Enhancing One-class Support Vector Machines for Unsupervised Anomaly Detection

Mennatallah Amer[1], **Markus Goldstein[2]**, Slim Abdennadher[1]

[1] German University in Cairo, Egypt
[2] German Research Center for Artificial Intelligence (DFKI)
http://www.dfki.de/

August 11th, 2013

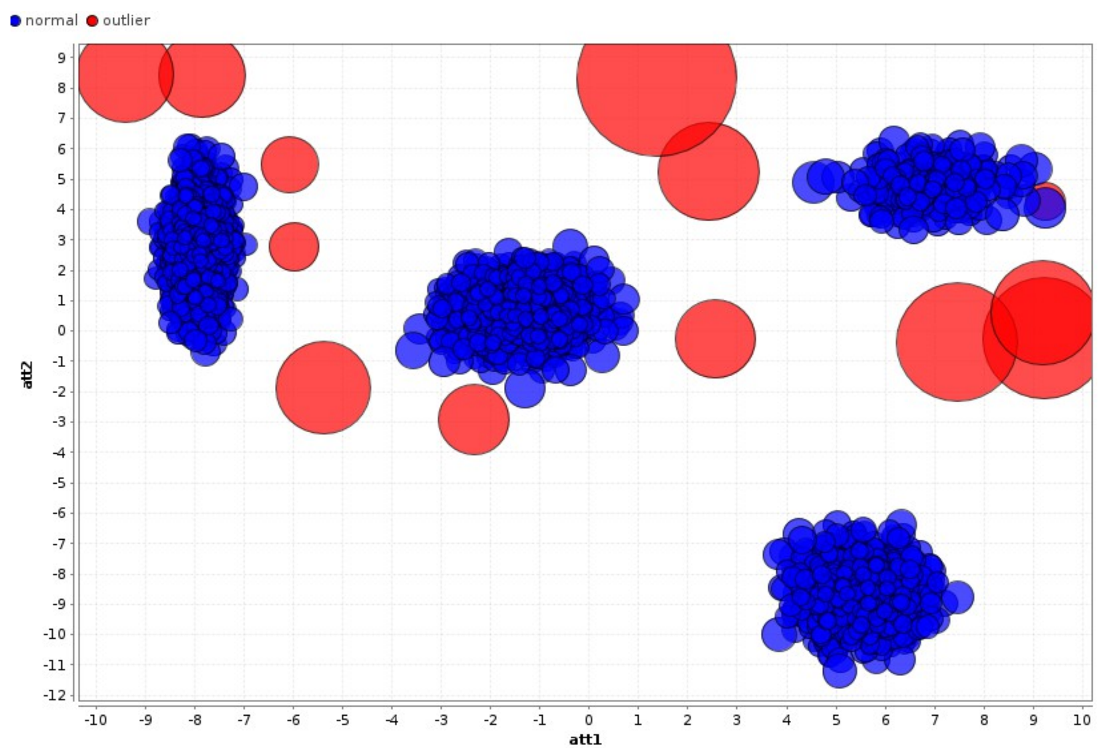
Outline

- Introduction
- Enhanced one-class SVMs
  - Robust one-class SVM
  - Eta one-class SVM
- Experiments
- Conclusion

An outlying observation, or **outlier**, is one that appears to deviate markedly from other members of the sample in which it occurs.

(Grubbs,1969)

- Unsupervised anomaly detection



- Support Vector Machines
  - Supervised
  - Learns a decision boundary by maximizing the margin
  - Applies the kernel trick for non-linear decision boundaries

