DMW Assignment-6

Submitted By - [Akhil Shukla, IIT2018112] [Akhil Singh, IIT2018198][Javed Ali, IIT2018501][Manan Bajaj, IIT2018502][Lokesh, IIT2018503]

6 th Semester, B.Tech, Department of Information Technology, IIIT Allahabad

You have to understand the algorithm proposed in the paper "Deep Support Vector Data Description for Unsupervised and Semi-Supervised Anomaly Detection".

Run the algorithm on the shared given two datasets and show the accuracy in terms of the attached image table: (make one more column in the last name SS-Deep-SVDD with the new algorithm and give the result).

Semi Supervised Deep SVDD (Literature):

This implementation is used if we have some labelled data and some unlabelled data. Suppose if we have *m* samples of $(\tilde{x}_1, \tilde{y}_1), \ldots, (\tilde{x}_m, \tilde{y}_m) \in X \times Y$ along with $n \in N$ unlabeled samples $x_1, \ldots, x_n \in X$ with $X \subseteq R^d$ and $Y = \{-1, +1\}$. Here, y=+1 denotes normal samples and y=-1 denotes anomaly sample.

Soft-Boundary SS-DSVDD problem:

$$\min_{R,\mathcal{W}} \; R^2 + \frac{1}{\nu(n+m)} \sum_{i=1}^{n} l\left(R^2 - \|\phi(\boldsymbol{x}_i;\mathcal{W}) - \boldsymbol{c}\|^2\right) + \frac{\eta}{\nu(n+m)} \sum_{j=1}^{m} l\left(\tilde{y}_j \left(R^2 - \|\phi(\tilde{\boldsymbol{x}}_j;\mathcal{W}) - \boldsymbol{c}\|^2\right)\right),$$

where $l(z) = \max\{0, -z\}$ is the hinge loss.

Here we require normal samples(y=+1) to be present inside and anomaly ones(y=-1) to be outside of the hypersphere.

Penalty is given by penalty is given by $R^2 - ||\varphi(\tilde{x}_j;W) - c||^2$ and $||\varphi(\tilde{x},;W) - c||^2 - R^2$.

Algorithm used in the paper for Optimization of SS-DSVDD:

Input:

Unlabeled data: x1, . . . , xn

Labeled data: $(\tilde{x}_1, \tilde{y}_1), \ldots, (\tilde{x}_m, \tilde{y}_m)$

Hyperparameters: v, η, λ

SGD learning rate: ε

Output:

Trained model: (R^*, W^*)

Initialize:

Neural network weights: W

Hypersphere parameters: R, c

for each epoch do

for each mini-batch do

Draw mini-batch B

$W \leftarrow W - \varepsilon \cdot \nabla W_J(R, W; B)$

Solve for R on mini-batch B

end for
end for

Architecture Used in this Paper:

1. We use LeNet type architecture CNNs, where each CNN layer consist of three modules, where the modules consist of
   a. $32 \times (5 \times 5 \times 3)$-filters
   b. $64\times(5\times5\times3)$-filters
   c. $128\times(5\times5\times3)$-filters
   and finally a dense layer of 128 units.
2. Batch size is 200 and lambda is 10^{-6}.

Dataset

Here , we are using CIFAR-10 dataset.
Link - https://www.cs.toronto.edu/~kriz/cifar.html

Observation:

Input Parameters for pretraining -

1. ν parameter: 0.10
2. Pretraining optimizer: adam
3. Pre-training learning rate: 0.0001
4. Pre-training epochs: 350
5. Pre-training batch size: 200

Pre-training time: 26565.082

After pre-training is being done , then auto-encoder is tested.

Auto-encoder testing -

1. Test set Loss: 3.13629933
2. Test set AUC: 57.40%

Input Parameters for training -

1. Training optimizer: adam
2. Training learning rate: 0.0001
3. Training epochs: 150
4. Training batch size: 200

Training Time : 5026.234

Testing Results :

1. Testing time: 27.702
2. Test set AUC: 56.70%

References

[1] Ruff, L., Robert A. Vandermeulen, N., Deecke, L., Siddiqui, S. A., Binder, A., Emmanuel Muller, Kloft, M. (2018, July). "Deep Support Vector Data Description for Unsupervised and Semi-Supervised Anomaly Detection", Proceedings of the ICML 2019 Workshop on Uncertainty and Robustness in Deep Learning, Long Beach, CA, USA (pp. 9-15)

[2] Dataset: https://www.cs.toronto.edu/~kriz/cifar.html