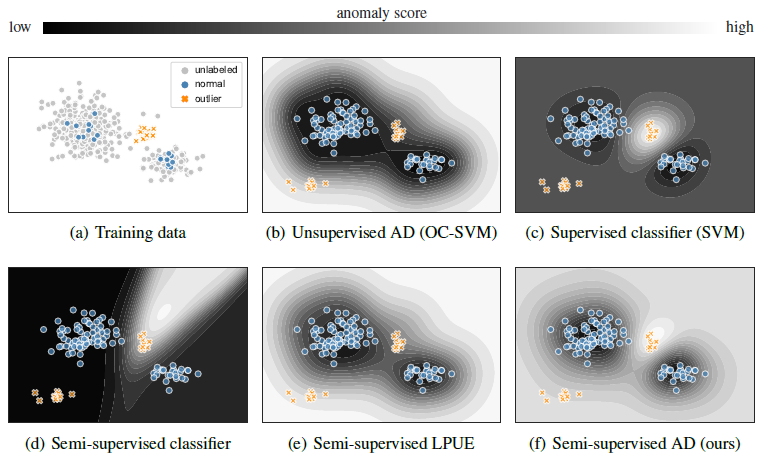
Deep SVDD

Link: <https://medium.com/analytics-vidhya/deep-semi-supervised-anomaly-detection-ab1db59d7820#:~:text=An%20existing%20unsupervised%20technique%20named,a%20pre%2Ddetermined%20point%20c>.

Anomaly detection (AD) is the task of identifying outliers in a given dataset. There are existing shallow supervised, as well as deep unsupervised techniques which are limited in either scalability or their ability to use labeled anomalous data.

In this article, we will be discussing deep semi-supervised anomaly detection (Deep SAD) as originally published by Ruff, et al. [1].

**Motivation behind Deep SAD**

Figure 1. (a) shows the distribution of training data, along with labels, (b-e) show 4 existing techniques and the contour maps learned by their model. (f) shows deep semi-supervised anomaly detection

At a high level, we can compare the performance of the existing techniques with Deep SAD, and look at the representation learned by each model. The training data is presented as i) labeled normal ii) labeled outlier iii) unlabeled.

Unsupervised AD (b) only learns the distribution of the normal representation. This leads to blurry boundaries between normal and anomalous data. The outcome of this shortfall are low confidence detection of anomalies.

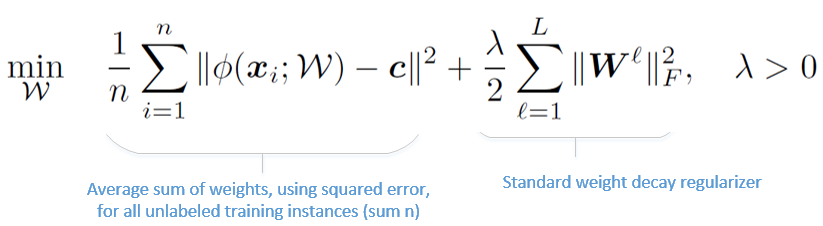
Supervised classifier (SVM) (c) successfully creates a maximal separating hyperplane between normal and anomalous boundaries, but would fail to detect anomalies in the case they are out-of-distribution (OOD).

The semi-supervised classifier (d) only learns the anomalous class, and would very likely fail at detecting any anomaly distributions that it was not pre-trained to detect.

Semi-supervised LPUE (Learning from Positive and Unlabeled Examples) (e) is an existing method that does not consider labeled anomalous data for training. The short coming is evident in the lack of a clear separation between the two distributions of labeled anomalous data (part of training set) and normal data.

**Deep SVDD**

An existing unsupervised technique named Deep Support vector data description (SVDD) is used as the motivation for Deep SAD. The objective of this technique is to train the neural network (*phi*) to learn a transformation that minimizes the volume of a data-enclosing hypersphere centered on a pre-determined point *c*.

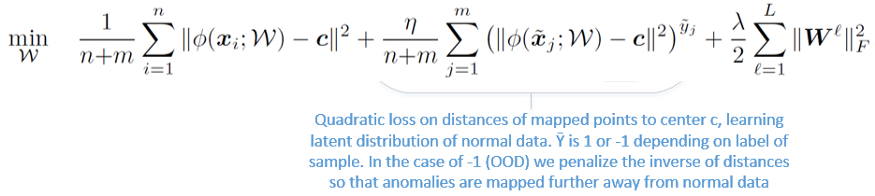


Equation 1: Deep SVDD equation representing an unsupervised AD framework based on OOD samples

Penalizing the mean squared forces the network to extract common factors of data variation which are the most stable in the dataset. As a consequence, normal data points get mapped near the hypersphere center, while anomalies are mapped further out.