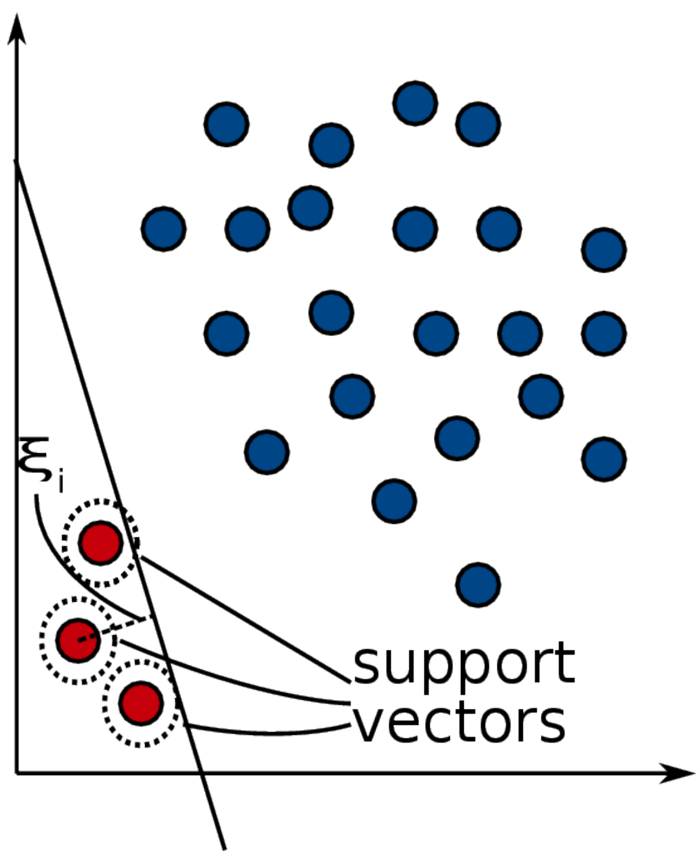
Image has been removed due to missing copyright for re-publishing. Please see: http://www.cac.science.ru.nl/people/ustun/SVM.JPG

- One-class SVMs[1]
  - Semi-supervised (trained with normal class only)
  - Learns a decision function $f(x_i) = w^T \phi(x_i) - \rho$
  - Applies the kernel and separates data from the origin

Image has been removed due to missing copyright for re-publishing.
Please see:
https://www.ntt-review.jp/archive_html/200711/images/sf2_fig02.gif

[1]B Schölkopf, J C Platt, J Shawe-Taylor, a J Smola, and R C Williamson. Estimating the support of a high-dimensional distribution. Neural computation, 13(7):1443–71, July 2001

- One-class SVMs
  - Outliers are allowed by a slack variable ξ (soft margin)
  - Outliers still may contribute to the decision boundary



$$\min_{w,\xi,\rho} \frac{\|w\|^2}{2} + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho$$

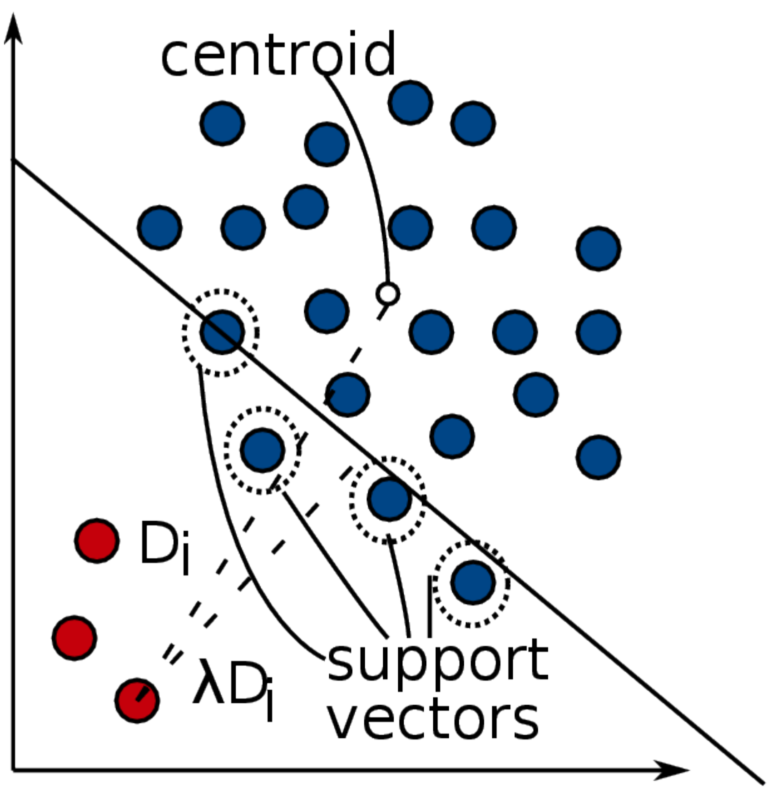
$$\text{s.t. } w^T\phi(x_i) \geq \rho - \xi_i,\ \xi_i \geq 0$$

