Centre for
Data Analytics

Insight

Elliptical Basis Function Data Descriptor (EBFDD) Networks: An Anomaly Detection Approach

Authors:

Mehran H. Bazargani

Brian Mac Namee

DCU NUI Galway OÉ Gaillimh UCC UCD Dublin sfi

Introduction

- “Anomaly detection is the problem of finding patterns in data that do not conform to expected behavior.“[1]

- We are interested in Streaming Data:
  - Where new data keeps coming in for judgement
  - Where training uses ***ONLY*** the normal data

- Examples of Streaming Data:
  - Incoming Data Packets in a network
  - Incoming video frames received from a CCTV camera

[1] V. Chandola and A. Banerjee and V. Kumar, “Anomaly Detection: A survey”, ACM Computing Surveys (CSUR), 2009

Motivation

- Our Approach: One-Class Classification based on ***Radial Basis Function (RBF) networks***

- RBF networks have been around for a long time and can be used for ***classification*** and ***regression*** → ***Supervised Learning!***

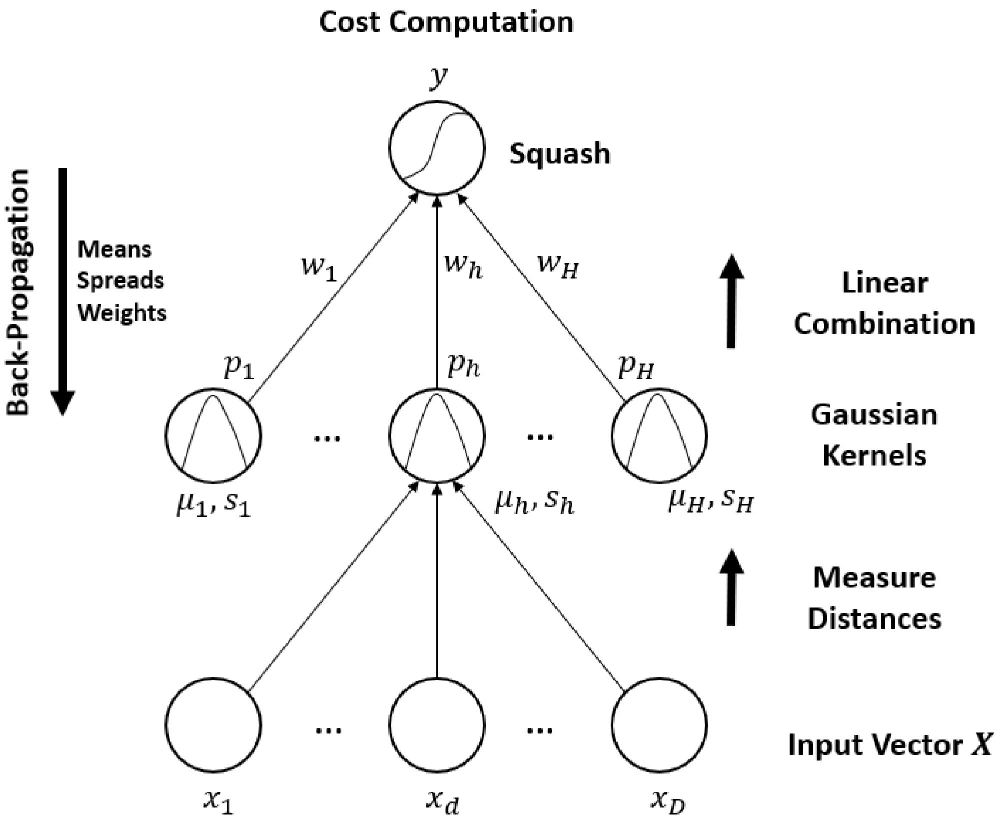
- RBF networks have been used for time-series classification

- Characteristics of RBF networks are:

    - RBF networks are interpretable!

    - RBF networks lend themselves to handling concept drift/shift!

