Reproducing Kernel PCA for Novelty Detection [1]

Nicholas Merrill

November 22, 2019

The goal of an anomaly (outlier or novelty) detection method is to detect anomalous points within a data set dominated by the presence of ordinary background points. Anomalies are by definition rare and are often generated by different underlying processes [2,3]. Numerous algorithms have been devised toward this goal, the results of which have been applied to a variety of fields to improve upon domain-specific, rule-based detection methods. Anomaly detection has applications in medicine, fraud detection, fault detection, and remote sensing, among others [1, 2].

Background data is often produced by non-linear processes [4]. Kernel-based learning methods are motivated by the idea that there exists a better model of the data in a transformed, non-linear, feature space ($\mathcal{F}$). A kernel function allows the efficient computation of inner products in $\mathcal{F}$ without the explicit calculation of the mapping. The most popular amongst these methods for anomaly detection is the One-Class Support Vector Machine (OC-SVM), which separates the data from the origin in $\mathcal{F}$ [5]. For a Gaussian radial-basis-function (rbf) kernel, this process is equivalent to spherically enclosing the data in $\mathcal{F}$.

Hoffman claims that because samples are treated independently a OC-SVM produces a boundary that is too large to tightly model the background data, causing false positives [1].Hoffman draws from the benefits of kernel techniques and the potential limitations of SVMs, by using kernel PCA (kPCA) to better model the relationship between background points [4, 6, 7]. The separation of points in $\mathcal{F}$ and the background model serves as an anomaly score. In an Hoffman's evaluation on a number of real-world and toy data sets, kPCA demonstrated better generalization, accuracy, and robustness over linear PCA, the Parzen density estimator, and OC-SVMs [1].

1. I plan to reproduce Hoffman's evaluations on the *Cancer* and *Digit 0* data sets. In addition I will extend the evaluation to two additional toy data sets that I will generate and two more real anomaly data sets, *ionosphere* and *glass*, retrieved from [8]

2. I retrieve Hoffman's original MATLAB code from [9] and port it to Python. I will write my own function for the Parzen Density Window and use PYOD's implementation of OC-SVM [10].

3. I will use a k-folds cross-validation procedure, for selecting parameters. I will generate Receiver operating characteristic (ROC) curves to quantify the detection rates and report the Area Under the Curve (AUC) for each method and data set at the optimum parameters.

In addition to Hoffman's work I will read a recent review of multivariate anomaly detection [2] and [11] which details the OC-SVM Hoffman used for a comparison. I will be working on my own.

References

[1] H. Hoffmann, "Kernel pca for novelty detection," *Pattern Recognition*, vol. 40, no. 3, pp. 863 – 874, 2007.

[2] M. Goldstein and S. Uchida, “A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data,” *PLoS ONE*, Apr 2016.

[3] G. Enderlein, “Hawkins, d. m.: Identification of outliers. chapman and hall, london – new york 1980, 188 s., £ 14, 50,” *Biometrical Journal*, vol. 29, no. 2, pp. 198–198, 1987.

[4] C. C. Olson, M. Coyle, and T. Doster, “A study of anomaly detection performance as a function of relative spectral abundances for graph- and statistics-based detection algorithms,” in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIII* (M. Velez-Reyes and D. W. Messinger, eds.), vol. 10198, pp. 309 – 320, International Society for Optics and Photonics, SPIE, 2017.

[5] D. M. Tax and R. P. Duin, “Support vector domain description,” *Pattern Recognition Letters*, vol. 20, no. 11, pp. 1191 – 1199, 1999.

[6] B. Shen, B.-D. Liu, Q. Wang, Y. Fang, and J. P. Allebach, “Sp-svm: Large margin classifier for data on multiple manifolds,” in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI’15, pp. 2965–2971, AAAI Press, 2015.

[7] B. Schölkopf, A. Smola, and K. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Computation*, vol. 10, pp. 1299–1319, July 1998.

[8] S. Rayana, “Outlier detection datasets: ODDS,” 2016.

[9] http://www.heikohoffmann.de/kpca.html.

[10] Y. Zhao, Z. Nasrullah, and Z. Li, “Pyod: A python toolbox for scalable outlier detection,” *Journal of Machine Learning Research*, vol. 20, no. 96, pp. 1–7, 2019.

[11] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, “Support vector method for novelty detection,” in *Proceedings of the 12th International Conference on Neural Information Processing Systems*, NIPS’99, (Cambridge, MA, USA), pp. 582–588, MIT Press, 1999.