LEARNING MINIMUM VOLUME SETS WITH SUPPORT VECTOR MACHINES

Mark A. Davenport, Richard G. Baraniuk

Rice University
Department of Electrical and Computer Engineering

Clayton D. Scott

University of Michigan
Department of Electrical Engineering and Computer Science

Overview

Use support vector machines to estimate *minimum volume sets (MV-sets)*

- anomaly detection
- clustering

Key idea: reduce MV-set estimation to *Neyman-Pearson classification*

- treat MV-set estimation (*one-class* problem) as a *two-class* problem like classification
- draw second class from *uniform* distribution

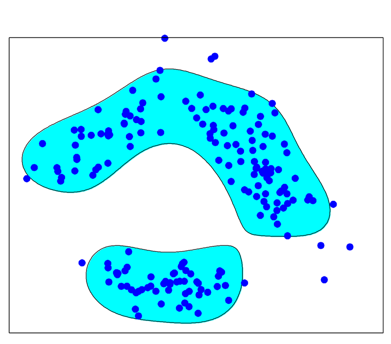
Minimum Volume Sets

Given

- Probability measure P
- Reference measure μ (typically Lebesgue)
- Target mass β

The *minimum volume set* is

$$G^*_\beta = \arg\min\{\mu(G) : P(G) \geq \beta, G \text{ measurable}\}$$



Neyman-Pearson Classification

Given

- Probability measures Q_+ and Q_-
- Target power α

Let
$$P_F(f) = Q_-(\{x : f(x) = +1\})$$
$$P_M(f) = Q_+(\{x : f(x) = -1\})$$

The *Neyman-Pearson classifier* is

$$f^*_\alpha = \arg\min\{P_M(f) : P_F(f) \leq \alpha\}$$

Reduction to Neyman-Pearson Classification

Any technique for estimating an NP classifier can be adapted to estimate an MV-set

Set
$$Q_- = 1 - P$$
$$Q_+ = \mu$$
$$\alpha = 1 - \beta$$

Then, if f^*_α is the optimal NP classifier,

$$G^*_\beta = \{x : f^*_\alpha = -1\}$$

Challenge: we only have samples from P

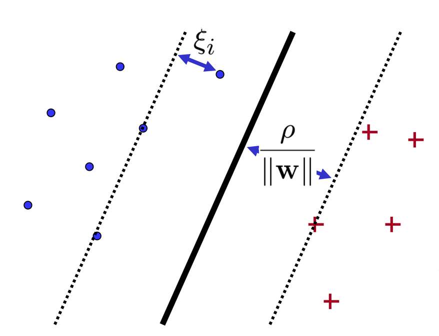
→ we can *sample* from μ

Support Vector Machines

Method for learning classifiers from training data

- Use "kernel-trick"
- Maximize the "margin"

$$\min_{\mathbf{w},b,\xi,\rho} \quad \tfrac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{n}\sum_{i=1}^{n}\xi_i \qquad \nu \in [0,1]$$
$$\text{s.t.} \quad (\langle \mathbf{w}, \mathbf{x_i}\rangle + b)y_i \geq \rho - \xi_i$$



Neyman-Pearson SVMs

Consider *cost-sensitive* SVM

- Introduce class-specific weights
- Adjust weights to achieve desired error rates

Relies on accurate error estimation

- cross-validation

$$\min_{\mathbf{w},b,\xi,\rho} \quad \tfrac{1}{2}\|\mathbf{w}\|^2 - 2\nu_-\nu_+\rho + \frac{\nu_-}{n_+}\sum_{i\in I_+}\xi_i + \frac{\nu_+}{n_-}\sum_{i\in I_-}\xi_i$$
$$\text{s.t.} \quad (\langle \mathbf{w}, \mathbf{X_i}\rangle + b)Y_i \geq \rho - \xi_i$$

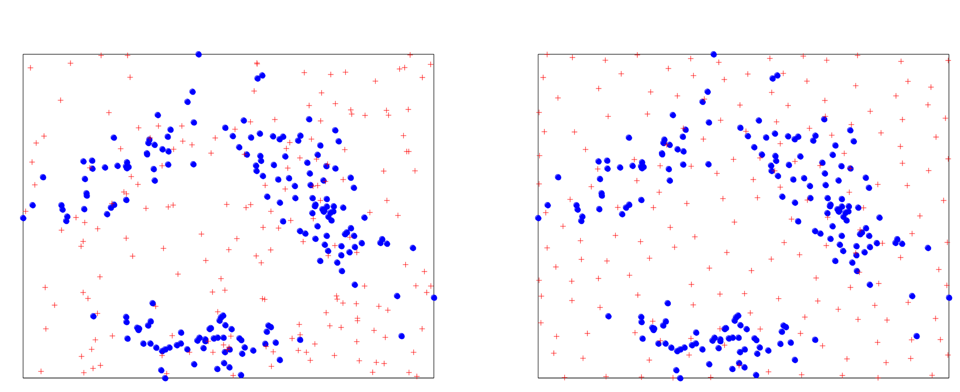
$$(\nu_+, \nu_-) \in [0,1]^2$$

Uniform Data: Thinning

In high dimensions we must confront the "curse of dimensionality"

