The Limit of One-Class SVM

Régis Vert

`Regis.Vert@lri.fr`

Laboratoire de Recherche en Informatique

Université Paris Sud, Orsay, France.

****

Jean-Philippe Vert

`Jean-Philippe.Vert@mines.org`

Computational Biology Group

ENSMP, Fontainebleau, France.

One-Class SVM (Schölkopf&al, 2001)

- Given $X_1, \ldots, X_n \in \mathcal{X}$, $\lambda > 0$, RKHS $\mathcal{H}$,

$$\hat{f} = \arg\min_{f \in \mathcal{H}} \ \lambda \| f \|_{\mathcal{H}}^2 + \frac{1}{n} \sum_{i=1}^{n} (1 - f(X_i))_+$$

- Acceptance region: $\left\{ \hat{f} \geq 1 - \delta \right\}$

Quantile Estimation (QE)

- P distribution on $\mathbb{R}^d$
- α-Quantile ($0 \leq \alpha \leq 1$)

$$Q(\alpha) = \arg \min_{C \in \text{mes}\left(\mathbb{R}^d\right),\ P(C) \geq \alpha} \text{Leb}(C)$$

- Given $X_1, \ldots, X_n$ i.i.d. from unknown P, estimate $Q(\alpha)$

Density Level Set Estimation (DLSE)

- P has density ρ w.r.t. the Lebesgue measure
- Density level set at level μ

$$\Gamma(\mu) = \{\rho \geq \mu\}$$

- Given $X_1, \ldots, X_n$ i.i.d. from P, estimate $\Gamma(\mu)$

QE = DLSE

- If ρ has no flat part, then QE is equivalent to DLSE:

$$\alpha \leftrightarrow \mu, \quad Q(\alpha) = \Gamma(\mu)$$

- Main feature of $Q(\alpha)$: highly concentrated region of P-probability mass

- Applications: unsupervised learning, anomaly detection, cluster analysis, ...

One-Class SVM (Schölkopf&al, 2001)

- Given $X_1, \ldots, X_n \in \mathcal{X}$, $\lambda > 0$, RKHS $\mathcal{H}$,

$$\hat{f} = \arg\min_{f \in \mathcal{H}} \ \lambda \| f \|_{\mathcal{H}}^2 + \frac{1}{n} \sum_{i=1}^{n} \left(1 - f\left(X_i\right)\right)_+$$

- Acceptance region: $\left\{ \hat{f} \geq 1 - \delta \right\}$

One-Class SVM (Schölkopf&al, 2001)

- Given $X_1, \ldots, X_n \in \mathcal{X}$, $\lambda > 0$, RKHS $\mathcal{H}$,

$$\hat{f} = \arg\min_{f \in \mathcal{H}} \;\; \lambda \| f \|_{\mathcal{H}}^2 + \frac{1}{n} \sum_{i=1}^{n} \left(1 - f\left(X_i\right)\right)_+$$

- Acceptance region: $\left\{ \hat{f} \geq 1 - \delta \right\}$

- Links between $\left\{ \hat{f} \geq 1 - \delta \right\}$ and $Q\left(\alpha\right)$? $\Gamma\left(\mu\right)$?

Main Contribution

- fixed $\lambda > 0$
- Gaussian kernel
- bandwidth $\sigma \to 0$ as $n \to \infty$, certain speed
- **One-class SVM is Consistent for DLSE (and QE)**

Plan

- Setup and Main Results
- Global scheme of the proof
- Bounding the Estimation Error
- Bounding the Regularization Error
- Bounding the Approximation Error

Plan

- Setup and Main Results
- Global scheme of the proof
- Bounding the Estimation Error
- Bounding the Regularization Error
- Bounding the Approximation Error

Some Notation

- $X_1, \dots, X_n \sim P$, $X_i \in \mathbb{R}^d$, support $\subset \mathcal{X}$ compact
- $k_\sigma(x, x') = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^d} \exp\left(\frac{-\| x - x' \|^2}{2\sigma^2}\right)$
- $\mathcal{H}_\sigma$ Gaussian RKHS, norm $\| \, . \, \|_{\mathcal{H}_\sigma}$
- $\hat{f}_\sigma$ minimizes $\lambda \| f \|^2_{\mathcal{H}_\sigma} + \frac{1}{n} \sum_{i=1}^n \left(1 - f\left(X_i\right)\right)_+$ over $\mathcal{H}_\sigma$
- $R_\sigma\left(f\right) = \lambda \| f \|^2_{\mathcal{H}_\sigma} + \mathbb{E}_P\left[\left(1 - f\left(X_1\right)\right)_+\right]$

