Lecture 6: Minimum encoding ball and Support vector data description (SVDD)

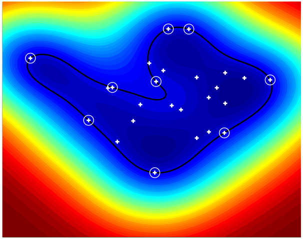
Stéphane Canu
stephane.canu@litislab.eu

Sao Paulo 2014

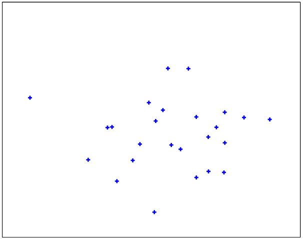
May 12, 2014

Plan

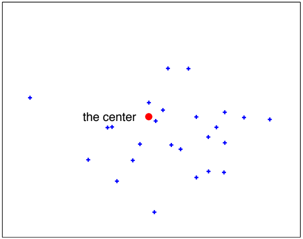
1. Support Vector Data Description (SVDD)
   - SVDD, the smallest enclosing ball problem
   - The minimum enclosing ball problem with errors
   - The minimum enclosing ball problem in a RKHS
   - The two class Support vector data description (SVDD)



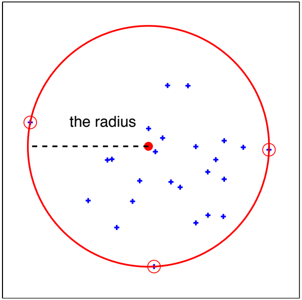
The minimum enclosing ball problem [Tax and Duin, 2004]



The minimum enclosing ball problem [Tax and Duin, 2004]



The minimum enclosing ball problem [Tax and Duin, 2004]



Given n points, $\{\mathbf{x}_i, i = 1, n\}$.

$$\begin{cases} \min\limits_{R \in \mathbb{R}, \mathbf{c} \in \mathbb{R}^d} & R^2 \\ \text{with} & \|\mathbf{x}_i - \mathbf{c}\|^2 \leq R^2, \quad i = 1, \dots, n \end{cases}$$

What is that in the convex programming hierarchy?

LP, QP, QCQP, SOCP and SDP

The convex programming hierarchy (part of)

LP

$$\begin{cases} \min\limits_{\mathbf{x}} & \mathbf{f}^\top \mathbf{x} \\ \text{with} & A\mathbf{x} \leq \mathbf{d} \\ \text{and} & 0 \leq \mathbf{x} \end{cases}$$

QCQP

$$\begin{cases} \min\limits_{\mathbf{x}} & \frac{1}{2}\mathbf{x}^\top G\mathbf{x} + \mathbf{f}^\top \mathbf{x} \\ \text{with} & \mathbf{x}^\top B_i\mathbf{x} + \mathbf{a}_i^\top \mathbf{x} \leq \mathbf{d}_i \\ & i = 1, n \end{cases}$$

QP

$$\begin{cases} \min\limits_{\mathbf{x}} & \frac{1}{2}\mathbf{x}^\top G\mathbf{x} + \mathbf{f}^\top \mathbf{x} \\ \text{with} & A\mathbf{x} \leq \mathbf{d} \end{cases}$$

SOCP

$$\begin{cases} \min\limits_{\mathbf{x}} & \mathbf{f}^\top \mathbf{x} \\ \text{with} & \|\mathbf{x} - \mathbf{a}_i\| \leq \mathbf{b}_i^\top \mathbf{x} + \mathbf{d}_i \\ & i = 1, n \end{cases}$$

The convex programming hierarchy?

Model generality: LP < QP < QCQP < SOCP < SDP

MEB as a QP in the primal

Theorem (MEB as a QP)

The two following problems are equivalent,

$$\begin{cases} \min\limits_{R\in\mathbb{R},\mathbf{c}\in\mathbb{R}^d} & R^2 \\ \text{with} & \|\mathbf{x}_i-\mathbf{c}\|^2 \le R^2, \quad i=1,\dots,n \end{cases} \qquad \begin{cases} \min\limits_{\mathbf{w},\rho} & \frac{1}{2}\|\mathbf{w}\|^2 - \rho \\ \text{with} & \mathbf{w}^\top\mathbf{x}_i \ge \rho + \frac{1}{2}\|\mathbf{x}_i\|^2 \end{cases}$$

with $\rho = \frac{1}{2}(\|\mathbf{c}\|^2 - R^2)$ *and* $\mathbf{w}=\mathbf{c}$.