Enhanced one-class SVMs

- Motivation
  - For unsupervised anomaly detection, is there a good way to cope with outliers?
- In supervised SVMs, there exist approaches:
  - Robust SVMs dealing with noise in the data
  - Identify and remove outliers during training
- Idea: Use these approaches for unsupervised anomaly detection

Enhanced one-class SVMs

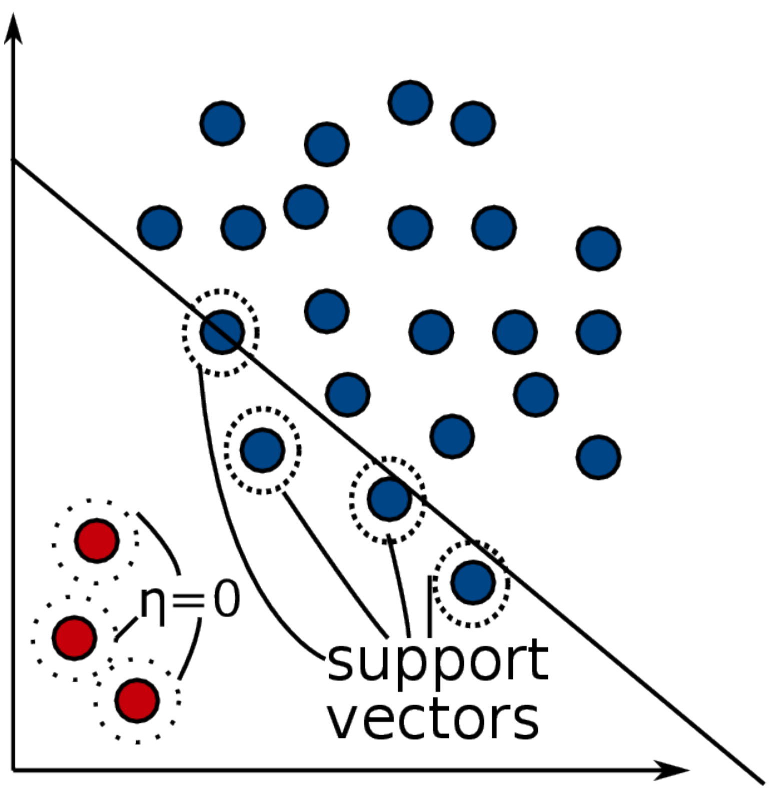
- Robust[2] one-class SVMs
  - Slack variable proportional to the distance to the centroid



$$\min_{w,\rho} \frac{\|w\|^2}{2} - \rho$$
$$\text{s.t. } w^T \phi(x_i) \geq \rho - \lambda * \bar{D}_i$$

[2]Qing Song, Wenjie Hu, and Wenfang Xie. Robust support vector machine with bullet hole image classification. IEEE Transactions on Systems, Man and Cybernetics, 32(4):440–448, November 2002

- Eta[3] one-class SVMs
  - Eta represents the estimate if a point is normal or not
  - Beta is the expected percentage of normal data



$$\min_{w,\xi,\rho} \min_{\eta \in \{0,1\}} \frac{\|w\|^2}{2} + \eta^T \xi - \rho$$

$$\text{s.t. } e^T \eta \geq \beta n$$

$$\xi_i \geq 0$$

$$\xi_i \geq (\rho - w^T \phi(x_i))$$

[3] Linli Xu, K Crammer, and Dale Schuurmans. Robust support vector machine training via convex outlier ablation. Proceedings of the National Conference On Artificial Intelligence, pages 536–542, 2006

