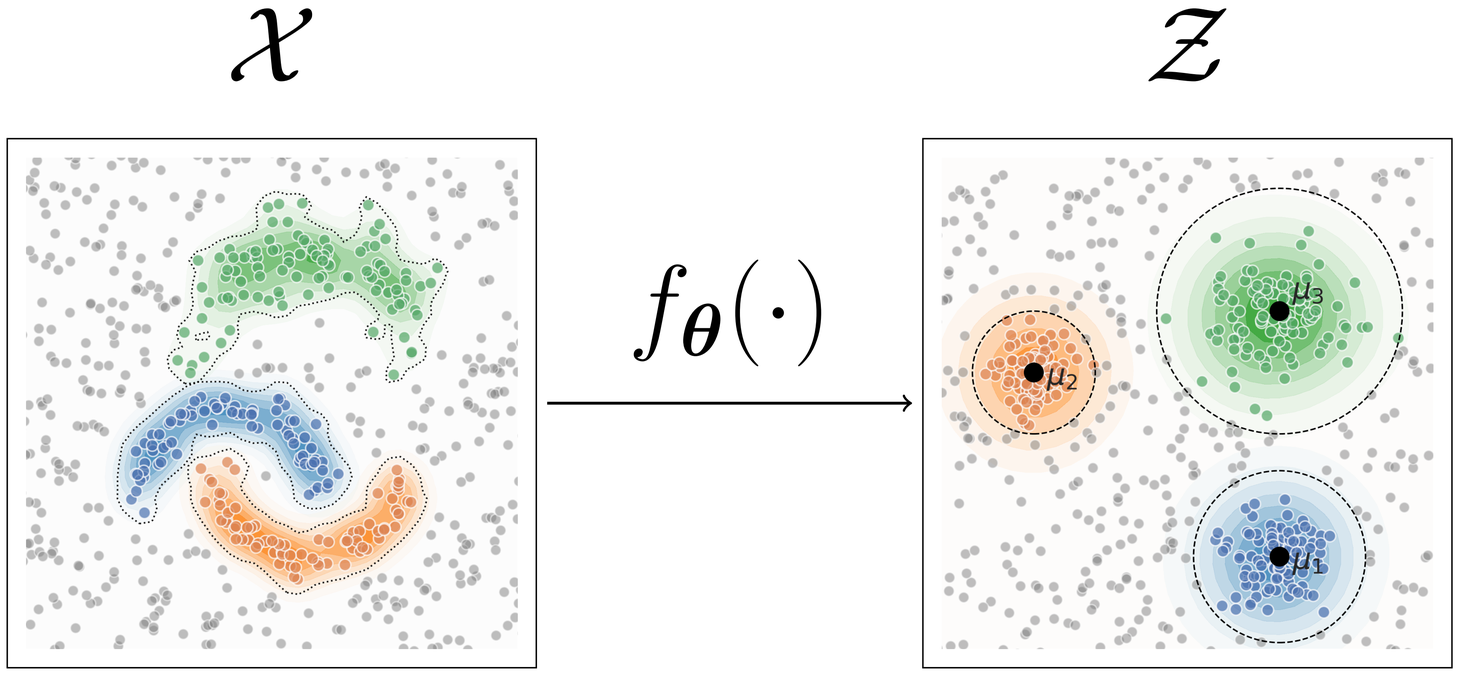
Multi-Class Hypersphere Anomaly Detection

Konstantin Kirchheim, Marco Filax, Frank Ortmeier
surname.lastname@ovgu.de
Department of Computer Science

August 21-25 • Montréal, Québec
ICPR 2022
26TH International Conference on Pattern Recognition

Machine learning-based classification algorithms typically operate under assumptions that assert that the underlying data generating distribution is stationary and draws from a finite set of categories. In some scenarios, these assumptions might not hold, but identifying violating inputs - here referred to as anomalies - is a challenging task. Recent publications propose deep learning-based approaches that perform anomaly detection and classification jointly by (implicitly) learning a mapping that projects data points to a lower-dimensional space, such that the images of points of one class reside inside of a hypersphere, while others are mapped outside of it. In this work, we propose Multi-Class Hypersphere Anomaly Detection (MCHAD), a new hypersphere learning algorithm for anomaly detection in classification settings, as well as a generalization of existing hypersphere learning methods that allows incorporating example anomalies into the training. Extensive experiments on competitive benchmark tasks, as well as theoretical arguments, provide evidence for the effectiveness of our method.