Proof:

$$\begin{aligned} \|\mathbf{x}_i-\mathbf{c}\|^2 &\le R^2 \\ \|\mathbf{x}_i\|^2 - 2\mathbf{x}_i^\top\mathbf{c} + \|\mathbf{c}\|^2 &\le R^2 \\ -2\mathbf{x}_i^\top\mathbf{c} &\le R^2 - \|\mathbf{x}_i\|^2 - \|\mathbf{c}\|^2 \\ 2\mathbf{x}_i^\top\mathbf{c} &\ge -R^2 + \|\mathbf{x}_i\|^2 + \|\mathbf{c}\|^2 \\ \mathbf{x}_i^\top\mathbf{c} &\ge \underbrace{\frac{1}{2}(\|\mathbf{c}\|^2 - R^2)}_{\rho} + \frac{1}{2}\|\mathbf{x}_i\|^2 \end{aligned}$$

MEB and the one class SVM

SVDD: $\begin{cases} \min\limits_{\mathbf{w},\rho} & \frac{1}{2}\|\mathbf{w}\|^2 - \rho \\ \text{with} & \mathbf{w}^\top \mathbf{x}_i \geq \rho + \frac{1}{2}\|\mathbf{x}_i\|^2 \end{cases}$

SVDD and linear OCSVM (Supporting Hyperplane)

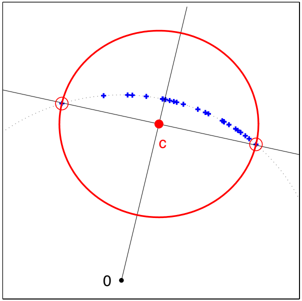
if $\forall i = 1, n, \ \|\mathbf{x}_i\|^2 = \text{constant}$, it is the the linear one class SVM (OC SVM)

The linear one class SVM [Schölkopf and Smola, 2002]

$$\begin{cases} \min\limits_{\mathbf{w},\rho'} & \frac{1}{2}\|\mathbf{w}\|^2 - \rho' \\ \text{with} & \mathbf{w}^\top \mathbf{x}_i \geq \rho' \end{cases}$$

with $\rho' = \rho + \frac{1}{2}\|\mathbf{x}_i\|^2$ $\Rightarrow$ OC SVM is a particular case of SVDD

When $\forall i = 1, n, \; \|\mathbf{x}_i\|^2 = 1$



$$\|\mathbf{x}_i - \mathbf{c}\|^2 \leq R^2 \qquad \Leftrightarrow \qquad \mathbf{w}^\top \mathbf{x}_i \geq \rho$$

with

$$\rho = \frac{1}{2}(\|\mathbf{c}\|^2 - R + 1)$$

SVDD and OCSVM

"Belonging to the ball" is also "being above" an hyperplane

MEB: KKT

$$\mathcal{L}(\mathbf{c}, R, \alpha) = R^2 + \sum_{i=1}^{n} \alpha_i \big(\|\mathbf{x}_i - \mathbf{c}\|^2 - R^2 \big)$$

KKT conditionns :

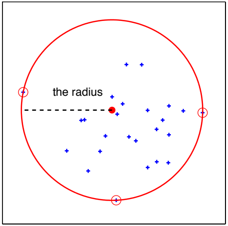
stationarty
- $2\mathbf{c} \sum_{i=1}^{n} \alpha_i - 2 \sum_{i=1}^{n} \alpha_i \mathbf{x}_i = 0 \quad \leftarrow$ The representer theorem
- $1 - \sum_{i=1}^{n} \alpha_i = 0$

primal admiss. $\|\mathbf{x}_i - \mathbf{c}\|^2 \le R^2$

dual admiss. $\alpha_i \ge 0 \qquad i = 1, n$

complementarity $\alpha_i \big(\|\mathbf{x}_i - \mathbf{c}\|^2 - R^2 \big) = 0 \qquad i = 1, n$

MEB: KKT



$$\mathcal{L}(\mathbf{c}, R, \alpha) = R^2 + \sum_{i=1}^{n} \alpha_i \big(\|\mathbf{x}_i - \mathbf{c}\|^2 - R^2\big)$$

KKT conditionns :

stationarty
- $2\mathbf{c}\sum_{i=1}^{n} \alpha_i - 2\sum_{i=1}^{n} \alpha_i \mathbf{x}_i = 0$ ← The representer theorem
- $1 - \sum_{i=1}^{n} \alpha_i = 0$

primal admiss. $\|\mathbf{x}_i - \mathbf{c}\|^2 \leq R^2$

dual admiss. $\alpha_i \geq 0$ $\quad i = 1, n$

complementarity $\alpha_i \big(\|\mathbf{x}_i - \mathbf{c}\|^2 - R^2\big) = 0$ $\quad i = 1, n$

Complementarity tells us: two groups of points

the support vectors $\|\mathbf{x}_i - \mathbf{c}\|^2 = R^2$ and the insiders $\alpha_i = 0$

MEB: Dual

The representer theorem:

$$\mathbf{c} = \frac{\sum_{i=1}^{n} \alpha_i \mathbf{x}_i}{\sum_{i=1}^{n} \alpha_i} = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i$$

$$\mathcal{L}(\alpha) = \sum_{i=1}^{n} \alpha_i \Big(\| \mathbf{x}_i - \sum_{j=1}^{n} \alpha_j \mathbf{x}_j \|^2 \Big)$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j x_i^\top x_j = \alpha^\top G \alpha \qquad \text{and} \qquad \sum_{i=1}^{n} \alpha_i x_i^\top x_i = \alpha^\top \text{diag}(G)$$

with $G = XX\top$ the Gram matrix: $G_{ij} = x_i^\top x_j$,

$$\begin{cases} \min\limits_{\alpha \in \mathbb{R}^n} & \alpha^\top G \alpha - \alpha^\top diag(G) \\ \text{with} & e^\top \alpha = 1 \\ \text{and} & 0 \leq \alpha_i, \qquad\qquad i = 1 \dots n \end{cases}$$