One option is *thinning* the data to ensure a large distance between any pair of points

- results in an approximate "packing set"

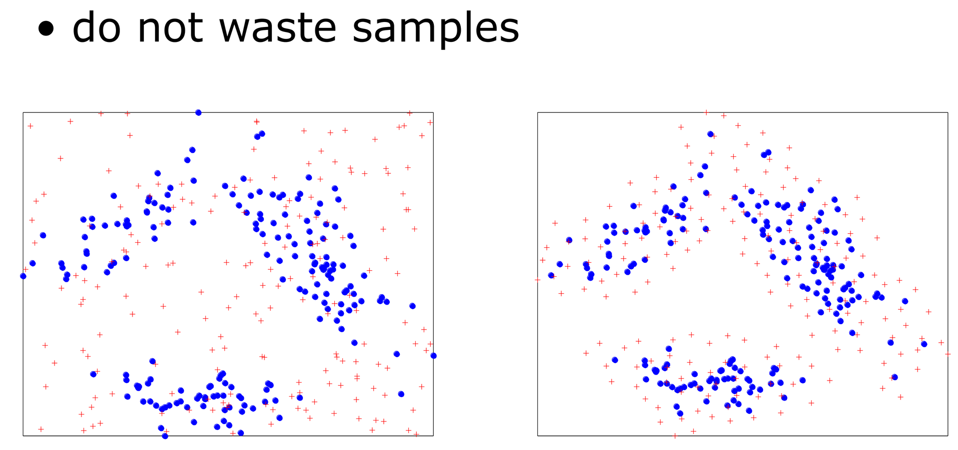


Uniform Data: Manifold Sampling

Thinning does not directly overcome the "vastness of space" in high dimensions

What if our data lies on a *manifold*?

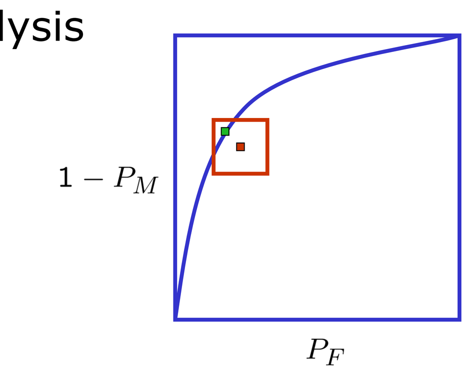
- adapt to this structure
- do not waste samples



Measuring Performance

Algorithms for MV/level set estimation of NP classification are typically analyzed using

ROC analysis



We want to operate at a *specific point* of the ROC curve

$$\mathcal{E}(G) := \frac{1}{1-\beta}\max\{\beta - P(G), 0\} + \mu(G)$$

Results: MV-set Estimation

Compare with one-class SVM

Modified LIBSVM software

Highlights:

- manifold sampling performs best
- two-class methods more reliable
- impact of discrete data

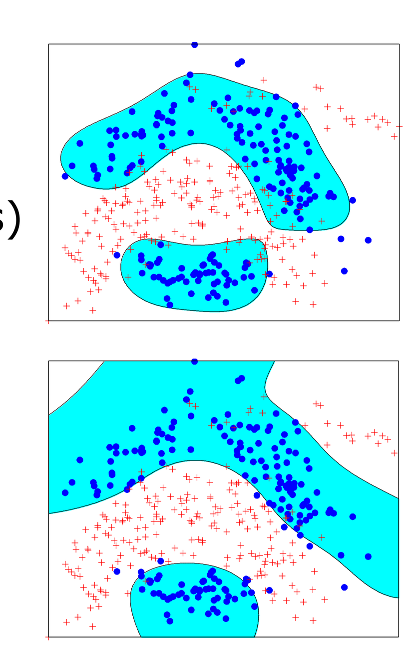
		$\mathcal{E}_\mu(G)$
banana	OC-SVM	1.36
	NP-IND	0.53
	NP-THIN	0.47
	NP-MAN	**0.44**
breast-cancer	OC-SVM	0.55
	NP-IND	0.29
	NP-THIN	1.75
	NP-MAN	**0.06**
heart	OC-SVM	0.63
	NP-IND	0.43
	NP-THIN	1.26
	NP-MAN	**0.16**
thyroid	OC-SVM	0.77
	NP-IND	**0.63**
	NP-THIN	0.79
	NP-MAN	0.7
ringnorm	OC-SVM	0.11
	NP-IND	0.17
	NP-THIN	0.11
	NP-MAN	**0.06**

Results: Anomaly Detection

Test validity of uniform prior

Compare

- MV-set (one class)
- NP-classifier (both classes)



		$\mathcal{E}_+(G)$
banana	without	0.29
	with	**0.24**
breast-cancer	**without**	**0.83**
	with	0.99
heart	without	0.76
	with	**0.50**
thyroid	without	0.44
	with	**0.22**
ringnorm	**without**	**0.015**
	with	0.021

Conclusions

Minimum volume sets are an effective way to approach anomaly detection

We can accurately estimate minimum volume sets using Neyman-Pearson SVMs

The procedure used for generating "uniform" samples can significantly impact performance

Our approach tends to perform

- better than the one-class SVM
- often nearly as well the NP classifier trained using *both* classes

dsp.rice.edu