Support Vector Data Description for Uncertainty Data Sets

Jorge Guevara[1], **Stephane Canu**[2], **Roberto Hirata Jr**[1]

[1]Department of Computer Science, Institute of Mathematics and Statistics-USP, São Paulo - Brazil.

[2]LITIS EA 4108, INSA de Rouen, 76800 Saint Etienne du Rouvray, France

Introduction

This work shows how to estimate the support of the distribution of some data when observations in the data have uncertainties. To model uncertainties, we consider each observation of the training set to be a random vector distributed according to a distribution with first and second moments in a local vicinity. To estimate the support, we used the support vector data description method.

Chance Constrain Approach

- Let $\{\mathbf{x}_i \sim (\hat{\mathbf{x}}_i, \Sigma_i)\}_{i=1}^n$ be the training set, the probabilistic levels $\kappa_i \in [0,1],\ i = 1, 2 \ldots, n$, and $C > 0$, USVDD seeks to minimize the radius of the hypersphere that encloses most of the uncertainty points. The chance constraint formulation is

Problem

$$\begin{aligned} \min_{c \in \mathcal{H}, R \in \mathbb{R}} \quad & R^2 + C \sum_{i=1}^n \xi_i \\ \text{subject to} \quad & \mathbb{P}_{\mathbf{x}_i \sim (\bar{\mathbf{x}}_i, \Sigma_i)} (\|\mathbf{x}_i - c\|^2 \leq R^2 + \xi_i) \geq 1 - \kappa_i,\ i = 1, \ldots, n \\ & \xi_i \geq 0,\ i = 1, \ldots, n. \end{aligned}$$

- **Interpretation** : *The probability that the random vector $\mathbf{x}_i$ takes its values outside the sphere of radius R and center $\mathbf{c}$ is less or equal than κ_i*. Example, if $\kappa_i = \kappa,\ i = 1, 2, \ldots, n$ are small values, then R will increase, i.e., the probability that $\mathbf{x}_i$ will be outside the sphere will be small.
- **Some Lemmas**

Lema

$$\mathbb{E}_{\mathbf{x} \sim (\hat{\mathbf{x}}, \Sigma)}[\|\mathbf{x} - \mathbf{c}\|^2] = tr(\Sigma) + \|\hat{\mathbf{x}} - \mathbf{c}\|^2,\ \mathbf{x} \in \mathbb{R}^d$$

Lema

The probabilistic constraint $\mathbb{P}_{\mathbf{x} \sim (\bar{\mathbf{x}}, \Sigma)} (\|\mathbf{x} - c\|^2 \geq R^2 + \xi)$, *is bounded by (Markov's inequality)*

$$\frac{tr(\Sigma) + \|\bar{\mathbf{x}} - \mathbf{c}\|^2}{R^2 + \xi}$$

Forcing this bound to be less or equal than a given value κ_i for each constraint, i.e.,

$$\frac{tr(\Sigma_i) + \|\hat{\mathbf{x}}_i - \mathbf{c}\|^2}{R^2 + \xi_i} \leq \kappa_i,\ i = 1, 2, \ldots, n, \tag{1}$$

permits us to control the size of the hypersphere that encloses the data

Definition (USVDD)

Given the training dataset $\{\mathbf{x}_i \sim (\bar{\mathbf{x}}_i, \Sigma_i)\}_{i=1}^n$, the probabilistic levels $\kappa_i \in [0,1],\ i = 1, 2 \ldots, n$, and $C > 0$, the support vector data description for uncertainty data is given by

Problem

$$\begin{aligned} \min_{\mathbf{c} \in \mathbb{R}^d, R \in \mathbb{R}, \boldsymbol{\xi} \in \mathbb{R}^n} \quad & R^2 + C \sum_{i=1}^n \xi_i \\ \text{subject to} \quad & \|\hat{\mathbf{x}}_i - \mathbf{c}\|^2 \leq (R^2 + \xi_i)\kappa_i - tr(\Sigma_i),\ i = 1, \ldots, n \\ & \xi_i \geq 0,\ i = 1, \ldots, n. \end{aligned}$$

References

[1] David M. J. Tax and Robert P. W. Duin. 2004. "**Support Vector Data Description**." *Mach. Learn. 54, 1 (January 2004)*, 45-66.

[2] Bernhard Schölkopf, John C. Platt, John C. Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. 2001. "**Estimating the Support of a High-Dimensional Distribution**." *Neural Comput. 13, 7 (July 2001)*, 1443-1471.