SVDD primal vs. dual

Primal

$$\begin{cases} \min\limits_{R \in \mathbb{R}, \mathbf{c} \in \mathbb{R}^d} & R^2 \\ \text{with} & \|\mathbf{x}_i - \mathbf{c}\|^2 \leq R^2, \\ & i = 1, \dots, n \end{cases}$$

- $d + 1$ unknown
- n constraints
- can be recast as a QP
- perfect when $d << n$

Dual

$$\begin{cases} \min\limits_{\alpha} & \alpha^\top G \alpha - \alpha^\top diag(G) \\ \text{with} & e^\top \alpha = 1 \\ \text{and} & 0 \leq \alpha_i, \\ & i = 1 \dots n \end{cases}$$

- n unknown with G the pairwise influence Gram matrix
- n box constraints
- easy to solve
- to be used when $d > n$

SVDD primal vs. dual

Primal

$$\begin{cases} \min\limits_{R\in\mathbb{R},\mathbf{c}\in\mathbb{R}^d} & R^2 \\ \text{with} & \|\mathbf{x}_i - \mathbf{c}\|^2 \leq R^2, \\ & i = 1, \dots, n \end{cases}$$

- $d + 1$ unknown
- n constraints
- can be recast as a QP
- perfect when $d << n$

Dual

$$\begin{cases} \min\limits_{\alpha} & \alpha^\top G\alpha - \alpha^\top diag(G) \\ \text{with} & e^\top \alpha = 1 \\ \text{and} & 0 \leq \alpha_i, \\ & i = 1 \dots n \end{cases}$$

- n unknown with G the pairwise influence Gram matrix
- n box constraints
- easy to solve
- to be used when $d > n$

But where is R^2?

Looking for R^2

$$\begin{cases} \min\limits_{\alpha} & \alpha^\top G\alpha - \alpha^\top diag(G) \\ \text{with} & e^\top \alpha = 1, \quad 0 \le \alpha_i, \qquad i = 1, n \end{cases}$$

The Lagrangian: $\mathcal{L}(\alpha, \mu, \beta) = \alpha^\top G\alpha - \alpha^\top diag(G) + \mu(e^\top \alpha - 1) - \beta^\top \alpha$

Stationarity cond.: $\nabla_\alpha \mathcal{L}(\alpha, \mu, \beta) = 2G\alpha - diag(G) + \mu e - \beta = 0$

The bi dual

$$\begin{cases} \min\limits_{\alpha} & \alpha^\top G\alpha + \mu \\ \text{with} & -2G\alpha + diag(G) \le \mu e \end{cases}$$

by identification

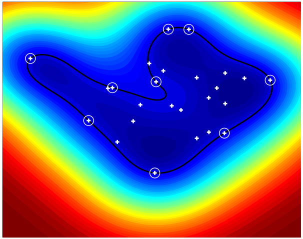
$$R^2 = \mu + \alpha^\top G\alpha = \mu + \|\mathbf{c}\|^2$$

μ is the Lagrange multiplier associated with the equality constraint $\sum_{i=1}^{n} \alpha_i = 1$

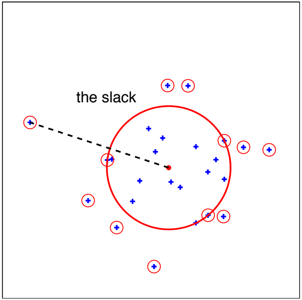
Also, because of the complementarity condition, if $\mathbf{x}_i$ is a support vector, then $\beta_i = 0$ implies $\alpha_i > 0$ and $R^2 = \|\mathbf{x}_i - \mathbf{c}\|^2$.

Plan

1. Support Vector Data Description (SVDD)
   - SVDD, the smallest enclosing ball problem
   - The minimum enclosing ball problem with errors
   - The minimum enclosing ball problem in a RKHS
   - The two class Support vector data description (SVDD)



The minimum enclosing ball problem with errors



The same road map:

- initial formuation
- reformulation (as a QP)
- Lagrangian, KKT
- dual formulation
- bi dual

Initial formulation: for a given C

$$\begin{cases} \min\limits_{R,a,\xi} & R^2 + C\sum\limits_{i=1}^{n} \xi_i \\ \text{with} & \|\mathbf{x}_i - \mathbf{c}\|^2 \leq R^2 + \xi_i, \quad i = 1, \dots, n \\ \text{and} & \xi_i \geq 0, \qquad\qquad\qquad i = 1, \dots, n \end{cases}$$

The MEB with slack: QP, KKT, dual and R^2

SVDD as a QP:

$$\begin{cases} \min_{\mathbf{w},\rho} & \frac{1}{2}\|\mathbf{w}\|^2 - \rho + \frac{C}{2}\sum_{i=1}^{n}\xi_i \\ \text{with} & \mathbf{w}^\top \mathbf{x}_i \geq \rho + \frac{1}{2}\|\mathbf{x}_i\|^2 - \frac{1}{2}\xi_i \\ \text{and} & \xi_i \geq 0, \\ & i = 1, n \end{cases}$$

again with OC SVM as a particular case.
With $G = XX^\top$

Dual SVDD:

$$\begin{cases} \min_{\alpha} & \alpha^\top G\alpha - \alpha^\top diag(G) \\ \text{with} & e^\top \alpha = 1 \\ \text{and} & 0 \leq \alpha_i \leq C, \\ & i = 1, n \end{cases}$$

for a given $C \leq 1$. If C is larger than one it is useless (it's the no slack case)

$$R^2 = \mu + \mathbf{c}^\top \mathbf{c}$$

with μ denoting the Lagrange multiplier associated with the equality constraint $\sum_{i=1}^{n} \alpha_i = 1$.

