((0,l),p\,dx)} \to \min.$

$\mathcal{T}_K : \mathcal{H}_K \to L_2((0,l), p\,dx)$ — inclusion operator.

available data

$$(x_i, y_i)\ i = 1, \ldots, n$$

available data

$$(x_i, y_i) \; i = 1, \ldots, n$$

sampling operator

$$S_x : \mathcal{H}_K \to \mathbb{R}^n$$
$$S_x : f \mapsto (f(x_i))_{i=1}^n$$

modified *regularized* goal

Find $f \in \mathcal{H}_K$ such that $\|S_x f - y\|^2 + \alpha \|f\|_{\mathcal{H}_K} \to \min$.

modified *regularized* goal

Find $f \in \mathcal{H}_K$ such that $\|S_x f - y\|^2 + \alpha \|f\|_{\mathcal{H}_K} \to \min$.

unique solution:

$$f_\alpha(\cdot) = \sum_{i=1}^{n} c_i K(x_i, \cdot)$$

nature of the problem

	ideally	in practice
operator	$\mathcal{T}_K : \mathcal{H}_K \to L_2((0,l), p\,dx)$	$S_x : \mathcal{H}_K \to \mathbb{R}^n$
data	$f_0 \in L_2((0,l), p\,dx)$	$y \in \mathbb{R}^n$

nature of the problem

	ideally	in practice
operator	$\mathcal{T}_K : \mathcal{H}_K \to L_2((0,l), p\,dx)$	$S_x : \mathcal{H}_K \to \mathbb{R}^n$
data	$f_0 \in L_2((0,l), p\,dx)$	$y \in \mathbb{R}^n$

ideal	considered
minimization problem	
$\Vert\mathcal{T}_K f - f_0\Vert_{L_2} \to \min$	$\Vert S_x f - y\Vert_{\mathbb{R}^n} \to \min$
operator equation	
$\mathcal{T}_K^* \mathcal{T}_K f = \mathcal{T}_K^* f_0$	$S_x^* S_x f = S_x^* y$

from LS to DRTLS 1

$A_0 : X \to Y,\ R(A_0) \subset Y$
$A_0 f = f_0,\ f_0 \notin R(A_0)$

from LS to DRTLS 1

$A_0 : X \to Y$, $R(A_0) \subset Y$
$A_0 f = f_0$, $f_0 \notin R(A_0)$

Least Squares solution:
Find $f \in X$ such that $\|A_0 f - f_0\| \to \min$

from LS to DRTLS 2

$\|f_0 - f_\delta\| \leq \delta$, $\|A_0 - A_h\| \leq h$

from LS to DRTLS 2

$\|f_0 - f_\delta\| \leq \delta$, $\|A_0 - A_h\| \leq h$

[S.V.Huffel, J. Vanderwalle, 1991], [G.H.Golub, P.C.Hansen,...]
Regularized Total Least Squares

$$\min_{A, f \in \{\|Bf\| \leq R\}} \{\|A - A_h\| + \|Af - f_\delta\|\}$$

$\|f_0 - f_\delta\| \leq \delta$, $\|A_0 - A_h\| \leq h$

[S.V.Huffel, J. Vanderwalle, 1991], [G.H.Golub, P.C.Hansen,...]
Regularized Total Least Squares

$$\min_{A,f \in \{\|Bf\| \leq R\}} \{\|A - A_h\| + \|Af - f_\delta\|\}$$

[S.Lu, S.Pereverzev Sen.,U.Tautenhahn, 2007]
Dual RTLS

$$\min_{\substack{A, f\, : \\ Af = g, \|A - A_h\| \leq h, \\ \|g - f_\delta\| \leq \delta.}} \{\|Bf\|\}$$

application to the prediction problem

$$\min_{\substack{A, f\,:\\ Af = g,\, \|A - T_x\| \le h,\\ \|g - S_x^* y\| \le \delta.}} \{\|Bf\|\}$$