The Big Picture

- Pointwise convergence for fixed f in $\mathcal{H}_{\sigma_1}$

$$\lambda \| f \|^2_{\mathcal{H}_\sigma} \qquad + \frac{1}{n} \sum_{i=1}^{n} \left(1 - f\left(X_i\right)\right)_+$$

$$\sigma \leq \sigma_1, \ \sigma \to 0 \qquad \qquad n \to \infty$$

$$\lambda \| f \|^2_{L_2} \qquad \mathbb{E}_P \left[\left(1 - f\left(X_1\right)\right)_+ \right]$$

The Big Picture

- Pointwise convergence for fixed f in $\mathcal{H}_{\sigma_1}$

$$\lambda \| f \|^2_{\mathcal{H}_\sigma} \qquad + \frac{1}{n}\sum_{i=1}^{n} \left(1 - f\left(X_i\right)\right)_+$$

$$\sigma \leq \sigma_1,\ \sigma \to 0 \qquad\qquad n \to \infty$$

$$\lambda \| f \|^2_{L_2} \qquad \mathbb{E}_P\left[\left(1 - f\left(X_1\right)\right)_+\right]$$

- Asymptotically: $R_0\left(f\right) = \lambda \| f \|^2_{L_2} + \mathbb{E}_P\left[\left(1 - f\left(X_1\right)\right)_+\right]$

The Shape of f_0

- Define f_0 as follows

$$f_0(x) = \begin{cases} \frac{\rho(x)}{2\lambda} & \text{if } \rho(x) \leq 2\lambda, \\ 1 & \text{if } \rho(x) \geq 2\lambda \end{cases}$$

- Then f_0 is a minimizer of R_0 over L_2

- Asymptotically, $\hat{f}_\sigma \longrightarrow f_0$

Smoothness Assumption

- P has a density ρ w.r.t. the Lebesgue measure
- $\| \rho \|_{L_\infty} = M < \infty$
- $\omega(\rho, \delta) = \sup_{0 \leq \| t \| \leq \delta} \| \rho(. + t) - \rho(.) \|_{L_1}$

$$\omega(\rho, \delta) \leq c\delta^\beta$$

for some $0 < \beta \leq 1$

Main Result

for all $x, \epsilon > 0$, choosing $\sigma = \left(\frac{1}{n}\right)^{\frac{2+\beta}{4\beta+(2+\beta)d} - \frac{\epsilon(2+\beta)}{2\beta}}$, the following holds w.p. $\geq 1 - e^{-x}$

$$0 \leq R_0\left(\hat{f}_\sigma\right) - R_0\left(f_0\right) \leq K\left(\left(\frac{1}{n}\right)^{\frac{2\beta}{4\beta+(2+\beta)d} - \epsilon}\right)$$

and

$$\| \hat{f}_\sigma - f_0 \|_{L_2}^2 \leq K\left(\left(\frac{1}{n}\right)^{\frac{2\beta}{4\beta+(2+\beta)d} - \epsilon}\right)$$

Plan

- Setup and Main Results
- Global scheme of the proof
- Bounding the Estimation Error
- Bounding the Regularization Error
- Bounding the Approximation Error

Split

For $0 < \sigma \leq \sigma_1$,

$$\begin{aligned} R_0\left(\hat{f}_\sigma\right) - R_0\left(f_0\right) = & \left[R_0\left(\hat{f}_\sigma\right) - R_\sigma\left(\hat{f}_\sigma\right)\right] \\ & + \left[R_\sigma\left(\hat{f}_\sigma\right) - R_\sigma^*\right] \\ & + \left[R_\sigma^* - R_\sigma\left(k_{\sigma_1} * f_0\right)\right] \\ & + \left[R_\sigma\left(k_{\sigma_1} * f_0\right) - R_0\left(k_{\sigma_1} * f_0\right)\right] \\ & + \left[R_0\left(k_{\sigma_1} * f_0\right) - R_0\left(f_0\right)\right] \end{aligned}$$

Split

For $0 < \sigma \leq \sigma_1$,

$$
\begin{aligned}
R_0\left(\hat{f}_\sigma\right) - R_0\left(f_0\right) = & \left[R_0\left(\hat{f}_\sigma\right) - R_\sigma\left(\hat{f}_\sigma\right)\right] \\
& + \left[R_\sigma\left(\hat{f}_\sigma\right) - R_\sigma^*\right] \\
& + \left[R_\sigma^* - R_\sigma\left(k_{\sigma_1} * f_0\right)\right] \\
& + \left[R_\sigma\left(k_{\sigma_1} * f_0\right) - R_0\left(k_{\sigma_1} * f_0\right)\right] \\
& + \left[R_0\left(k_{\sigma_1} * f_0\right) - R_0\left(f_0\right)\right]
\end{aligned}
$$