Hypersphere Learning



Training

Objective Function:

$$\sum_{(\mathbf{x},y)\in\mathcal{D}^{in}} \alpha\mathcal{L}_{\Lambda}(y,\mathbf{x}) + \beta\mathcal{L}_{\Delta}(y,\mathbf{x}) \tag{1}$$

- Intra-Class Compactness $\mathcal{L}_{\Lambda}$: capture factors of common variance in a class
- Inter-Class Variation $\mathcal{L}_{\Delta}$: capture distinguishing factors between classes

Inference

Classification:

- Assign class of nearest class center $\arg\min_y \|\mu_y - f(\mathbf{x})\|$ in output space

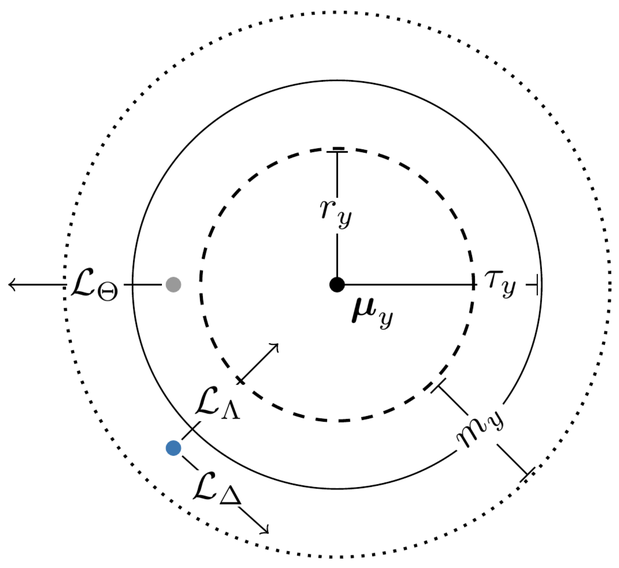
Anomaly Detection:

- Threshold distance to nearest class center $\min_y \|\mu_y - f(\mathbf{x})\|$ in output space

Problems

Existing Hypersphere Learning methods:

- Require additional parameters/auxiliary classifier
- Make assumptions on output space $\mathcal{Z}$ or centers
- Can not integrate examples of anomalies



Multi Class Hypersphere Anomaly Detection (MCHAD)

We propose MCHAD:

- Class Centers μ_y are learable parameters of the model
- Inter-Class Variance $\mathcal{L}_{\Delta}(\mathbf{x},y) = \log\left(1 + \sum_{j\neq y} e^{\|\boldsymbol{\mu}_y - f(\mathbf{x})\|^2 - \|\boldsymbol{\mu}_j - f(\mathbf{x})\|^2}\right)$
- Intra-Class Compactness $\mathcal{L}_{\Lambda}(\mathbf{x},y) = \|\mu_y - f(\mathbf{x})\|$

Generalized Hypersphere Learning

- Include example anomalies
- Introduce loss term $\mathcal{L}_{\Theta}$ to learn factors that discriminate normal from anomalous data
- Applicable to other hypersphere learning methods

$$\sum_{(\mathbf{x},y)\in\mathcal{D}^{in}} \alpha\mathcal{L}_{\Lambda}(y,\mathbf{x}) + \beta\mathcal{L}_{\Delta}(y,\mathbf{x}) + \sum_{\mathbf{x}\in\mathcal{D}^{out}} \gamma\mathcal{L}_{\Theta}(\mathbf{x}) \tag{2}$$

