Support Vector Data Description for Uncertainty Data Sets

Jorge Guevara, Stephane Canu, Roberto Hirata Jr
1Department of Computer Science, Institute of Mathematics and Statistics-USP, São Paulo - Brazil.
2LITIS EA 4108, INSA de Rouen, 78680 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 \( \{ \mathbf{x}_i \sim (\xi_i, \Sigma_i) \}_{i=1}^n \) 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

\[
\min_{c \in \mathbb{R}^d, \xi_i \in \mathbb{R}} \quad R^2 + C \sum_{i=1}^n \xi_i \\
\text{subject to} \quad \mathbb{P}(\|\mathbf{x}_i - c\|^2 \leq R^2 + \xi_i) \geq 1 - \kappa_i, \quad i = 1, 2, \ldots, n.
\]

- Interpretation: The probability that the random vector \( \mathbf{x}_i \) takes its values outside the sphere of radius \( R \) and center \( c \) is less or equal than \( \kappa_i \). Example: if \( \kappa_i = 0 \), \( i = 1, 2, 3 \ldots n \) 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

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

Lema

The probabilistic constraint \( \mathbb{P}(\|\mathbf{x}_i - c\|^2 \geq R^2 + \xi_i) \) is bounded by (Markov’s inequality):

\[ tr(\Sigma) + \|\mathbf{c}\|^2 \]

Forcing this bound to be less or equal than a given value \( \kappa_i \) for each instance, i.e.,

\[ \frac{tr(\Sigma) + \|\mathbf{c}\|^2}{R^2 + \xi_i} \leq \kappa_i, \quad i = 1, 2, \ldots, n. \] (1)

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

Definition (USVDD)

Given the training dataset \( \{ \mathbf{x}_i \sim (\xi_i, \Sigma_i) \}_{i=1}^n \), 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

\[
\min_{c \in \mathbb{R}^d, \xi_i \in \mathbb{R}} \quad R^2 + C \sum_{i=1}^n \xi_i \\
\text{subject to} \quad \mathbb{P}(\|\mathbf{x}_i - c\|^2 \leq R^2 + \xi_i - tr(\Sigma_i)) \geq 1 - \kappa_i, \quad i = 1, 2, \ldots, n.
\]

Lagragian and KKT’s Conditions

Lagragian

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

Karush–Kuhn–Tucker (KKT) conditions

\[ \begin{align*}
\frac{\partial L}{\partial R} &= 0 \quad \Rightarrow \quad R = \left( \sum_{i=1}^n \alpha_i \right)^{\frac{1}{2}} \sum_{i=1}^n \alpha_i \mathbf{x}_i = \sum_{i=1}^n \alpha_i \mathbf{x}_i \\
\frac{\partial L}{\partial \xi_i} &= 0 \quad \Rightarrow \quad -2 \mathbf{c}^T \mathbf{c} + 2 \mathbf{c}^T \mathbf{c} + \alpha_i = 0 \\
\frac{\partial L}{\partial \alpha_i} &= 0 \quad \Rightarrow \quad C = \mathbb{P}(\|\mathbf{x}_i - c\|^2 \geq R^2 + \xi_i) = 0, \quad i = 1, 2, \ldots, n
\end{align*} \]

Definition (USVDD Dual Form)

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

Problem

\[ \max_{\alpha \in \mathbb{R}^n} \quad \sum_{i=1}^n \alpha_i \xi_i - \frac{1}{2} \sum_{i=1}^n \alpha_i \sum_{i=1}^n \alpha_i \kappa_i - \sum_{i=1}^n \beta_i \xi_i \]

subject to \( 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, n \)

\[ \alpha_i \xi_i = 0 \quad i \in \{0 < \alpha_i \leq C\} \quad \text{From KKT’s} \]

Analysis

\[ \alpha_i = 0, \beta_i > 0 \Rightarrow \xi_i = 0. \quad \text{Then } \mathbf{x}_i \sim (\hat{\xi}_i, \hat{\Sigma}_i), \ i \in \{0 < \alpha_i \leq C\} \text{ lies inside the hypersphere no matter the value for } \kappa_i. \quad \text{Points } \{ \mathbf{x}_i | \| \mathbf{x}_i - c \|^2 = (\| \mathbf{x}_i - c \|^2 + tr(\Sigma_i))/\kappa_i \} \text{ will be inside the hypersphere.} \]

\[ \alpha_i > 0, \beta_i = 0 \Rightarrow \xi_i > 0. \quad \text{Then } \mathbf{x}_i \sim (\hat{\xi}_i, \hat{\Sigma}_i), \ i \in \{0 < \alpha_i \leq C\} \text{ lies outside the hypersphere with probability } \kappa_i. \quad \text{Points } \{ \mathbf{x}_i | \| \mathbf{x}_i - c \|^2 = (\| \mathbf{x}_i - c \|^2 + tr(\Sigma_i))/\kappa_i \} \text{ will be outside the hypersphere.} \]

\[ \alpha_i > 0, \beta_i > 0 \Rightarrow \xi_i = 0 \text{ and } 0 < \alpha_i \kappa_i \leq C. \quad \text{From this and (2)} \]

\[ \frac{R^2}{\kappa_i} \geq \| \mathbf{x}_i - c \|^2 + tr(\Sigma_i) \quad i \in \{0 < \alpha_i \kappa_i < C\} \]

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

Figures

![Two different USVDD solutions for the same dataset and same C but with different probabilistic levels between both problems. The first one has values \( \kappa_1 = 0.35 \), \( \kappa_2 = 0.30 \) and the second one has values \( \kappa_1 = 0.1, \kappa_2 = 0.2 \).](image)

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

Conclusion

References