The RBF Network



- A local representation learning technique
- Each Kernel is responsible for a part of the input space
- The most commonly used activation function is ***Gaussian***:

$$P_h = \exp[\frac{-\|X - \mu_h\|^2}{2s_h^2}]$$

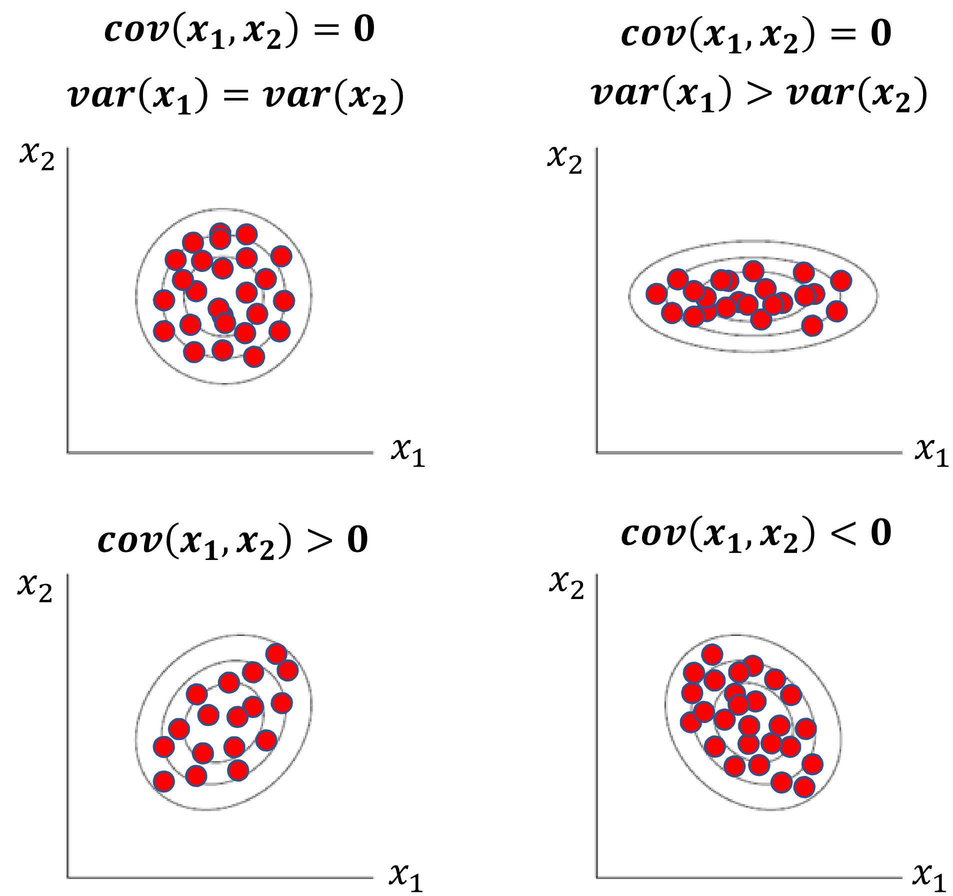
The Elliptical Basis Function Data Descriptor (EBFDD) network

Transforming the RBF Network to a One-Class Classifier

- Classical RBF networks cannot be applied to Anomaly Detection
- We need to cover the normal region with the tightest set of Gaussians
    - Solution: Regularize the size of these Gaussians
- We modify RBF networks in 2 ways:
    - We introduce elliptical Kernels
    - We propose a novel cost function

