Lecture 10: Robust outlier detection with L0-SVDD

Stéphane Canu
stephane.canu@litislab.eu

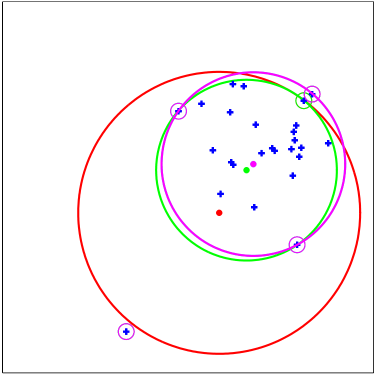
Sao Paulo 2014

February 28, 2014

Roadmap

1. Robust outlier detection with L0-SVDD
   - L_0 SVDD

4 iterations of Adaptive L0 SVDD



Recall SVDD

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

SVDD + outlier

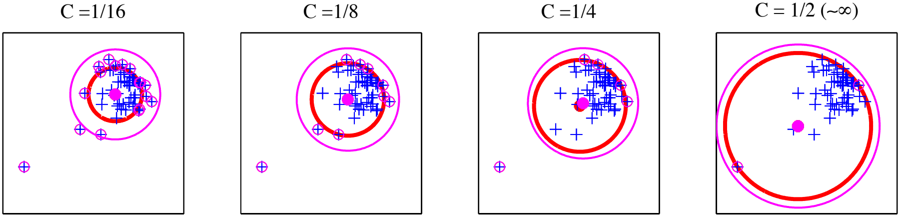


Figure: Example of SVDD solutions with different C values, $m = 0$ (red) and $m = 5$ (magenta). The circled data points represent support vectors for both m.

The L_0 norm

$$\|\xi\|_0 \leq t$$

$$\left\{ \begin{array}{ll} \min\limits_{c \in \mathbb{R}^p, R \in \mathbb{R}, \xi \in \mathbb{R}^n} & R + C\|\xi\|_0 \\ \text{with} & \|\mathbf{x}_i - c\|^2 \leq R + \xi_i \\ & \xi_i \geq 0 \qquad i = 1, n \end{array} \right.$$

L_0 relaxations

- p norm
- exponenetial
- piecewise linear
- log

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

DC programing

$$\log(\gamma + t) = f(t) - g(t) \qquad \text{with } f(t) = t \quad \text{and} \quad g(t) = t - \log(\gamma + t),$$

both functions f and g being convex. The DC framework consists in minimizing iteratively (R plus a sum of) the following convex term:

$$f(\xi) - g'(\xi)\xi = \xi - \left(1 - \frac{1}{\gamma + \xi^{\text{old}}}\right)\xi = \frac{\xi}{\gamma + \xi^{\text{old}}},$$

where ξ_i^{old} denotes the solution at the previous iteration.

The DC idea applied to our L_0 SVDD approximation consists in building a sequence of solutions of the following adaptive SVDD:

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

Stationary conditions of the KKT give: $c = \sum_{i=1}^n \alpha_i x_i$ and $\sum_{i=1}^n \alpha_i = 1$ where the α_i are the Lagrange multipliers associated with the inequality constraints $\|\mathbf{x}_i - c\|^2 \leq R + \xi_i$. The dual of this problem is

$$\begin{cases} \min\limits_{\alpha \in \mathbb{R}^n} & \alpha^\top X X^\top \alpha - \alpha^\top diag(XX^\top) \\ \text{with} & \sum_{i=1}^n \alpha_i = 1 \qquad 0 \leq \alpha_i \leq C w_i \qquad i = 1, n \end{cases} \tag{2}$$

Algorithm 1 L_0 SVDD for the linear kernel

Data: X, y, C , γ
Result: R , c, ξ , α
$w_i = 1; \quad i = 1, n$
while *not converged* **do**
  $(\alpha, \lambda) \leftarrow$ solve_QP(X, C, w) % solve problem (2)
  $c \leftarrow X^\top \alpha$
  $R \leftarrow \lambda + c^\top c$
  $\xi_i \leftarrow \max(0, \|\mathbf{x}_i - c\|^2 - R) \quad i = 1, n$
  $w_i \leftarrow 1/(\gamma + \xi_i) \quad i = 1, n$
end

Bibliography