[3] Kendrick, David A. "**Stochastic control for economic models**." *New York: McGraw-Hill, 1981.*

Lagrangian and KKT's Conditions

Lagrangian

$$\mathcal{L}(R, \mathbf{c}, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = R^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \{(R^2 + \xi_i)\kappa_i - \|\hat{\mathbf{x}}_i - \mathbf{c}\|^2 - tr(\Sigma_i)\} - \sum_{i=1}^n \beta \xi_i$$

Karush–Kuhn–Tucker (KKT) conditions

$$\begin{array}{lcr} \partial_R \mathcal{L} = 0 : & \Sigma_{i=1}^n \alpha_i \kappa_i & 1 \\ \nabla_{\mathbf{c}} \mathcal{L} = 0 : & -2 \Sigma_{i=1}^n \alpha_i \hat{\mathbf{x}}_i + 2 \Sigma_{i=1}^n \alpha_i \mathbf{c} = & 0 \\ \nabla_{\boldsymbol{\xi}} \mathcal{L} = 0 : & C\mathbf{1}_n - diag(\boldsymbol{\alpha\kappa}\top) - \boldsymbol{\beta} & \mathbf{0}_n \end{array}$$

$$\begin{array}{c} \alpha_i \{(R^2 + \xi_i)\kappa_i - \|\hat{\mathbf{x}}_i - \mathbf{c}\|^2 - tr(\Sigma_i)\} \\ \beta_i \xi_i \end{array} = \left.\begin{array}{c} 0 \\ 0 \end{array}\right\} 1, 2, \ldots, n$$

Definition (USVDD Dual Form)

Given the training dataset $\{\mathbf{x}_i \sim (\bar{\mathbf{x}}_i, \Sigma_i)\}_{i=1}^n$, $\kappa_i \in [0,1],\ i = 1, 2 \ldots, n$, and $C > 0$, the dual form of USVDD is

Problem

$$\begin{aligned} \max_{\boldsymbol{\alpha} \in \mathbb{R}^n} \quad & \sum_{i=1}^n \alpha_i \langle \hat{\mathbf{x}}_i, \hat{\mathbf{x}}_i \rangle - \frac{\Sigma_{i,j=1}^n \alpha_i \alpha_j \langle \hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j \rangle}{\Sigma_{i=1}^n \alpha_i} + \sum_{i=1}^n \alpha_i tr(\Sigma_i) \\ \text{subject to} \quad & 0 \leq \sum_{i=1}^n \alpha_i \kappa_i = 1, \quad \alpha_i \kappa_i \leq C, \quad i = 1, \ldots, n \end{aligned}$$

$$\mathbf{c} = \frac{\Sigma_i \alpha_i \hat{\mathbf{x}}_i}{\Sigma_{i \in} \alpha_i},\ i \in \{i | 0 < \alpha_i \kappa_i \leq C\},\ \text{From KKT's :} \tag{2}$$