Generalized MCHAD:

$$\mathcal{L}_{\Theta}(\mathbf{x}) = \sum_{y}^{K} \max\{0, (r_y + m_y)^2 - \|f(\mathbf{x}) - \mu_y\|^2\} \tag{3}$$

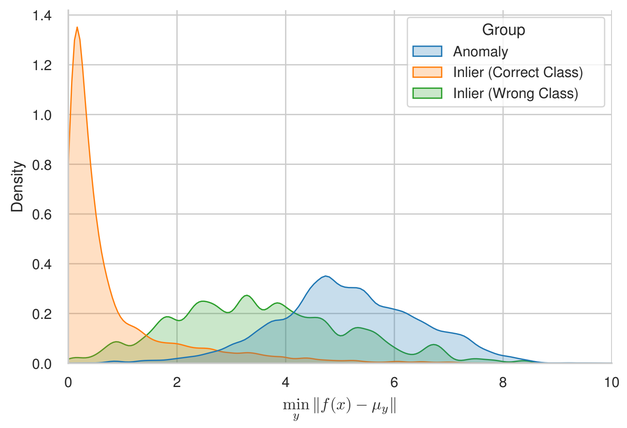
Results

- Each model trained 21 times with different random seeds
- Performance averaged over 7 outlier datasets

Dataset	Model	Accuracy ↑ Mean	$\pm\sigma_{\bar{x}}$	AUROC ↑ Mean	$\pm\sigma_{\bar{x}}$	AUPR-IN ↑ Mean	$\pm\sigma_{\bar{x}}$	AUPR-OUT ↑ Mean	$\pm\sigma_{\bar{x}}$	FPR95 ↓ Mean	$\pm\sigma_{\bar{x}}$
CIFAR10	II	28.83	0.14	57.88	1.21	56.18	1.05	60.73	1.28	83.19	1.81
	CAC	**95.14**	0.01	93.33	0.25	88.63	0.60	95.13	0.19	18.45	0.74
	Center	94.43	0.01	92.14	0.33	88.50	0.47	92.26	0.40	31.48	1.66
	MCHAD (ours)	94.84	0.01	94.33	0.34	89.94	0.64	95.95	0.22	15.81	0.77
	G-MCHAD (ours)	94.72	0.01	**96.65**	0.19	**94.23**	0.40	**97.54**	0.14	**10.49**	0.53
	G-Center (ours)	94.29	0.01	93.51	0.49	89.68	0.80	94.85	0.40	19.19	1.21
	G-CAC (ours)	83.83	1.91	85.74	1.73	84.13	1.57	88.10	1.40	31.32	2.55
CIFAR100	II	5.78	0.09	48.70	1.23	47.83	0.91	54.46	1.16	87.82	1.62
	CAC	75.80	0.03	76.00	1.00	70.89	1.20	79.49	0.91	57.68	1.96
	Center	76.54	0.02	75.01	1.30	69.54	1.33	78.64	1.20	57.47	2.26
	MCHAD (ours)	**77.49**	0.02	80.59	0.99	73.32	1.13	84.68	0.85	47.24	2.16
	G-MCHAD (ours)	77.24	0.03	**83.44**	1.05	**80.46**	1.09	**85.76**	0.95	**45.97**	2.43
	G-Center (ours)	68.18	0.10	70.60	2.37	76.23	1.72	70.82	2.04	64.15	3.41
	G-CAC (ours)	70.49	1.05	69.61	1.54	66.89	1.44	73.97	1.21	65.15	2.06

Ablation Studies

- Observed distance consistent with imposed learning objectives
- Model can also detect classification errors
- All loss terms are required



Code

Code for all experiments is available online under the MIT license. The implementation is based on PyTorch, and we use the `pytorch-ood` package.