Energy Based Out-Of-Distribution Detection

20d070020 : Aziz Shameem
200050154 : Vedang Asgaonkar
200070077 : Shivam Patel
200050131 : Shikhar Mundra

Description

- Our project involved using energy based models to classify data as ID or OOD
- We train the model to assign high energy values to OOD data and low values to ID data.
- We attempt to create a clear delineation between the energy levels of ID and OOD data by making use of various regularizers
- We also attempt to modify the architecture by incorporating a LSTMs to allow it to extend to text data.

Experiments

- We experimented with the energy based model on the CIFAR 10 dataset. We attempted to fine tune the energy based regularizer as proposed in the paper

$$L_{\text{energy}} = \mathbb{E}_{(\mathbf{x}_{\text{in}}, y) \sim \mathcal{D}_{\text{in}}^{\text{train}}} \left(\max(0, E(\mathbf{x}_{\text{in}}) - m_{\text{in}})\right)^2 + \mathbb{E}_{\mathbf{x}_{\text{out}} \sim \mathcal{D}_{\text{out}}^{\text{train}}} \left(\max(0, m_{\text{out}} - E(\mathbf{x}_{\text{out}}))\right)^2$$

We observed that the values m_out = 23 and m_in = 5 work well in practice and we present their results in the result section

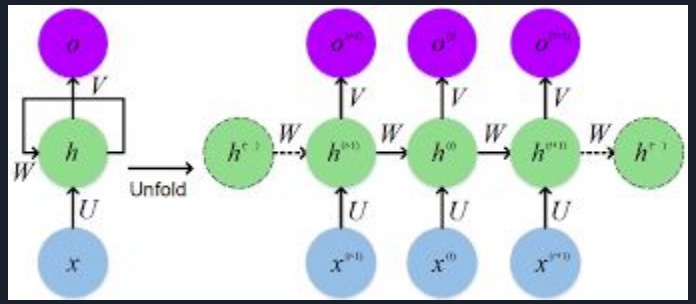
Experiments

- We also replace the energy based regularizer in the paper with a ranking loss regularizer with a margin of 20, similar to the RELU based ranking loss. The ranking loss encourages the energies of ID and OOD samples to be separated by a margin of m.

$$L_{energy} = \sum_{x_{in}} \sum_{x_{out}} ReLU(m + E(x_{in}) - E(x_{out}))$$

Experiments

- We attempt to generalize this approach to text data by using an LSTM in our architecture
- Experiments are done with different kinds of regularizers to create a lower error rate



Experimental Results

We summarize the results of our experiments with the CIFAR10 dataset:

Texture dataset as OOD

	Ranking (our)	Energy Ft (fine tuned by us)	Energy Baseline
FPR 95	0.24	0.34	0.52
AUROC	99.81	99.81	85.27
AUPR	99.96	99.96	95.38

Experimental Results

Speckle Noise images as OOD

	Ranking (our)	Energy Ft (fine tuned by us)	Energy Baseline
FPR 95	0.31	0.24	0.66
AUROC	99.91	99.93	89.37
AUPR	99.97	99.97	97.96

Text Classification

- Attempt at using Energy framework for OOD detection on text data
- Tried on 20news dataset, using LSTM based architecture
- Results not great, but some improvement using ranking loss

Num Epochs	Method	Test Error
1	Energy	89.78
1	Ranking	89.70

Observations

- We observe that using the ranking regularizer leads a reduction in the false positivity rate (FPR 95), as compared to the energy based regularizer
- We also observe an improvement in the FPR for 20news datasets using the ranking loss. However performance of the LSTM on the 20news dataset is not that great
- We conjecture that using a transformer instead of an LSTM would perform better on the 20news dataset.

References

[1] S. Elflein, B. Charpentier, D. Z¨ugner, and S. G¨unnemann, “On out-of-distribution detection with energy-based models,” CoRR, vol. abs/2107.08785, 2021.

[2] F. Tonin, A. Pandey, P. Patrinos, and J. A. K. Suykens, “Unsupervised energy-based out-of-distribution detection using stiefel-restricted kernel machine,” in 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, 2021.