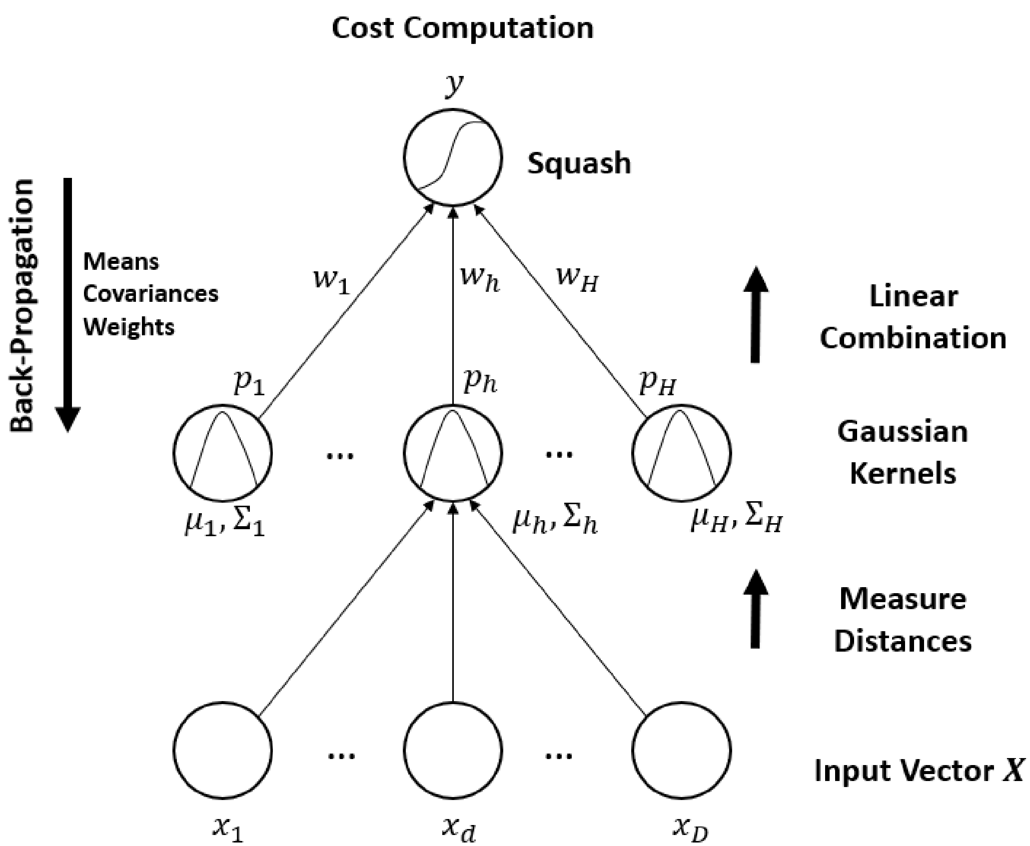
EBFDD in More Depth …

- What is nice about the EBFDD network?
  - Non-diagonal covariance matrices
  - We can have elliptical kernels which can ***rotate*** and ***elongate***, at will
  - This can help for better coverage and a better understanding of the distribution of the normal data



$$p_h(X) = exp\left[-\frac{1}{2}(X-\mu_h)^T {\Sigma_h}^{-1}(X-\mu_h)\right]$$

The EBFDD Network



- The cost function to be **minimized**:

$$E = \frac{1}{2}\left[(1-y)^2 + \beta R_\Sigma + \lambda R_W\right]$$

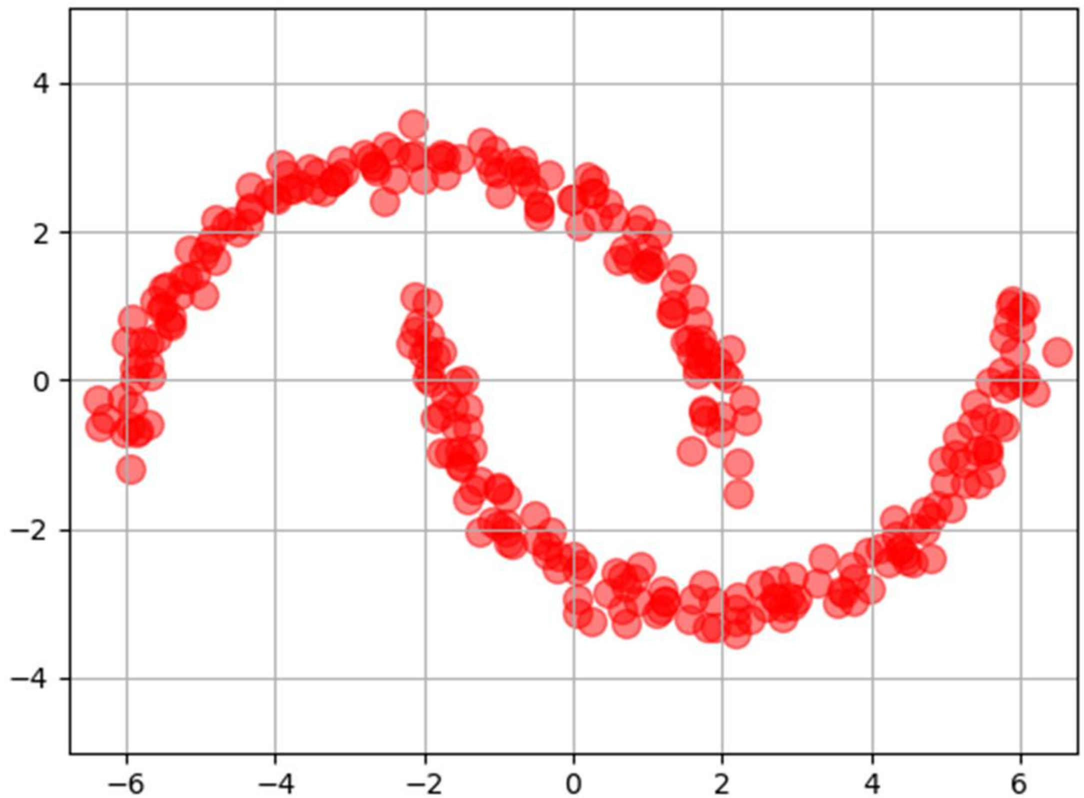
- Where:

$$R_\Sigma = \sum_{h=1}^{H}\sum_{d=1}^{D}\left(\Sigma_h\left[d,d\right]\right)^2$$

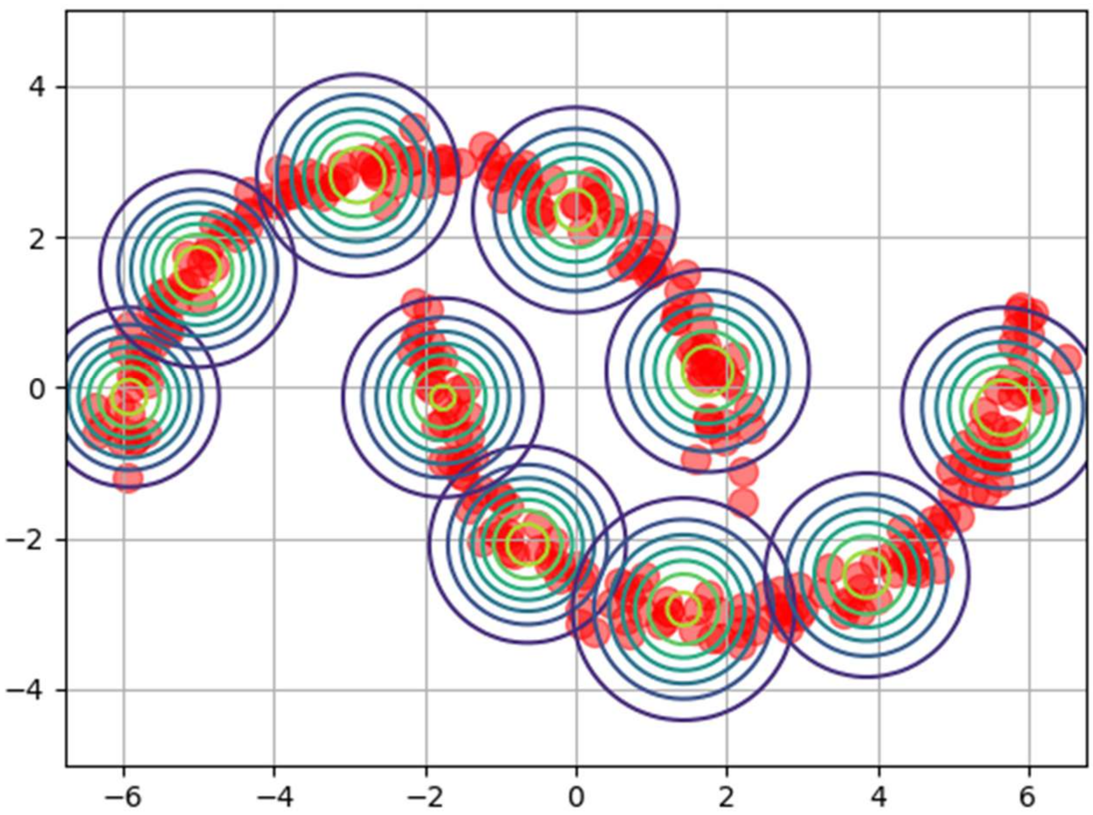
$$R_W = \sum_{h=1}^{H} w_h^2$$

A Visual Representation of the EBFDD Network

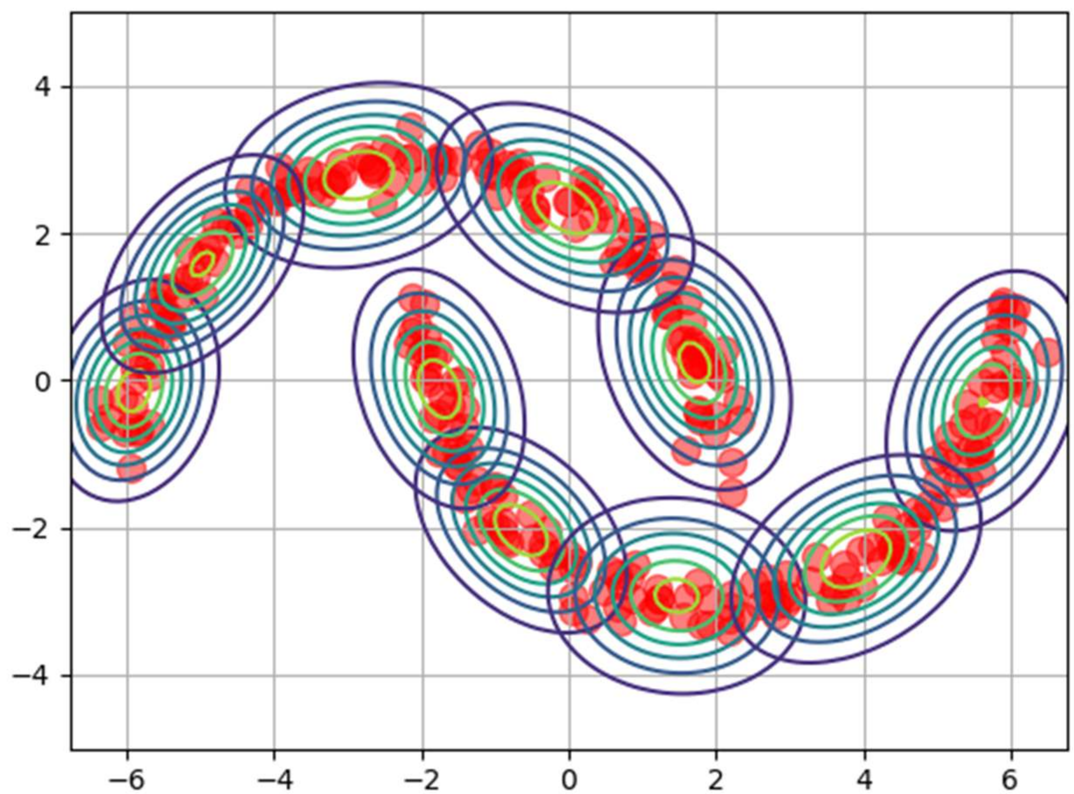
2-Dimensional Normal Data



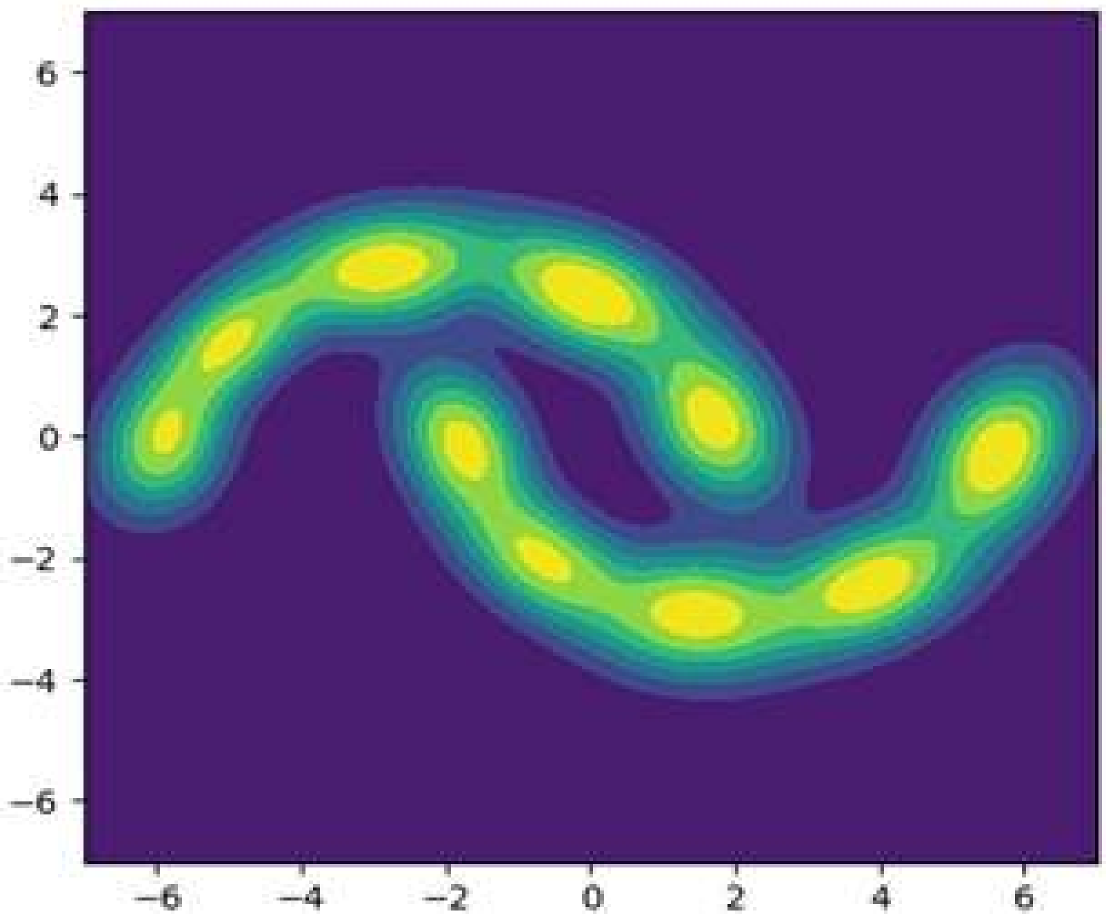
After K-means has Converged



After EBFDD has Converged



The Decision Boundary of EBFDD



The Datasets
and
Experiment Setup

The Datasets We Have Used

Dataset Name	Number of Rows	Number of Classes	Number of Features	Number of Generated Scenarios
Magic Gamma Telesacope	19021	2	10	1
Spambase	4602	2	57	1
Skin Segmentation	245058	2	3	1
Steel Plates Faults	1941	7	27	15
Image Segmentation	2311	7	18	15
Page Blocks Classification	5473	5	10	11
Statlog (Landsat Satellite)	6436	6	36	13