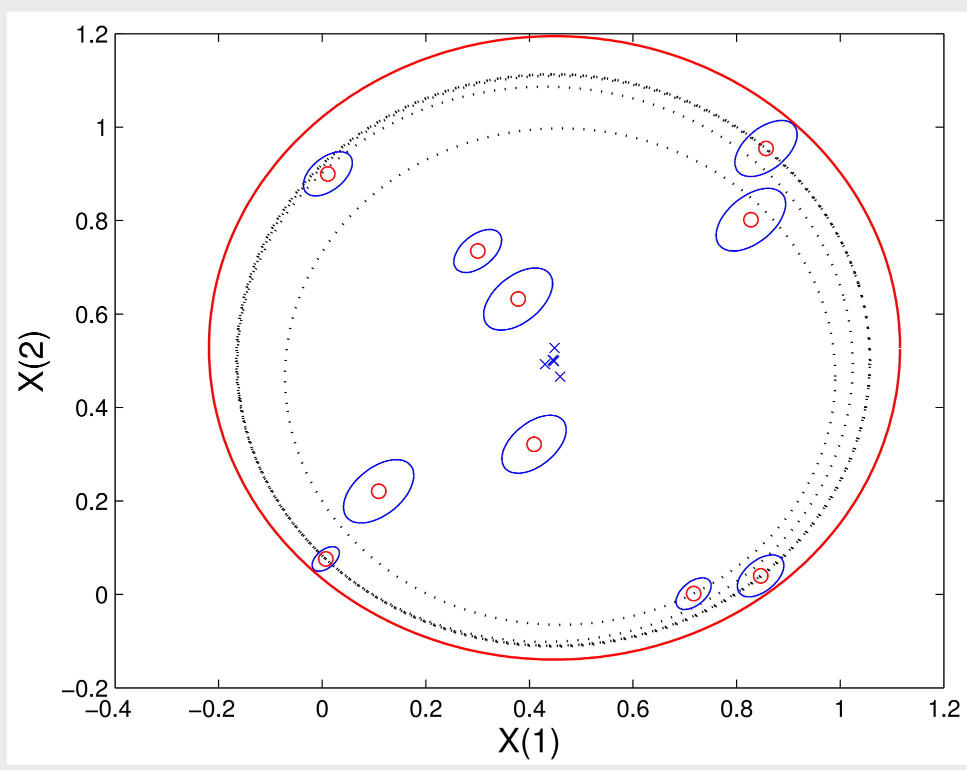
Analysis

- $\alpha_i = 0, \beta_i > 0 \implies \xi_i = 0$. Then $\mathbf{x}_i \sim (\hat{\mathbf{x}}_i, \Sigma_i),\ i \in \{i | \alpha_i = 0\}$ lies *inside* the hypersphere *no matters the value for* κ_i, Points $\{\mathbf{x}' | \|\mathbf{x}' - \mathbf{c}\|^2 = (\|\hat{\mathbf{x}}_i - \mathbf{c}\|^2 + tr(\Sigma_i))/\kappa_i\}$, will be *inside* the hypersphere.
- $\alpha_i > 0, \beta_i = 0 \implies \xi_i > 0$. Then $\mathbf{x}_i \sim (\hat{\mathbf{x}}_i, \Sigma_i),\ i \in \{i | \alpha_i \kappa_i = C\}$ lies *outside* the hypersphere with probability κ_i. Points $\{\mathbf{x}' | \|\mathbf{x}' - \mathbf{c}\|^2 = (\|\hat{\mathbf{x}}_i - \mathbf{c}\|^2 + tr(\Sigma_i))/\kappa_i\}$, will be *outside* the hypersphere.
- $\alpha_i > 0, \beta_i > 0 \implies \xi_i = 0$ and $0 < \alpha_i \kappa_i < C$. From this and (2)

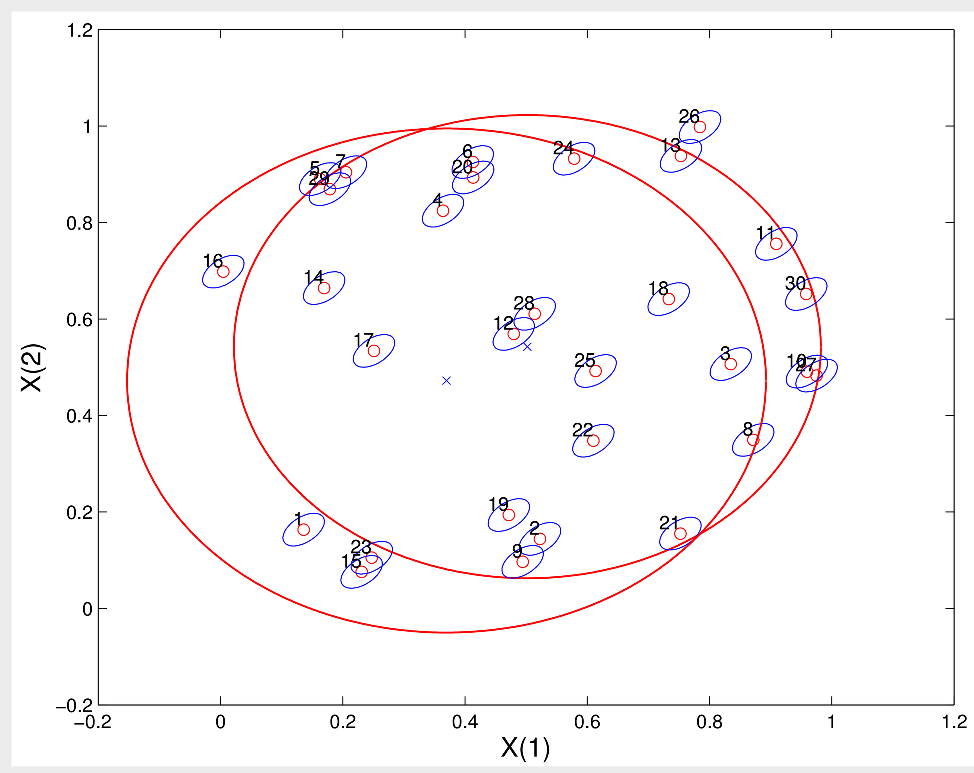
$$R^2 = \frac{\|\hat{\mathbf{x}}_i - \mathbf{c}\|^2 + tr(\Sigma_i)}{\kappa_i},\ i \in \{i | 0 < \alpha_i \kappa_i < C\}. \tag{3}$$

Points $\{\mathbf{x}' | \|\mathbf{x}' - \mathbf{c}\|^2 = (\|\hat{\mathbf{x}}_i - \mathbf{c}\|^2 + tr(\Sigma_i))/\kappa_i\}$, will be *in* the hypersphere.

Figures



USVDD vs SVDD. Red line: USVDD solution, Dashed lines: several SVDD solutions for $C = \{2^{-3}, \ldots 2^{15}\}$



Two different USVDD solutions for the same dataset and same C but with different probabilistic levels between both problems, The first one has values $\kappa_1 = 0.95$, $kappa_3 = 0.99$, $\kappa_{16} = 0.93$ and the other one has values $\kappa_1 = 0.1$, $kappa_3 = 1$, $\kappa_{16} = 0.2$

Conclusion

a) Probabilistic values (κ_i) values associated $\mathbf{x}_i \sim (\hat{\mathbf{x}}_i, \Sigma_i)$ inside the hypersphere are not important. b) By controlling some specific probabilistic value, i.e., κ_i, we will construct a hypersphere that encloses or not the associated point, i.e.,$\mathbf{x}_i$. c) USVDD equals to SVDD solutions if $\kappa_i = 1$, $tr(\Sigma_i) = 0$, $\forall i = 1, 2 \ldots, n$, i.e., it is not uncertainty in the data.