Id155: Distributional Smoothing with Virtual Adversarial Training
Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, Ken Nakae, Shin Ishii (Kyoto University, Japan)

Introduction

Construct a “Good” classifier with new regularization method.

Input: \( x \)

Model

\[ p(y|x, \theta) \]

Output: \( y \)

Advantages of our method

- Applicability to both supervised and semi-supervised training.
- At most two hyperparameters (\( \varepsilon \) and \( \lambda \)).
- Parametrization invariant formulation.
- Low computational cost.

What are the properties of a good classifier?

- Recognition should be robust against minor perturbations!

Adversarial example

\[ \theta \]

57.7% confidence

\[ \text{“noise”} \]

8.2% confidence

99.3% confidence


Local Distributional Smoothness (LDS)

- Linear regression, \( \text{LDS}(x, \theta) = e^T \theta^T x \)
- Logistic regression, \( \text{LDS}(x, \theta) = \frac{1}{2} e^T \theta^T x + \log 1 + e^T \theta^T x \)

Computation of LDS

Seemingly insurmountable bottleneck of computing LDS.

- Approximate \( \Delta_{KL} \) with 2nd Taylor expansion:

\[ \Delta_{KL}(x, \theta) \approx \frac{1}{2} \| H(x, \theta) \|_2^2 \]

where \( H(x, \theta) = \nabla_x \Delta_{KL}(x, \theta) \approx 0 \)

- Approximate \( \Delta_{KL} \) with finite difference method:

\[ \Delta_{KL}(x, \theta) \approx \frac{1}{2} \| \nabla_x \Delta_{KL}(x, \theta) \|_2^2 \]

\( r_{\varepsilon, \theta} \) is also the dominant eigenvector of \( H(x, \theta) \) with magnitude \( \varepsilon \) :

\[ r_{\varepsilon, \theta} = \arg \max \| H(x, \theta) \|_2 \]

Virtual Adversarial Training (VAT)

- Local Distributional Smoothness (LDS):

\[ \text{LDS}(x, \theta) \equiv - \Delta_{KL}(x, \theta) \]

- LDS regularized objective function

\[ \frac{1}{N} \sum_{i=1}^{N} \log p(y|x_i, \theta) + \frac{1}{N} \sum_{i=1}^{N} \text{LDS}(x_i, \theta) \]

Advantages of VAT

- Construct a “Good” classifier with new regularization method.
- Applicability to both supervised and semi-supervised learning.
- At most two hyperparameters (\( \varepsilon \) and \( \lambda \)).
- Low computational cost.
- Semi-supervised learning tasks (permutation invariant tasks)

MNIST

- SVM (street view housing numbers)
- SVHN (street view housing numbers)
- NORB (NYU Object Recognition Benchmark)

Our approach was effective for supervised and semi-supervised learning for benchmark datasets.

- With only 1 hyperparameter \( \varepsilon \), our method achieved good performance.

References