Waveform Database Generator (Version 1)	5000	3	21	7

Emmott, A.F., Das, S., Dietterich, T., Fern, A., Wong, W.K.: Systematic construction of anomaly detection benchmarks from real data. In: Proceedings of the ACM SIGKDD workshop on outlier detection and description. pp. 16–21. ACM (2013)

Experiment Scenarios

- We have followed 3 approaches for generating anomaly detection scenarios from labelled datasets, each with ***N classes***

1. One vs. All: Each of N classes is considered as Normal and everything else as anomalous → ***N Experiments***
2. All vs. One: Each of N classes is considered as Anomalous and everything else as normal → ***N Experiments***
3. Difficult Scenarios: Determine the 2 most difficult separable subsets, and consider one as Normal and the other as Anomalous → ***1 Experiment***

Emmott, A.F., Das, S., Dietterich, T., Fern, A., Wong, W.K.: Systematic construction of anomaly detection benchmarks from real data. In: Proceedings of the ACM SIGKDD workshop on outlier detection and description. pp. 16–21. ACM (2013)

Experiment Description

- Algorithms: EBFDD, RBFDD, Gaussian Mixture Model, Auto-Encoder, Isolation Forest, and One-Class SVM

- For every algorithm and scenario:
  - For every combination of hyper-parameters:
    - For **10** rounds:
      - ***Training Data*** *= Sample (no replacement) 80% of ALL Normal*
      - ***Test Data*** *= Remaining 20% of Normal + All Anomalous*
      - *Compute the AUC of the ROC curve*
    - Average the AUC's across all **10** rounds
- Report the best averaged AUC for each algorithm with its corresponding winning hyper-parameters