application to the prediction problem

$$\min_{\substack{A,f:\\ Af=g,\|A-T_x\|\le h,\\ \|g-S_x^*y\|\le\delta.}} \{\|Bf\|\}$$

solution of the minimization problem satisfies:

$$(T_x^*T_x+\alpha B^*B+\beta I)f = T_x^*S_x^*y$$

α,β depend on h,δ;
choice is realized using an iterative method
based on model function approach
[S.Lu, S.Pereverzev Sen.,U.Tautenhahn, 2008]

h, δ, B

Theorem. With probability $(1-\eta)$, $\eta \in (0,1)$

$$\|\mathcal{T}_K^* f_0 - S_x^* y\| \leq \frac{c_1(\eta)}{\sqrt{n}} = \delta$$

$$\| T - T_x\| \leq \frac{c_2(\eta)}{\sqrt{n}} = h$$

$c_1, c_2 \sim \log \frac{1}{\eta}$.

h, δ, B

Theorem. With probability $(1-\eta)$, $\eta \in (0,1)$

$$\|\mathcal{T}_K^* f_0 - S_x^* y\| \leq \frac{c_1(\eta)}{\sqrt{n}} = \delta$$

$$\|T - T_x\| \leq \frac{c_2(\eta)}{\sqrt{n}} = h$$

$c_1, c_2 \sim \log \frac{1}{\eta}$.
Proof is based on [E.De Vito, L.Rosasco, et.al., 2005],
[I.F.Pinelis, A.I.Sakhanenko, 1985]

h, δ, B

Theorem. With probability $(1-\eta)$, $\eta \in (0,1)$

$$\|\mathcal{T}_K^* f_0 - S_x^* y\| \leq \frac{c_1(\eta)}{\sqrt{n}} = \delta$$

$$\|T - T_x\| \leq \frac{c_2(\eta)}{\sqrt{n}} = h$$

$c_1, c_2 \sim \log \frac{1}{\eta}$.
Proof is based on [E.De Vito, L.Rosasco, et.al., 2005], [I.F.Pinelis, A.I.Sakhanenko, 1985]

$B = S_x$

h, δ, B

Theorem. With probability $(1-\eta)$, $\eta \in (0,1)$

$$\|\mathcal{T}_K^* f_0 - S_x^* y\| \leq \frac{c_1(\eta)}{\sqrt{n}} = \delta$$

$$\|T - T_x\| \leq \frac{c_2(\eta)}{\sqrt{n}} = h$$

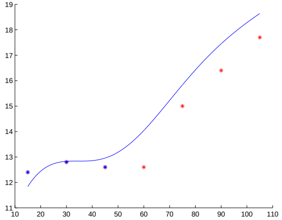
$c_1, c_2 \sim \log \frac{1}{\eta}$.
Proof is based on [E.De Vito, L.Rosasco, et.al., 2005],
[I.F.Pinelis, A.I.Sakhanenko, 1985]

$B = S_x$

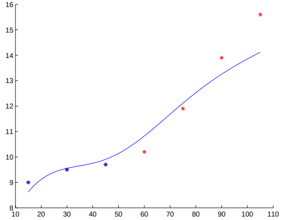
Sanaa Moussa, Master Thesis, 2009

numerical examples from EU-Project DIAdvisor

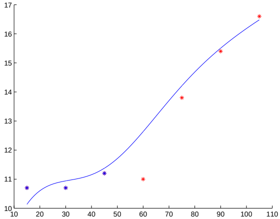
Day1:



Day2:



Day3:



Kernel $K(x,t) = (xt)^{0.35} + 0.005e^{-0.001(x-t)^2}$ in the interval [15, 105]; $h = 0.0001$, $\delta = 0.0001$

outlook

1. other choices of B?
2. $B = S_z$ with points $\{z_i\}$ that are different from $\{x_i\}$;
3. extensive comparison with Tikhonov ($B = I$);
4. another way of DRTLS-application:
   noisy operator is not $S_x^* S_x$ but S_x.