Split

For $0 < \sigma \leq \sigma_1$,

$$\begin{aligned} R_0\left(\hat{f}_\sigma\right) - R_0\left(f_0\right) = & \left[R_0\left(\hat{f}_\sigma\right) - R_\sigma\left(\hat{f}_\sigma\right)\right] \\ & + \left[R_\sigma\left(\hat{f}_\sigma\right) - R_\sigma^*\right] \\ & + \left[R_\sigma^* - R_\sigma\left(k_{\sigma_1} * f_0\right)\right] \\ & + \left[R_\sigma\left(k_{\sigma_1} * f_0\right) - R_0\left(k_{\sigma_1} * f_0\right)\right] \\ & + \left[R_0\left(k_{\sigma_1} * f_0\right) - R_0\left(f_0\right)\right] \end{aligned}$$

Split

For $0 < \sigma \leq \sigma_1$,

$$R_0\left(\hat{f}_\sigma\right) - R_0\left(f_0\right) \leq \left[R_\sigma\left(\hat{f}_\sigma\right) - R^*_\sigma\right]$$

$$+ \left[\lambda\left(\| k_{\sigma_1} * f_0 \|^2_{\mathcal{H}_\sigma} - \| k_{\sigma_1} * f_0 \|^2_{L_2}\right)\right]$$

$$+ \left[R_0(k_{\sigma_1} * f_0) - R_0\left(f_0\right)\right]$$

Split

For $0 < \sigma \leq \sigma_1$,

$$R_0\left(\hat{f}_\sigma\right) - R_0\left(f_0\right) \leq \left[R_\sigma\left(\hat{f}_\sigma\right) - R_\sigma^*\right]$$

$$+ \left[\lambda\left(\| k_{\sigma_1} * f_0 \|^2_{\mathcal{H}_\sigma} - \| k_{\sigma_1} * f_0 \|^2_{L_2}\right)\right]$$

$$+ \left[R_0(k_{\sigma_1} * f_0) - R_0\left(f_0\right)\right]$$

estimation error

Split

For $0 < \sigma \leq \sigma_1$,

$$R_0\left(\hat{f}_\sigma\right) - R_0\left(f_0\right) \leq \left[R_\sigma\left(\hat{f}_\sigma\right) - R^*_\sigma\right]$$

$$+ \left[\lambda\left(\| k_{\sigma_1} * f_0 \|^2_{\mathcal{H}_\sigma} - \| k_{\sigma_1} * f_0 \|^2_{L_2}\right)\right]$$

$$+ \left[R_0(k_{\sigma_1} * f_0) - R_0\left(f_0\right)\right]$$

regularization error

Split

For $0 < \sigma \leq \sigma_1$,

$$R_0\left(\hat{f}_\sigma\right) - R_0\left(f_0\right) \leq \left[R_\sigma\left(\hat{f}_\sigma\right) - R_\sigma^*\right]$$

$$+ \left[\lambda\left(\| k_{\sigma_1} * f_0 \|_{\mathcal{H}_\sigma}^2 - \| k_{\sigma_1} * f_0 \|_{L_2}^2\right)\right]$$

$$+ \left[R_0(k_{\sigma_1} * f_0) - R_0\left(f_0\right)\right]$$

approximation error

Plan

- Setup and Main Results
- Global scheme of the proof
- Bounding the Estimation Error
- Bounding the Regularization Error
- Bounding the Approximation Error

Estimation Error

- $\hat{f}_\sigma \in \mathcal{H}_\sigma$ *a priori*

- Actually, $\hat{f}_\sigma \in \left(1/\sqrt{\lambda}\right) \mathcal{B}_\sigma^{\mathcal{X}}$, where $\mathcal{B}_\sigma^{\mathcal{X}}$ is the unit ball of the completed span of

$$\{k_\sigma(x, \bullet) \ : \ x \in \mathcal{X}\}$$

- Covering numbers: for any $0 < p \leq 2$, $\delta > 0, \ \epsilon > 0$,

$$\log \mathcal{N}\left(\mathcal{B}_\sigma^{\mathcal{X}}, \sqrt{\kappa_\sigma}\epsilon, L_2(T)\right) \leq c_{p,\delta,d}\sigma^{(1-p/2)(1+\delta)d}\epsilon^{-p}$$

where $\kappa_\sigma = 1/\left(\sqrt{2\pi}\sigma\right)^d$

Estimation Error

For all $x > 0$, $0 < p \leq 2$, this holds w.p. $\geq 1 - e^{-x}$

$$R_\sigma\left(\hat{f}_\sigma\right) - R_\sigma^* \leq K_1 \left(\frac{1}{\sigma}\right)^{\frac{[2+(2-p)(1+\delta)]d}{2+p}} \left(\frac{1}{n}\right)^{\frac{2}{2+p}}$$

$$+ K_2 \left(\frac{1}{\sigma}\right)^d \frac{x}{n}$$

Here we are

$$R_0\left(\hat{f}_\sigma\right) - R_0\left(f_0\right) \leq K_1 \left(\frac{1}{\sigma}\right)^{\frac{[2+(2-p)(1+\delta)]d}{2+p}} \left(\frac{1}{n}\right)^{\frac{2}{2+p}}$$

$$+K_2 \left(\frac{1}{\sigma}\right)^d \frac{x}{n}$$

$+$ regularization error

$+$ approximation error

Plan

- Setup and Main Results
- Global scheme of the proof
- Bounding the Estimation Error
- Bounding the Regularization Error
- Bounding the Approximation Error