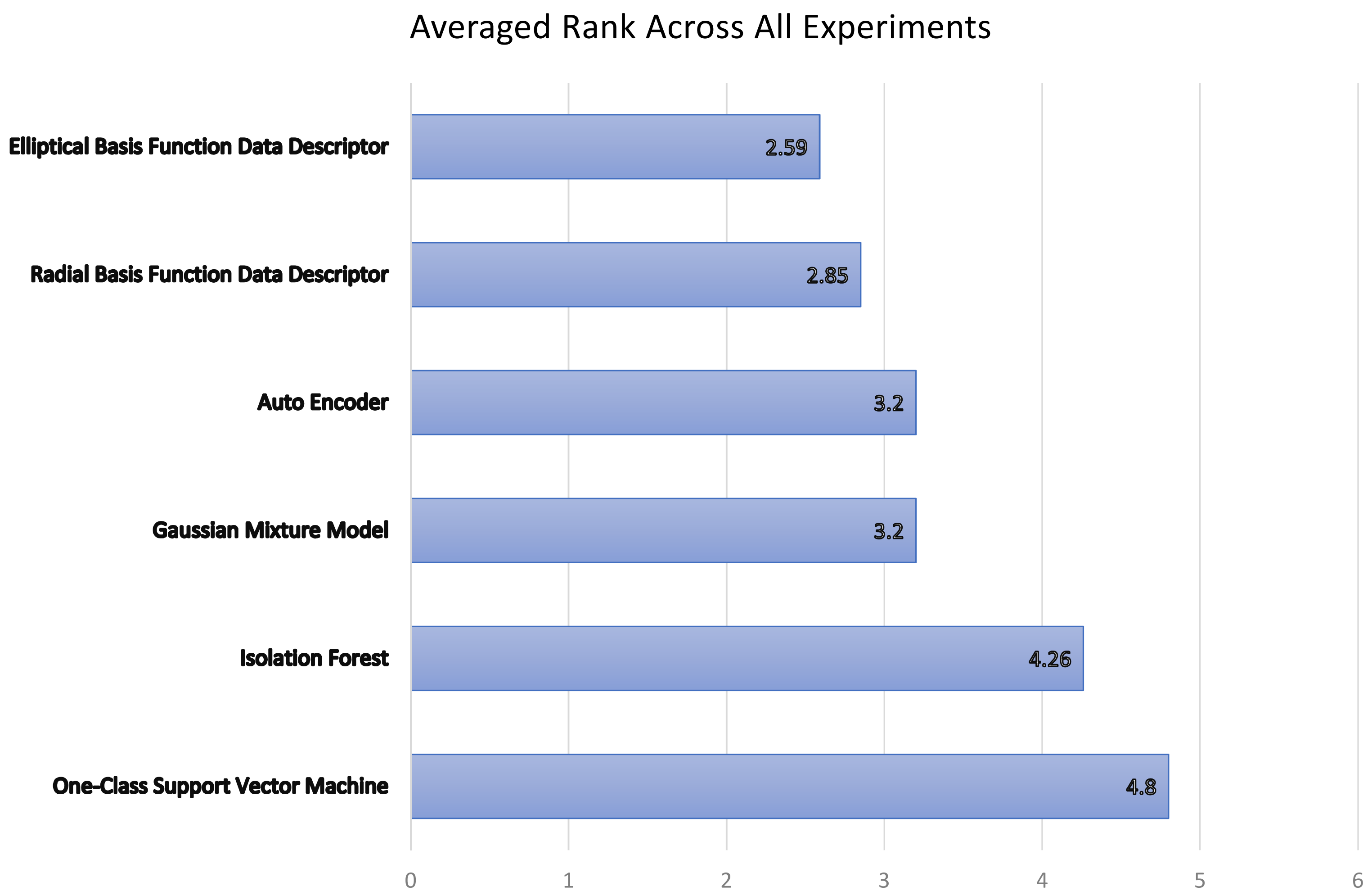
The Experiment Results

Results using AUC's and Ranks

Multi-Class Datasets: One vs All Scenarios

	Class ID	RBFDD	OCSVM	EBFDD	AEN	GMM	iForest
Fault	1	0.87388 (3)	0.85343 (6)	0.87555 (2)	0.86261 (5)	0.88637 (1)	0.87033 (4)
	2	0.97667 (2)	0.97224 (4)	0.97772 (1)	0.96593 (5)	0.97570 (3)	0.92684 (6)
	3	0.98721 (3)	0.96846 (6)	0.98856 (2)	0.98017 (4)	0.98888 (1)	0.98001 (5)
	4	0.99480 (3)	0.991842 (5)	0.99561 (2)	0.99325 (4)	0.98924 (6)	0.99766 (1)
	5	0.98214 (2)	0.93395 (5)	0.99376 (1)	0.93515 (4)	0.97002 (3)	0.91264 (6)
	6	0.82843 (4)	0.82729 (5)	0.84157 (1)	0.82484 (6)	0.83205 (2)	0.83173 (3)
	7	0.62015 (3)	0.59953 (5)	0.62293 (2)	0.59474 (6)	0.65096 (1)	0.61955 (4)
Image Segmentation	1	0.99394 (4)	0.97861 (6)	0.99535 (3)	0.99328 (5)	0.99920 (1)	0.99815 (2)
	2	1.00000 (1)	1.00000 (1)	1.00000 (1)	1.00000 (1)	0.99881 (5)	0.99810 (6)
	3	0.93342 (2)	0.87113 (6)	0.92810 (3)	0.91278 (4)	0.95734 (1)	0.90678 (5)
	4	0.94126 (1)	0.90427 (5)	0.93665 (3)	0.92766 (4)	0.94099 (2)	0.90318 (6)
	5	0.91362 (4)	0.86126 (6)	0.91953 (3)	0.90077 (5)	0.95195 (1)	0.93139 (2)
	6	0.99823 (2)	0.98943 (5)	0.99605 (3)	0.99440 (4)	0.99974 (1)	0.98015 (6)
	7	1.00000 (1)	0.99887 (5)	0.99993 (2)	0.99946 (4)	0.99848 (6)	0.99986 (3)
LandSat	1	0.98994 (4)	0.98820 (5)	0.99355 (2)	0.99423 (1)	0.98729 (6)	0.99136 (3)
	2	0.99469 (2)	0.98829 (4)	0.99577 (1)	0.98255 (5)	0.96337 (6)	0.98876 (3)
	3	0.97932 (1)	0.97379 (4)	0.97577 (2)	0.97434 (3)	0.95708 (6)	0.97054 (5)
	4	0.92693 (1)	0.92038 (5)	0.92691 (2)	0.92378 (4)	0.87460 (6)	0.92636 (3)
	5	0.91687 (1)	0.87032 (6)	0.90505 (2)	0.88969 (4)	0.87151 (5)	0.90125 (3)
	6	0.96146 (3)	0.95407 (6)	0.96158 (2)	0.95853 (4)	0.95448 (5)	0.96252 (1)
Page	1	0.93654 (4)	0.90640 (6)	0.93995 (3)	0.94363 (2)	0.95736 (1)	0.92717 (5)
	2	0.84573 (3)	0.61670 (6)	0.71468 (5)	0.74304 (4)	0.93233 (1)	0.87221 (2)
	3	0.81629 (4)	0.92336 (3)	0.79569 (5)	0.93125 (2)	0.94769 (1)	0.57743 (6)
