Large Margin Meta-Learning for Few-Shot Classification

The University of Hong Kong[1], The Hong Kong Polytechnic University[2]

Yong Wang[1], Xiao-Ming Wu[2], Qimai Li[2], Jiatao Gu[1], Wangmeng Xiang[2], Lei Zhang[2], Victor O.K. Li[1]

Large Margin Principle

$$\mathcal{L} = \mathcal{L}_{\text{softmax}} + \lambda * \mathcal{L}_{\text{large-margin}}$$

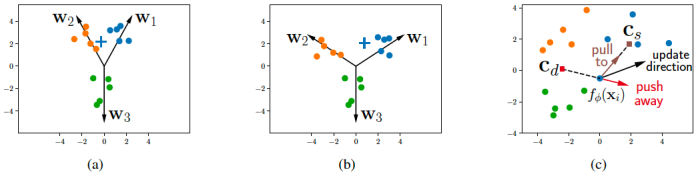

Fig. 1: Large margin meta-learning. (a) Classifier trained without the large margin constraint. (b) Classifier trained with the large margin constraint. (c) Gradient of the triplet loss.

One Implementation: Triplet Loss

$$\mathcal{L}_{\text{large-margin}} = \frac{1}{N_t}\sum_{i=1}^{N_t}\left[\| f_\phi(\mathbf{x}_i^a) - f_\phi(\mathbf{x}_i^p)\|_2^2 - \| f_\phi(\mathbf{x}_i^a) - f_\phi(\mathbf{x}_i^n)\|_2^2 + m\right]_+.$$

Case study

- Graph Neural Network (GNN)
- Prototypical Network (PN)

Analysis

After rearrangement:

$$\mathcal{L}_{\text{large-margin}} = \frac{1}{N_t}\left(\sum_{\mathbf{x}_s\in S_s}\| f_\phi(\mathbf{x}_i) - f_\phi(\mathbf{x}_s)\|_2^2 - \sum_{\mathbf{x}_d\in S_d}\| f_\phi(\mathbf{x}_i) - f_\phi(\mathbf{x}_d)\|_2^2\right) + \text{const.}$$

The gradient:

$$\begin{aligned}\frac{\partial\mathcal{L}_{\text{large-margin}}}{\partial f_\phi(\mathbf{x}_i)} &= \frac{2}{N_t}\left(\sum_{\mathbf{x}_s\in S_s}(f_\phi(\mathbf{x}_i) - f_\phi(\mathbf{x}_s)) - \sum_{\mathbf{x}_d\in S_d}(f_\phi(\mathbf{x}_i) - f_\phi(\mathbf{x}_d))\right)\\ &= -\frac{2|S_s|}{N_t}\left(\frac{1}{|S_s|}\sum_{\mathbf{x}_s\in S_s} f_\phi(\mathbf{x}_s) - f_\phi(\mathbf{x}_i)\right) - \frac{2|S_d|}{N_t}\left(f_\phi(\mathbf{x}_i) - \frac{1}{|S_d|}\sum_{\mathbf{x}_d\in S_d} f_\phi(\mathbf{x}_d)\right)\\ &= \underbrace{-\frac{2|S_s|}{N_t}(\mathbf{c}_s - f_\phi(\mathbf{x}_i))}_{\text{pull towards its own class}} \underbrace{-\frac{2|S_d|}{N_t}(f_\phi(\mathbf{x}_i) - \mathbf{c}_d)}_{\text{push away from other classes}}.\end{aligned}$$

Features

- We implement and compare several of other large margin methods for few-shot learning.
- Our framework is simple, efficient, and can be applied to improve existing and new meta-learning methods with very little overhead.