Variations over SVDD

- Adaptive SVDD: the weighted error case for given $w_i, i = 1, n$

$$\begin{cases} \min\limits_{c \in \mathbb{R}^p, R \in \mathbb{R}, \xi \in \mathbb{R}^n} & R + C \sum_{i=1}^{n} w_i \xi_i \\ \text{with} & \|\mathbf{x}_i - c\|^2 \leq R + \xi_i \\ & \xi_i \geq 0 \qquad i = 1, n \end{cases}$$

The dual of this problem is a QP [see for instance Liu et al., 2013]

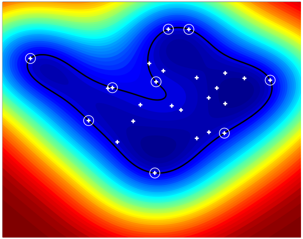
$$\begin{cases} \min\limits_{\alpha \in \mathbb{R}^n} & \alpha^\top XX^\top \alpha - \alpha^\top diag(XX^\top) \\ \text{with} & \sum_{i=1}^n \alpha_i = 1 \qquad\qquad 0 \leq \alpha_i \leq C w_i \qquad i = 1, n \end{cases}$$

- Density induced SVDD (D-SVDD):

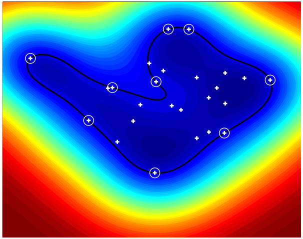
$$\begin{cases} \min\limits_{c \in \mathbb{R}^p, R \in \mathbb{R}, \xi \in \mathbb{R}^n} & R + C \sum_{i=1}^{n} \xi_i \\ \text{with} & w_i \|\mathbf{x}_i - c\|^2 \leq R + \xi_i \\ & \xi_i \geq 0 \qquad i = 1, n \end{cases}$$

Plan

1. Support Vector Data Description (SVDD)
   - SVDD, the smallest enclosing ball problem
   - The minimum enclosing ball problem with errors
   - The minimum enclosing ball problem in a RKHS
   - The two class Support vector data description (SVDD)



SVDD in a RKHS



The feature map:

$$\begin{array}{rcl} \mathbb{R}^p & \longrightarrow & \mathcal{H} \\ c & \longrightarrow & f(\bullet) \\ \mathbf{x}_i & \longrightarrow & k(\mathbf{x}_i, \bullet) \\ \|\mathbf{x}_i - c\|_{\mathbb{R}^p} \leq R^2 & \longrightarrow & \|k(\mathbf{x}_i, \bullet) - f(\bullet)\|^2_{\mathcal{H}} \leq R^2 \end{array}$$

Kernelized SVDD (in a RKHS) is also a QP

$$\begin{cases} \displaystyle\min_{f\in\mathcal{H}, R\in\mathbb{R}, \xi\in\mathbb{R}^n} & R^2 + C\displaystyle\sum_{i=1}^n \xi_i \\ \text{with} & \|k(\mathbf{x}_i, \bullet) - f(\bullet)\|^2_{\mathcal{H}} \leq R^2 + \xi_i \qquad i = 1, n \\ & \xi_i \geq 0 \qquad i = 1, n \end{cases}$$

SVDD in a RKHS: KKT, Dual and R^2

$$\begin{aligned}\mathcal{L} &= R^2 + C\sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i\big(\|k(\mathbf{x}_i,.) - f(.)\|_{\mathcal{H}}^2 - R^2 - \xi_i\big) - \sum_{i=1}^n \beta_i\xi_i \\ &= R^2 + C\sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i\big(k(\mathbf{x}_i,\mathbf{x}_i) - 2f(\mathbf{x}_i) + \|f\|_{\mathcal{H}}^2 - R^2 - \xi_i\big) - \sum_{i=1}^n \beta_i\xi_i\end{aligned}$$

KKT conditions

- Stationarity
  - $2f(.)\sum_{i=1}^n \alpha_i - 2\sum_{i=1}^n \alpha_i k(.,\mathbf{x}_i) = 0$ ← The representer theorem
  - $1 - \sum_{i=1}^n \alpha_i = 0$
  - $C - \alpha_i - \beta_i = 0$
- Primal admissibility: $\|k(\mathbf{x}_i,.) - f(.)\|^2 \leq R^2 + \xi_i$, $\xi_i \geq 0$
- Dual admissibility: $\alpha_i \geq 0$, $\beta_i \geq 0$
- Complementarity
  - $\alpha_i\big(\|k(\mathbf{x}_i,.) - f(.)\|^2 - R^2 - \xi_i\big) = 0$
  - $\beta_i\xi_i = 0$