	4	0.97299 (1)	0.91514 (6)	0.96094 (3)	0.94404 (5)	0.96387 (2)	0.95872 (4)
	5	0.83315 (2)	0.81884 (3)	0.81344 (4)	0.87817 (1)	0.79179 (6)	0.81032 (5)
Wave	1	0.86507 (2)	0.86207 (3)	0.87200 (1)	0.84739 (5)	0.84634 (6)	0.86151 (4)
	2	0.91489 (3)	0.91239 (4)	0.91652 (2)	0.90627 (5)	0.89499 (6)	0.92074 (1)
	3	0.91360 (4)	0.91934 (2)	0.91854 (3)	0.90830 (5)	0.90010 (6)	0.92011 (1)

Results using AUC's and Ranks



Conclusions

- This paper presents a novel cost function, whose minimization can adapt the Radial Basis Function (RBF) network into a one-class classifier, i.e., EBFDD.
- EBFDD utilizes elliptical kernels that can elongate and rotate to allow it to learn sophisticated decision surfaces.
- The empirical results show that the EBFDD network has a better overall performance than leading Anomaly Detection techniques across all the experiments.
- EBFDD suffers from:
  - Large number of trainable parameters → Flexibility comes with a price!
  - Covariance matrix inversion is expensive
  - Sensitivity to the number of kernels

Future Work

- Add recurrent connections to the EBFDD network architecture to allow contextual anomalies within streams to be identified

- Deepening the EBFDD network (Esp. for ***Dimensionality Reduction***):
    - Transfer Learning!
    - A hybrid model with a unified cost function
    - A separate feature extractor + EBFDD network

- Investigating streaming scenarios with concept drift / concept shift

Acknowledgements

- This work was supported by Science Foundation Ireland under Grant No. 15/CDA/3520 and Grant No. 12/RC/2289.
- Many thanks to our colleagues, Arjun Pakrashi and Ellen Rushe, for all their support.

www.mldawn.com

Mehran Bazargani

https://github.com/MLDawn

ML Dawn

mehran.hosseinzadehbazarga@ucdconnect.ie

@MLDawn2018

Institution: Insight Centre for Data Analytics, University College Dublin, Ireland