

Deep One-Class Classification

July 11, 2022

1 Related Work

1. Kernel Based Models:

- (a) One Class SVM (OC-SVM): Finding a maximum margin hyperplane in feature space, separating mapped data from origin:
- (b) Support Vector Data Description (SVDD): Finding minimum radius hypersphere separating mapped data in feature space:

$$\begin{aligned} \min_{R, c, \xi} R^2 + \frac{1}{\nu n} \sum_i \xi_i \\ \text{s.t. } \forall i : \|\phi_k(x_i) - c\|_{\mathcal{F}_k}^2 \leq R^2 + \xi_i, \xi_i \geq 0 \end{aligned} \quad (1)$$

With R : radius, c : hypersphere's center in feature space, ξ_i slack variables and $\nu \in (0, 1]$ hyperparameter controlling trade-off between radius and slack variable penalties. Points outside sphere are considered anomalous.

2. Deep Approaches:

- (a) : Deep Autoencoders: These networks can extract common features of normal samples in their intermediate representation and reconstruct them accurately. Hence reconstruction error is a good metric for anomaly score.
- (b) Generative Adversarial Networks (GANs)

2 Deep SVDD

Let $\phi(\cdot; \mathcal{W}) : \mathcal{X} \rightarrow \mathcal{F}$ be a neural network with L hidden layers with weights \mathcal{W} . **Soft-boundary Deep SVDD** loss function:

$$\min_{R, \mathcal{W}} R^2 + \frac{1}{\nu n} \sum_{i=1}^n \max(0, \|\phi_k(x_i) - c\|^2 - R^2) + \frac{\lambda}{2} \sum_{l=1}^L \|\mathbf{W}^l\|_F^2 \quad (2)$$

with radius $R > 0$ and center $c \in \mathcal{F}$ and hyperparameter ν controlling trade-off between volume of sphere and boundary violations.

If most of training data is normal, we can use a simplified form of above objective, **One Class Deep SVDD** loss function:

$$\min_{\mathcal{W}} \frac{1}{n} \sum_{i=1}^n \|\phi_k(x_i) - c\|^2 + \frac{\lambda}{2} \sum_{l=1}^L \|\mathbf{W}^l\|_F^2 \quad (3)$$

Properties of Deep SVDD

- If \mathcal{W}_0 be the set of all-zero weights and $c_0 = \phi(x; \mathcal{W})$ (for any input point x) setting $c = c_0$ will result in trivial solution $\mathcal{W}^* = \mathcal{W}_0$ and $R^* = 0$ with zero loss (hypersphere collapse). Similarly, including hypersphere center in optimization variables will result in the same behavior. Fixing c as the mean of the result of performing an initial (untrained) forward pass on some training data is a good candidate and making convergence faster and more robust.
- Having a bias term or a bounded activation function can result in learning center of hypersphere with zero weights. Thus bias terms and bounded activations should not be used in neural networks with Deep SVDD.
- **ν -property:** Hyperparameter ν in soft-boundary Deep SVDD objective in (2), is an upper bound on the fraction of outliers.