SVDD in a RKHS: Dual and R^2

$$\begin{aligned}\mathcal{L}(\alpha) &= \sum_{i=1}^{n} \alpha_i k(\mathbf{x}_i, \mathbf{x}_i) - 2\sum_{i=1}^{n} f(\mathbf{x}_i) + \|f\|_{\mathcal{H}}^2 \qquad \text{with } f(.) = \sum_{j=1}^{n} \alpha_j k(., \mathbf{x}_j) \\ &= \sum_{i=1}^{n} \alpha_i k(\mathbf{x}_i, \mathbf{x}_i) - \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j \underbrace{k(\mathbf{x}_i, \mathbf{x}_j)}_{G_{ij}}\end{aligned}$$

$G_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$

$$\begin{cases} \min_{\alpha} & \alpha^\top G \alpha - \alpha^\top diag(G) \\ \text{with} & e^\top \alpha = 1 \\ \text{and} & 0 \leq \alpha_i \leq C, \qquad i = 1 \dots n \end{cases}$$

As it is in the linear case:

$$R^2 = \mu + \|f\|_{\mathcal{H}}^2$$

with μ denoting the Lagrange multiplier associated with the equality constraint $\sum_{i=1}^{n} \alpha_i = 1$.

SVDD train and val in a RKHS

Train using the dual form (in: G, C; out: α, μ)

$$\begin{cases} \min_{\alpha} & \alpha^\top G\alpha - \alpha^\top diag(G) \\ \text{with} & e^\top \alpha = 1 \\ \text{and} & 0 \leq \alpha_i \leq C, \qquad i = 1 \dots n \end{cases}$$

Val with the center in the RKHS: $f(.) = \sum_{i=1}^n \alpha_i k(., \mathbf{x}_i)$

$$\begin{aligned} \phi(\mathbf{x}) &= \|k(\mathbf{x}, .) - f(.)\|_{\mathcal{H}}^2 - R^2 \\ &= \|k(\mathbf{x}, .)\|_{\mathcal{H}}^2 - 2\langle k(\mathbf{x}, .), f(.)\rangle_{\mathcal{H}} + \|f(.)\|_{\mathcal{H}}^2 - R^2 \\ &= k(\mathbf{x}, \mathbf{x}) - 2f(\mathbf{x}) + R^2 - \mu - R^2 \\ &= -2f(\mathbf{x}) + k(\mathbf{x}, \mathbf{x}) - \mu \\ &= -2\sum_{i=1}^n \alpha_i k(\mathbf{x}, \mathbf{x}_i) + k(\mathbf{x}, \mathbf{x}) - \mu \end{aligned}$$

$\phi(\mathbf{x}) = 0$ is the decision border

An important theoretical result

For a well-calibrated bandwidth,
The SVDD estimates the underlying distribution level set [Vert and Vert, 2006]

The level sets of a probability density function $\mathbb{P}(\mathbf{x})$ are the set

$$C_p = \{\mathbf{x} \in \mathbb{R}^d \mid \mathbb{P}(\mathbf{x}) \geq p\}$$

It is well estimated by the empirical minimum volume set

$$V_p = \{\mathbf{x} \in \mathbb{R}^d \mid \|k(\mathbf{x}, .) - f(.)\|^2_{\mathcal{H}} - R^2 \geq 0\}$$

The frontiers coincides

SVDD: the generalization error

For a well-calibrated bandwidth,

$$(\mathbf{x}_1, \ldots, \mathbf{x}_n) \text{ i.i.d. from some fixed but unknown } \mathbb{P}(\mathbf{x})$$

Then [Shawe-Taylor and Cristianini, 2004] with probability at least $1 - \delta$, $(\forall \delta \in]0, 1[)$, for any margin $m > 0$

$$\mathbb{P}\big(\|k(\mathbf{x}, .) - f(.)\|^2_{\mathcal{H}} \geq R^2 + m\big) \leq \frac{1}{mn}\sum_{i=1}^{n} \xi_i + \frac{6R^2}{m\sqrt{n}} + 3\sqrt{\frac{ln(2/\delta)}{2n}}$$

Equivalence between SVDD and OCSVM for translation invariant kernels (diagonal constant kernels)

Theorem

Let $\mathcal{H}$ be a RKHS on some domain $\mathcal{X}$ endowed with kernel k. If there exists some constant c such that $\forall \mathbf{x} \in \mathcal{X},\ k(\mathbf{x}, \mathbf{x}) = c$, then the two following problems are equivalent,

$$\begin{cases} \displaystyle\min_{f,R,\xi} & R + C\sum_{i=1}^{n} \xi_i \\ \text{with} & \|k(\mathbf{x}_i, .) - f(.)\|_{\mathcal{H}}^2 \le R + \xi_i \\ & \xi_i \ge 0 \qquad i = 1, n \end{cases}$$

$$\begin{cases} \displaystyle\min_{f,\rho,\xi} & \frac{1}{2}\|f\|_{\mathcal{H}}^2 - \rho + C\sum_{i=1}^{n} \varepsilon_i \\ \text{with} & f(\mathbf{x}_i) \ge \rho - \varepsilon_i \\ & \varepsilon_i \ge 0 \qquad i = 1, n \end{cases}$$

with $\rho = \frac{1}{2}(c + \|f\|_{\mathcal{H}}^2 - R)$ *and* $\varepsilon_i = \frac{1}{2}\xi_i$.

