Theoretical understanding of CL

Theoretical Understanding of CL

framework:

setup

$\mathcal{X} \rightarrow$ Dataset

$D_{sim} \sim (x, x^+)$

$D_{neg} \sim x_1^-, x_2^-, \cdots x_k^-$

Encoder family $\mathcal{F}$

$\therefore f : \mathcal{X} \rightarrow \mathbb{R}^d$ such that, $\|f(\cdot)\| \leq R$; $R > 0$

Latent class: $(x, x^+) \rightarrow$ similar pair

more setup

family of latent class $\mathcal{C}$

$c \in \mathcal{C}$

D_c over the $\mathcal{X}$

$D_c(x)$: How relevant x to c ??.

ρ : how class occurs naturally.

Semantic Similarity:

$$D_{sim}(x, x^+) = \mathop{E}_{c \sim \rho} D_c(x) \, D_c(x^+)$$

$$D_{neg}(x^-) = \mathop{E}_{c \sim \rho} D_c(x^-)$$

Supervised Tasks: for a labeled pair (x, c) ; $c \in \{c_1 \cdots c_{k+1}\}$

$$D_{\mathcal{T}}(x, c) = D_c(x) \, D_{\mathcal{T}}(c)$$

Theoretical understanding of CL

framework:

setup

- $\mathcal{X} \to$ Dataset
- $D_{sim} \sim (x, x^+)$
- $D_{neg} \sim x_1^-, x_2^-, \cdots x_k^-$
- Encoder family $\mathcal{F}$
- $\therefore f : \mathcal{X} \to \mathbb{R}^d$ such that, $\|f(\cdot)\| \le R$; $R > 0$

Latent class: $(x, x^+) \to$ similar pair

more setup

- family of latent class $\mathcal{C}$
- $c \in \mathcal{C}$
- D_c over the $\mathcal{X}$
- $D_c(x)$: How relevant x to c .??
- ρ : how class occurs naturally.

Semantic Similarity:

$$D_{sim}(x, x^+) = \mathbb{E}_{c \sim \rho} D_c(x) D_c(x^+)$$

$$D_{neg}(x^-) = \mathbb{E}_{c \sim \rho} D_c(x^-)$$

Supervised Tasks'. for a labeled pair (x, c) ; $c \in \{c_1 \cdots c_{k+1}\}$

$$D_{\mathcal{T}}(x, c) = D_c(x) \, D_{\mathcal{T}}(c)$$

Evaluation metric for representation:

Task $\mathcal{T} = \{c_1 \cdots c_{k+1}\}$

functions $g : \mathcal{X} \to \mathbb{R}^{k+1}$ // linear classifier.

point $(x, y) \in \mathcal{X} \times \mathcal{T}$

loss $\triangleq L(\{g(x)_y - g(x)_{y'}\}_{y \neq y'})$ [Different from true class should be high]

// f (k dim vector of Differences in Coordinate)

By considering standard hinge loss:

$$l(v) = \max\{0, 1 + \max_i\{-v_i\}\}$$

logistic loss $l(v) = \log_2\left(1 + \sum_i \exp(-v_i)\right)$; $v \in \mathbb{R}^k$

$$L_{sup}(\mathcal{T}, g) := \mathop{E}_{(x,c) \sim \mathcal{D}_c}\left[l\{g(x)_c - g(x)_{c'}\}_{c \neq c'}\right]$$

For linear Classifier: $g(x) = W f(x)$

finally $(k+1)$

Further, $L_{sup}(\mathcal{T}, f) = \inf_{W \in \mathbb{R}^{(k+1) \times d}} L_{sup}(\mathcal{T}, Wf)$

TP: $W \Rightarrow$ mean for each class representation:

Mean classifier: W^μ

c-th row $\Rightarrow \mu_c = \mathop{E}_{x \sim \mathcal{D}_c}[f(x)]$

Now, $L^{\mu}_{sup}(\mathcal{T}, f) = L_{sup}(\mathcal{T}, W^\mu f)$

(P) Theor. understanding of CL

Avg Supervised loss:

$$L_{sup}(f) := \underset{\{c_i\}_{i=1}^{k+1} \sim \rho^{k+1}}{E} \left[L_{sup}\left(\{c_i\}_{i=1}^{k+1}, f\right) \mid c_i \neq c_j \right]$$

for mean class

$$L^{\mu}_{sup}(f) := \underset{\{c_i\}_{i=1}^{k+1} \sim \rho^{k+1}}{E} \left[L^{\mu}_{sup}\left(\{c_i\}_{i=1}^{k+1}, f\right) \mid c_i \neq c_j \right]$$

CL Algorithm:

unsupervised loss: Population loss:

Neg number

$$L_{un}(f) := E\left[\ell\left(\{ f(x)^T (f(x^+) - f(x_i^-)) \}_{i=1}^{k} \right) \right]$$

Empirical counterparts: $(x_j, x_j^+, x_{j1}^-, \ldots x_{jk}^-)_{j=1}^{M} \in D_{sim} \times D_{neg}^{k}$

$$\hat{L}_{un}(f) = \frac{1}{M} \sum_{j=1}^{M} \ell\left(\{ f(x_j)^T (f(x_j^+) - f(x_{ji}^-)) \}_{i=1}^{k} \right)$$

Now,

$$L_{un}(f) = \underset{c^+, c^- \sim \rho^{k+1}}{E} \; \underset{\substack{x, x^+ \sim D_{c^+}^2 \\ x_i^- \sim D_{c^-}}}{E} \left[\ell\left(\{ f(x)^T (f(x^+) - f(x_i^-)) \} \right) \right]$$

Results and theorems

Th. 1. $L_{sup}(\hat{f}) \leq \alpha L_{un}(f) + \eta\, Gen_M + \delta \quad \forall f \in \mathcal{F}$

upper bound?? generalization error.

$M \to \infty, \quad Gen_M \to 0$

$\alpha, \eta \to 1, \quad \delta \to 0$

[if c is large, $L_{un}(f)$ can be small]

$L_{sup}(\hat{f}) \leq L_{un}^{\neq}(f) + \beta s(f) + \eta\, Gen_M \quad \forall f \in \mathcal{F}$

ρ dependent.

$\rho \to$ uniform, $|e| \to \infty$ then $\beta \to 0, \eta \to 1$

Ideal result should be: $\boxed{L_{sup}(\hat{f}) \leq \alpha L_{sup}(f) + \eta\, Gen_M} \quad \forall f \in \mathcal{F}$

However not true??