Regularization Error

- Gaussian RKHS

$$\mathcal{H}_\sigma = \left\{ f \in \mathcal{C}_0\left(\mathbb{R}^d\right) \cap L_1\left(\mathbb{R}^d\right) \; : \; \int_{\mathbb{R}^d} \left|\mathcal{F}\left[f\right](\omega)\right|^2 e^{\frac{\sigma^2 \| \omega \|^2}{2}} d\omega < \infty \right\}$$

- RKHS norm

$$\| f \|^2_{\mathcal{H}_\sigma} = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left|\mathcal{F}\left[f\right](\omega)\right|^2 e^{\frac{\sigma^2 \| \omega \|^2}{2}} d\omega$$

Regularization Error

- For any $0 < \sigma \leq \tau$,

$$\mathcal{H}_\tau \subset \mathcal{H}_\sigma \subset L_2(\mathbb{R}^d)$$

- For any $0 < \sigma \leq \tau$, $f \in \mathcal{H}_\tau$,

$$0 \leq \| f \|^2_{\mathcal{H}_\sigma} - \| f \|^2_{L_2} \leq \frac{\sigma^2}{\tau^2} \left(\| f \|^2_{\mathcal{H}_\tau} - \| f \|^2_{L_2} \right)$$

Regularization Error

- For $0 < \sigma \leq \sigma_1$,

$$\| k_{\sigma_1} * f_0 \|^2_{\mathcal{H}_\sigma} - \| k_{\sigma_1} * f_0 \|^2_{L_2} \leq \frac{\sigma^2}{2\sigma_1^2} \| k_{\sigma_1} * f_0 \|^2_{\mathcal{H}_{\sqrt{2}\sigma_1}}$$

- $\| k_{\sigma_1} * f_0 \|^2_{\mathcal{H}_{\sqrt{2}\sigma_1}} = \| f_0 \|^2_{L_2}$

- Eventually,

$$\| k_{\sigma_1} * f_0 \|^2_{\mathcal{H}_\sigma} - \| k_{\sigma_1} * f_0 \|^2_{L_2} \leq \frac{\sigma^2}{2\sigma_1^2} \| f_0 \|^2_{L_2}$$

Here we are

$$R_0\left(\hat{f}_\sigma\right) - R_0\left(f_0\right) \leq K_1 \left(\frac{1}{\sigma}\right)^{\frac{[2+(2-p)(1+\delta)]d}{2+p}} \left(\frac{1}{n}\right)^{\frac{2}{2+p}}$$

$$+ K_2 \left(\frac{1}{\sigma}\right)^d \frac{x}{n}$$

$$+ \frac{\sigma^2}{2\sigma_1^2} \| f_0 \|_{L_2}^2$$

$+$ approximation error

Plan

- Setup and Main Results
- Global scheme of the proof
- Bounding the Estimation Error
- Bounding the Regularization Error
- Bounding the Approximation Error

Approximation Error

- $R_0(k_{\sigma_1} * f_0) - R_0\,(f_0) \leq (2\lambda \| \, f_0 \, \|_{L_\infty} + M) \, \| \, k_{\sigma_1} * f_0 - f_0 \, \|_{L_1}$

- $\| \, k_{\sigma_1} * f_0 - f_0 \, \|_{L_1} \leq \left(1 + \sqrt{d}\right) \omega\,(f_0, \sigma_1)$

- $\omega\,(f_0, \sigma_1) \leq c\sigma_1^\beta$

Here we are

$$R_0\left(\hat{f}_\sigma\right) - R_0\left(f_0\right) \leq K_1\left(\frac{1}{\sigma}\right)^{\frac{[2+(2-p)(1+\delta)]d}{2+p}}\left(\frac{1}{n}\right)^{\frac{2}{2+p}}$$

$$+ K_2\left(\frac{1}{\sigma}\right)^d \frac{x}{n}$$

$$+ \frac{\sigma^2}{2\sigma_1^2}\| f_0 \|_{L_2}^2$$

$$+ \left(2\lambda\| f_0 \|_{L_\infty} + M\right)\left(1+\sqrt{d}\right) c\sigma_1^\beta$$

conclusion

- Regularization with the Gaussian kernel (parameter σ)
- One-class SVM = kernel density estimator in low density regions
- Our analysis is valid for general convex loss functions, and applies also to the case of binary classification