Proof of the Equivalence between SVDD and OCSVM

$$\begin{cases} \min\limits_{f\in\mathcal{H},R\in\mathbb{R},\xi\in\mathbb{R}^n} & R + C\sum\limits_{i=1}^{n}\xi_i \\ \text{with} & \|k(\mathbf{x}_i,.) - f(.)\|_{\mathcal{H}}^2 \le R + \xi_i, \quad \xi_i \ge 0 \qquad i = 1,n \end{cases}$$

since $\|k(\mathbf{x}_i,.) - f(.)\|_{\mathcal{H}}^2 = k(\mathbf{x}_i,\mathbf{x}_i) + \|f\|_{\mathcal{H}}^2 - 2f(\mathbf{x}_i)$

$$\begin{cases} \min\limits_{f\in\mathcal{H},R\in\mathbb{R},\xi\in\mathbb{R}^n} & R + C\sum\limits_{i=1}^{n}\xi_i \\ \text{with} & 2f(\mathbf{x}_i) \ge k(\mathbf{x}_i,\mathbf{x}_i) + \|f\|_{\mathcal{H}}^2 - R - \xi_i, \quad \xi_i \ge 0 \qquad i = 1,n. \end{cases}$$

Introducing $\rho = \frac{1}{2}(c + \|f\|_{\mathcal{H}}^2 - R)$ that is $R = c + \|f\|_{\mathcal{H}}^2 - 2\rho$, and since $k(\mathbf{x}_i,\mathbf{x}_i)$ is constant and equals to c the SVDD problem becomes

$$\begin{cases} \min\limits_{f\in\mathcal{H},\rho\in\mathbb{R},\xi\in\mathbb{R}^n} & \frac{1}{2}\|f\|_{\mathcal{H}}^2 - \rho + \frac{C}{2}\sum\limits_{i=1}^{n}\xi_i \\ \text{with} & f(\mathbf{x}_i) \ge \rho - \frac{1}{2}\xi_i, \qquad \xi_i \ge 0 \qquad i = 1,n \end{cases}$$

leading to the classical one class SVM formulation (OCSVM)

$$\begin{cases} \displaystyle\min_{f \in \mathcal{H}, \rho \in \mathbb{R}, \xi \in \mathbb{R}^n} & \frac{1}{2}\|f\|_{\mathcal{H}}^2 - \rho + C\sum_{i=1}^{n} \varepsilon_i \\ \text{with} & f(\mathbf{x}_i) \geq \rho - \varepsilon_i, \qquad \varepsilon_i \geq 0 \qquad i = 1, n \end{cases}$$

with $\varepsilon_i = \frac{1}{2}\xi_i$. Note that by putting $\nu = \frac{1}{nC}$ we can get the so called ν formulation of the OCSVM

$$\begin{cases} \displaystyle\min_{f' \in \mathcal{H}, \rho' \in \mathbb{R}, \xi' \in \mathbb{R}^n} & \frac{1}{2}\|f'\|_{\mathcal{H}}^2 - n\nu\rho' + \sum_{i=1}^{n} \xi'_i \\ \text{with} & f'(\mathbf{x}_i) \geq \rho' - \xi'_i, \qquad \xi'_i \geq 0 \qquad i = 1, n \end{cases}$$

with $f' = Cf$, $\rho' = C\rho$, and $\xi' = C\xi$.

Duality

Note that the dual of the SVDD is

$$\begin{cases} \min\limits_{\alpha \in \mathbb{R}^n} & \alpha^\top G \alpha - \alpha^\top g \\ \text{with} & \sum_{i=1}^n \alpha_i = 1 \quad\quad 0 \le \alpha_i \le C \quad\quad\quad i = 1, n \end{cases}$$

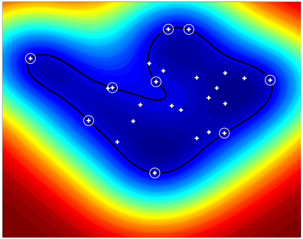
where G is the kernel matrix of general term $G_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$ and g the diagonal vector such that $g_i = k(\mathbf{x}_i, \mathbf{x}_i) = c$. The dual of the OCSVM is the following equivalent QP

$$\begin{cases} \min\limits_{\alpha \in \mathbb{R}^n} & \frac{1}{2} \alpha^\top G \alpha \\ \text{with} & \sum_{i=1}^n \alpha_i = 1 \quad 0 \le \alpha_i \le C \quad\quad\quad i = 1, n \end{cases}$$

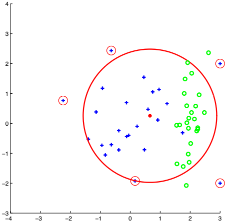
Both dual forms provide the same solution α, but not the same Lagrange multipliers. ρ is the Lagrange multiplier of the equality constraint of the dual of the OCSVM and $R = c + \alpha^\top G \alpha - 2\rho$. Using the SVDD dual, it turns out that $R = \lambda_{eq} + \alpha^\top G \alpha$ where λ_{eq} is the Lagrange multiplier of the equality constraint of the SVDD dual form.