- Outlier Score
  - Typically, a one-class SVM outputs class labels “normal”/”outlier”
  - In unsupervised anomaly detection, scores are preferred
  - The distance to the decision boundary is used as a score

$$f(x) = \frac{g_{max} - g(x)}{g_{max}}$$

  - Scores >1.0 indicate outliers

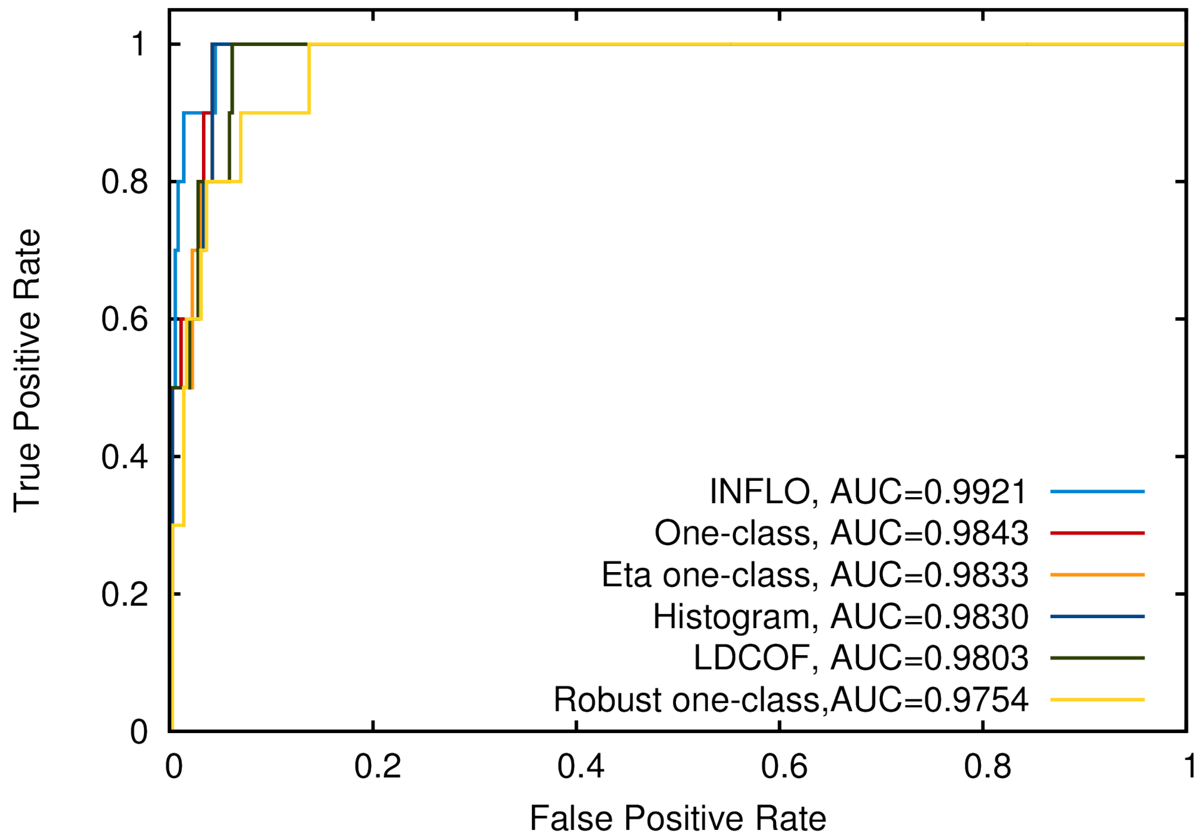
Experiments

Evaluation using UCI datasets

- Ionosphere (233 instances, 26 dim, 3.4% outlier)
- Breast-cancer (569 instances, 30 dim, 2.72% outlier)
- Satellite (6435 instances, 36 dim, 1.94% outlier)
- Shuttle (58000 instances, 9 dim, 1.89% outlier)

