DMW Assignment-5

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 One-Class Classification".

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 Deep-SVDD with the new algorithm and give the result).

Deep SVDD (Literature):

In this, some input space $X \subseteq R^d$ and output space $F \subseteq R^p$, let $\varphi(\cdot\ ; W) : X \rightarrow F$ be a neural network with $L \in N$ hidden layers and a set of weights $W = \{W^1, \ldots, W^L\}$. So we represent it as $\varphi(x; W) \in F$ is the feature representation of $x \in X$ given by network φ with parameters W.

Here we need to learn the parameters W, together with minimising the volume of a data-enclosing hypersphere in output space F that is characterized by radius $R > 0$ and center $c \in F$.

$D_n = \{x_1, \ldots, x_n\}$ on X

$$\min_{R,\mathcal{W}} \quad R^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\max\{0, \|\phi(\boldsymbol{x}_i;\mathcal{W}) - \boldsymbol{c}\|^2 - R^2\} + \frac{\lambda}{2}\sum_{\ell=1}^{L}\|\boldsymbol{W}^\ell\|_F^2.$$

Here we have to minimize the volume of the hypersphere. And here the second term is the penalty term for those points that are lying outside the hypersphere, i.e., the distance between the center of the hypersphere is greater than R. That last term has weight decay regularizer on the network parameters W with hyperparameter $\lambda > 0$.

In one class Deep SVDD, we use

$$\min_{\mathcal{W}} \quad \frac{1}{n}\sum_{i=1}^{n}\|\phi(\boldsymbol{x}_i;\mathcal{W}) - \boldsymbol{c}\|^2 + \frac{\lambda}{2}\sum_{\ell=1}^{L}\|\boldsymbol{W}^\ell\|_F^2. \quad (4)$$

In this, One-Class Deep SVDD simply applies a quadratic loss for penalizing the distance of every network representation $\varphi(x_i\ ; W)$ to $c \in F$. Hyperparameter $\lambda > 0$ is a network weight decay regularizer.

In one class classification, we find a hypersphere of minimum volume with center c. But in Soft boundary SVDD, the hypersphere is contracted by penalizing the radius directly and the data representations that fall outside the sphere.

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. $64 \times (5 \times 5 \times 3)$-filters
    c. $128 \times (5 \times 5 \times 3)$-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 one-class classification”, International conference on machine learning (pp. 4393-4402). PMLR.

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