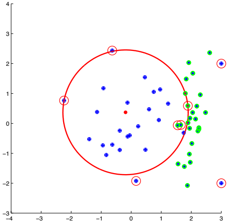
Plan

1. Support Vector Data Description (SVDD)
   - SVDD, the smallest enclosing ball problem
   - The minimum enclosing ball problem with errors
   - The minimum enclosing ball problem in a RKHS
   - The two class Support vector data description (SVDD)



The two class Support vector data description (SVDD)





$$\begin{cases} \min\limits_{\mathbf{c},R,\xi^+,\xi^-} & R^2 + C\Big(\sum\limits_{y_i=1} \xi_i^+ + \sum\limits_{y_i=-1} \xi_i^-\Big) \\ \text{with} & \|\mathbf{x}_i - \mathbf{c}\|^2 \leq R^2 + \xi_i^+, \qquad \xi_i^+ \geq 0 \qquad i \text{ such that } y_i = 1 \\ \text{and} & \|\mathbf{x}_i - \mathbf{c}\|^2 \geq R^2 - \xi_i^-, \qquad \xi_i^- \geq 0 \qquad i \text{ such that } y_i = -1 \end{cases}$$

The two class SVDD as a QP

$$\begin{cases} \displaystyle\min_{\mathbf{c},R,\xi^+,\xi^-} & R^2+C\Big(\sum_{y_i=1}\xi_i^+ + \sum_{y_i=-1}\xi_i^-\Big) \\ \text{with} & \|\mathbf{x}_i-\mathbf{c}\|^2 \le R^2+\xi_i^+, \qquad \xi_i^+\ge 0 \qquad i \text{ such that } y_i=1 \\ \text{and} & \|\mathbf{x}_i-\mathbf{c}\|^2 \ge R^2-\xi_i^-, \qquad \xi_i^-\ge 0 \qquad i \text{ such that } y_i=-1 \end{cases}$$

$$\begin{cases} \|\mathbf{x}_i\|^2 - 2\mathbf{x}_i^\top\mathbf{c} + \|\mathbf{c}\|^2 \le R^2+\xi_i^+, \qquad \xi_i^+\ge 0 \qquad i \text{ such that } y_i=1 \\ \|\mathbf{x}_i\|^2 - 2\mathbf{x}_i^\top\mathbf{c} + \|\mathbf{c}\|^2 \ge R^2-\xi_i^-, \qquad \xi_i^-\ge 0 \qquad i \text{ such that } y_i=-1 \end{cases}$$

$$\begin{aligned} 2\mathbf{x}_i^\top\mathbf{c} &\ge \|\mathbf{c}\|^2 - R^2 + \|\mathbf{x}_i\|^2-\xi_i^+, \qquad \xi_i^+\ge 0 \qquad i \text{ such that } y_i=1 \\ -2\mathbf{x}_i^\top\mathbf{c} &\ge -\|\mathbf{c}\|^2 + R^2 - \|\mathbf{x}_i\|^2-\xi_i^-, \qquad \xi_i^-\ge 0 \qquad i \text{ such that } y_i=-1 \end{aligned}$$

$$2y_i\mathbf{x}_i^\top\mathbf{c} \ge y_i(\|\mathbf{c}\|^2 - R^2 + \|\mathbf{x}_i\|^2)-\xi_i, \qquad \xi_i\ge 0 \qquad i=1,n$$

change variable: $\rho = \|\mathbf{c}\|^2 - R^2$

$$\begin{cases} \displaystyle\min_{\mathbf{c},\rho,\xi} & \|\mathbf{c}\|^2 - \rho + C\sum_{i=1}^n \xi_i \\ \text{with} & 2y_i\mathbf{x}_i{}^\top\mathbf{c} \ge y_i(\rho - \|\mathbf{x}_i\|^2)-\xi_i \qquad i=1,n \\ \text{and} & \xi_i \ge 0 \qquad i=1,n \end{cases}$$

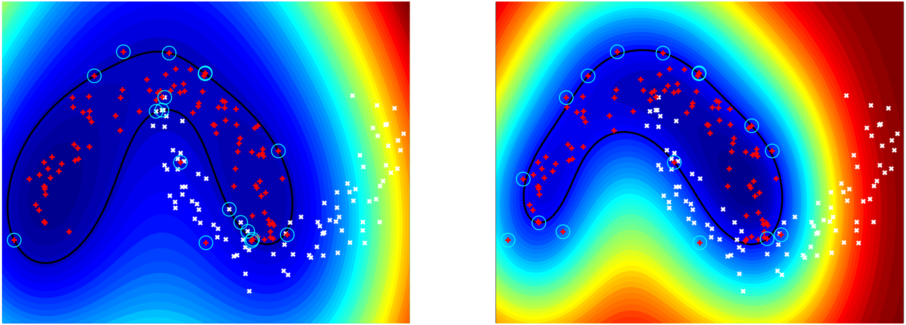
The dual of the two class SVDD

$G_{ij} = y_i y_j \mathbf{x}_i \mathbf{x}_j^\top$

The dual formulation:

$$\begin{cases} \min\limits_{\alpha \in \mathbb{R}^n} & \alpha^\top G \alpha - \sum_{i=1}^n \alpha_i y_i \|x_i\|^2 \\ \text{with} & \sum\limits_{i=1}^n y_i \alpha_i = 1 \\ & 0 \le \alpha_i \le C \qquad i = 1, n \end{cases}$$