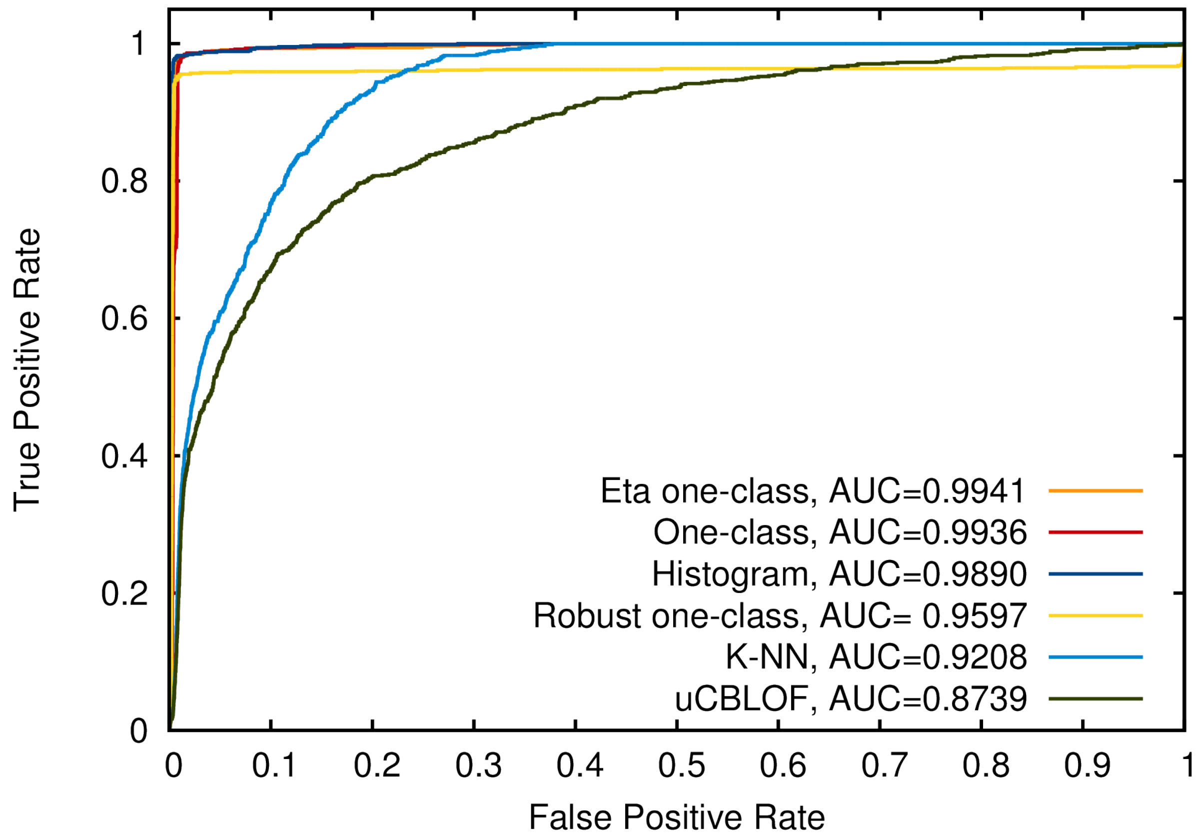
- ROC computation by varying the outlier threshold

- Results for breast-cancer



Algorithm	nSV	time[ms]
One-class	144	48.72 ± 1.01
Robust one-class	90	57.27 ± 2.29
Eta one-class	48	82.46 ± 0.42

- Results for shuttle



Algorithm	nSV	time[s]
One-class	21374	747.15 ± 10.94
Robust one-class	5	218.93 ± 3.17
Eta one-class	8	4.07 ± 0.14

Summary of AUC results

Dataset	One-class	Robust one-class	Eta one-class	k-NN	LOF	COF	INFLO	LoOP	Histogram	CBLOF	u-CBLOF	LDCOF
ionosphere	0.9878	0.9956	**0.9972**	0.9933	0.9178	0.9406	0.9406	0.9211	0.7489	0.3183	0.9822	0.9306
shuttle	0.9936	0.9597	**0.9941**	0.9208	0.6072	0.5612	0.5303	0.5655	0.9889	0.8700	0.8739	0.5312
breast-cancer	0.9843	0.9734	0.9833	0.9826	0.9916	0.9888	**0.9922**	0.9882	0.9829	0.8389	0.9743	0.9804
satellite	0.8602	0.8861	0.8544	**0.9003**	0.8964	0.8708	0.8592	0.8664	0.8862	0.4105	0.9002	0.8657

Conclusion

Key findings

- Eta one-class SVM seems most promising (among SVMs)
- One-class SVM approaches outperform clustering based methods
- SVMs not very good in detecting local outliers
- Can be faster than nearest-neighbor approaches

Implementation: RapidMiner Anomaly Detection Extension
http://madm.dfki.de/rapidminer/anomalydetection

Thank you for your attention!

Questions?
Demo!