Deep SVDD equation is equivalent to minimizing the empirical variance and thus an upper bound on the entropy of a latent Gaussian. In simple terms, this technique minimizes entropy around point*c*and within the arbitrary hypersphere.

**Deep SAD**



Equation 2: Deep SAD equation with a supervised learning expression the second term. This second term optimizes weights based on label

This equation is the same as Deep SVDD, except for the second expression.

In the second expression, η (eta) is hyperparameter controlling the amount of emphasis placed on labeled vs unlabeled data.

Parameter m represents labeled samples and the y-bar represents a value of -1 or 1 depending on whether it belongs to anomalous or normal distribution, respectively.

We can summarize Deep SAD objective as modeling the latent distribution of normal data, to have low entropy, and the latent distribution of anomalies, to have high entropy. As discussed in previous section, Deep SAD increases the affinity of weights to be closer to *c*for the normal distribution, and further out for the anomalies.

**Conclusion and Future Work**

Deep SAD is a generalization of unsupervised Deep SVDD method. The goal of Deep SAD is to extend the Deep SVDD approach to train with labeled anomalies. In doing so, it is more effectively able to anticipate anomalies that may be sampled from various distributions.

The results discussed in the experiments section suggest that general semi-supervised anomaly detection should be preferred when some labeled information on anomalous distribution is available.

Potential future works can include more rigorous analysis and studying deep anomaly detection under rate distortion curve, for example.

**References**

[1] Lukas Ruff, et al. “Deep Semi-Supervised Anomaly Detection.” International Conference on Learning Representations.

[2] Lukas Ruff, Robert A Vandermeulen, Nico Görnitz, Lucas Deecke, Shoaib A Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In ICML, volume 80, pp. 4390–4399, 2018.

Reference paper: <https://arxiv.org/abs/1906.02694>

Code for Experiments section: <https://github.com/lukasruff/Deep-SAD-PyTorch>

**Out-of-Distribution (ODD)**

Link: <https://deepchecks.com/glossary/out-of-distribution/>

For Language and Vision activities, the term “distribution” has slightly different meanings. Consider a task to classify cat breed photographs; photographs of cats would be in-distribution, while photographs of dogs, humans, balls, and other objects would be out-of-distribution.

The data distribution in real-world activities generally drifts with time, and tracking a developing data distribution is expensive.

* **OOD identification is critical in preventing AI systems from making predictions that are incorrect.**

**Various ODD detection techniques**

**Ensemble Learning**

This model is used to produce predictions for each data point in Ensemble Learning, and the decisions made by these models are merged to improve overall performance. There are several methods for combining decisions:

**Averaging**– for regression tasks, averaging the predictions of all models is simple, but for classification problems, we can average the softmax probabilities.

**Weighted Averaging** — In this methodology, different weights are allocated to models, and the final prediction is computed using a weighted average.

**Maximum Voting** – the final prediction is based on the majority of the models’ predictions.

Combining the judgments, in this case, refers to calculating the forecast confidence across various models.

Summary:

Deep SVDD is a type of deep learning algorithm used for anomaly detection, which aims to identify instances in a dataset that are significantly different from the majority of the data. The "SVDD" stands for "support vector data description", which is a type of one-class classification algorithm that is used as the basis for the Deep SVDD approach.

In the Deep SVDD approach, a neural network is trained to map the input data to a low-dimensional space where the normal instances are expected to be tightly clustered. This is achieved by minimizing the distance between the network output and the center of the normal instances in this space. The goal is to learn a compact representation of the normal data that can be used to detect anomalous instances as those that are significantly far away from the center of the normal instances in the low-dimensional space.

The Deep SVDD algorithm is trained in an unsupervised manner, meaning that only normal data is used during training. This is because the algorithm aims to learn a representation of the normal data that can be used to identify anomalies, rather than explicitly modeling the anomalies themselves.

The Deep SVDD algorithm has several advantages over traditional anomaly detection approaches. For example, it is able to handle high-dimensional data and can learn complex representations of the data that may not be possible with traditional methods. Additionally, it is able to identify anomalies in new data that may not have been seen during training, making it useful for detecting previously unknown types of anomalies.

In summary, Deep SVDD is a type of deep learning algorithm used for anomaly detection that aims to learn a compact representation of normal data and identify anomalies as those that are significantly far away from the center of the normal data in a low-dimensional space. It is trained in an unsupervised manner and is able to handle high-dimensional and complex data, making it a useful approach for detecting anomalies in various domains.