The two class SVDD vs. one class SVDD



The two class SVDD (left) vs. the one class SVDD (right)

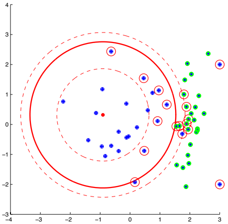
Small Sphere and Large Margin (SSLM) approach

Support vector data description with margin [Wu and Ye, 2009]

$$\begin{cases} \min\limits_{\mathbf{w},R,\xi\in\mathbb{R}^n} & R^2 + C\Big(\sum\limits_{y_i=1}\xi_i^+ + \sum\limits_{y_i=-1}\xi_i^-\Big) \\ \text{with} & \|\mathbf{x}_i-\mathbf{c}\|^2 \le R^2 - 1 + \xi_i^+, \quad \xi_i^+\ge 0 \qquad i \text{ such that } y_i = 1 \\ \text{and} & \|\mathbf{x}_i-\mathbf{c}\|^2 \ge R^2 + 1 - \xi_i^-, \quad \xi_i^-\ge 0 \qquad i \text{ such that } y_i = -1 \end{cases}$$

$$\|\mathbf{x}_i-\mathbf{c}\|^2 \ge R^2 + 1 - \xi_i^- \text{ and } y_i = -1 \quad \Longleftrightarrow \quad y_i\|\mathbf{x}_i-\mathbf{c}\|^2 \le y_i R^2 - 1 + \xi_i^-$$

$$\mathcal{L}(\mathbf{c},R,\xi,\alpha,\beta) = R^2 + C\sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i\big(y_i\|\mathbf{x}_i-\mathbf{c}\|^2 - y_iR^2 + 1 - \xi_i\big) - \sum_{i=1}^n \beta_i\xi_i$$



SVDD with margin – dual formulation

$$\mathcal{L}(\mathbf{c}, R, \xi, \alpha, \beta) = R^2 + C\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}\alpha_i\big(y_i\|\mathbf{x}_i - \mathbf{c}\|^2 - y_iR^2 + 1 - \xi_i\big) - \sum_{i=1}^{n}\beta_i\xi_i$$

Optimality: $\mathbf{c} = \sum_{i=1}^{n}\alpha_i y_i \mathbf{x}_i \quad ; \quad \sum_{i=1}^{n}\alpha_i\, y_i = 1 \quad ; \quad 0 \le \alpha_i \le C$

$$\begin{aligned}\mathcal{L}(\alpha) &= \sum_{i=1}^{n}\alpha_i\Big(y_i\|\mathbf{x}_i - \sum_{j=1}^{n}\alpha_i y_j \mathbf{x}_j\|^2\Big) + \sum_{i=1}^{n}\alpha_i \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_j\alpha_i y_i y_j \mathbf{x}_j^\top \mathbf{x}_i + \sum_{i=1}^{n}\|\mathbf{x}_i\|^2 y_i\alpha_i + \sum_{i=1}^{n}\alpha_i\end{aligned}$$

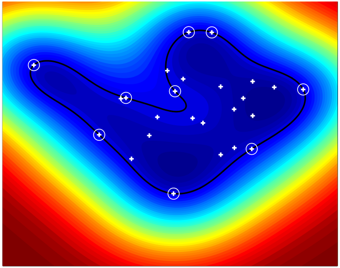
Dual SVDD is also a quadratic program

$$\text{problem } \mathcal{D} \quad \begin{cases} \min\limits_{\alpha \in \mathbb{R}^n} & \alpha^\top G\alpha - \mathbf{e}^\top\alpha - \mathbf{f}^\top\alpha \\ \text{with} & \mathbf{y}^\top\alpha = 1 \\ \text{and} & 0 \le \alpha_i \le C \qquad\qquad i = 1, n \end{cases}$$

with G a symmetric matrix $n \times n$ such that $G_{ij} = y_i y_j \mathbf{x}_j^\top \mathbf{x}_i$ and $f_i = \|\mathbf{x}_i\|^2 y_i$

Conclusion

- Applications
  - outlier detection
  - change detection
  - clustering
  - large number of classes
  - variable selection, ...
- A clear path
  - reformulation (to a standart problem)
  - KKT
  - Dual
  - Bidual
- a lot of variations
  - L^2 SVDD
  - two classes non symmetric
  - two classes in the symmetric classes (SVM)
  - the multi classes issue
- practical problems with translation invariant kernels



Bibliography

Bo Liu, Yanshan Xiao, Longbing Cao, Zhifeng Hao, and Feiqi Deng. Svdd-based outlier detection on uncertain data. *Knowledge and information systems*, 34(3):597–618, 2013.

B. Schölkopf and A. J. Smola. *Learning with Kernels*. MIT Press, 2002.

John Shawe-Taylor and Nello Cristianini. *Kernel methods for pattern analysis*. Cambridge university press, 2004.

David MJ Tax and Robert PW Duin. Support vector data description. *Machine learning*, 54(1):45–66, 2004.

Régis Vert and Jean-Philippe Vert. Consistency and convergence rates of one-class svms and related algorithms. *The Journal of Machine Learning Research*, 7:817–854, 2006.

Mingrui Wu and Jieping Ye. A small sphere and large margin approach for novelty detection using training data with outliers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(11):2088–2092, 2009.