Max-margin contrastive learning

preleminaries cont. learn.

$$\mathcal{D} = \{x_i\}_{i=1}^{N}$$

$$x_i \in \mathbb{R}^d$$

transformation: $\mathcal{T}: \mathbb{R}^d \to \mathbb{R}^d$

$$x = \{\mathcal{T}(x)\}$$

data dimension

embedding dim.

mapping $f_\theta : \mathbb{R}^d \to \mathbb{R}^{d'}$

CL objective: (NCE one)

$$-\sum_{x \in B} \log \frac{g(f_\theta(x), f_\theta(x^+))}{g(f_\theta(x), f_\theta(x^+)) + \sum_{x^- \in B \setminus x} g(f_\theta(x), f_\theta(x^-))}$$

CL objective: (max-margin triplet loss: margin t

$$\sum_{x, x^+, x^-} \left[t - \mathrm{sim}\left(f_\theta(x), f_\theta(x^+)\right) + \mathrm{sim}\left(f_\theta(x), f_\theta(x^-)\right)\right]_+$$

Support Vector Machines:

set 1 : $x^+ \to 1$

$x^- \to -1$

Soft-margin SVM solves the objective:

$$\begin{cases} \min\limits_{w,\, b,\, \xi \geq 0} \quad \frac{1}{2}\|w\|^2 + C \sum\limits_x \xi_x \\ \text{s.t.} \quad y_x \underbrace{(w^T x + b)}_{\text{linear project + bias}} \geq 1 - \underbrace{\xi_x}_{\text{margin}} \ ; \ \forall x \in X^+ \cup X^- \end{cases}$$

Hyperplane.

The Lagrangian dual:

$$\min_{0 < \alpha < C,\ \alpha^T y = 0} \quad \frac{1}{2} \alpha K(x^+, x^-) \alpha - \alpha^T 1$$

$K \in S_{++}^{|X^+ \cup X^-|}$ → symmetric positive semidefinite

→ Reproducible kernel hilbert space

$k_{ij} = y_{x_i} y_{x_j} K(x_i, x_j)$ (RKHS)

Exact decision boundary:

$$\boxed{w(\cdot) = \sum_{x \in X^+ \cup X^-} \alpha_x y_x K(x, \cdot)}$$

Proposed methodology:

Rewritten soft-constraint variant.

$$\begin{cases} \min_{\theta} \sum_{B \subset D'} \sum_{x \in B} \min_{w_x} \Big(\frac{1}{2} \|w_x\|^2 + \left[t - \langle w_x, f_\theta(x) \rangle \right]_+ \\ \qquad + \sum_{x^- \in B \setminus x} \left[t + \langle w_x, f_\theta(x^-) \rangle \right]_+ \Big) \end{cases}$$

How to optimize this guy??

Alternative formulation:

$$\min_{\theta} \mathcal{L}(\theta) := \sum_{B \in D'} \sum_{x \in B} \alpha_x^{*T} K\big(f_\theta(x), f_\theta(B \setminus x)\big) \alpha_x^*$$

$$\text{s.t.} \quad \alpha_x^* = \arg\min_{0 < \alpha < C} \frac{1}{2} \alpha^T K\big(\underset{=}{f_\theta(x)}, f_\theta(B \setminus x)\big) \alpha - \alpha^T \mathbb{1}$$

↳ so far trained f_θ

$$\alpha^T y = 0$$

$$\begin{bmatrix} K(z^+, z^+) \; , \; -K(z^+, z^-) \end{bmatrix}$$

where $\underbrace{k(z^+, z^-)}_{\text{kernel matrix}} = \left[-k(z^-, z^+) \quad , \quad k(z^-, z^-) \right]$

$$k(z, z') = \langle \phi(z), \phi(z') \rangle$$

Proposition:

solved by SVM solver. PGD / or InV

$$w(z) = \alpha_n^T \left(k(z^+, z)\mathbb{1} - k(z^-, z) \right)$$

// SVM Decision for new points

Proposed max-margin objective

$$\min_{\theta} \sum_{\substack{(x, x^+) \sim B \in D' \\ y^- = B \setminus (x, x^+)}} \alpha_x^T \left[k(f_\theta(y^-), f_\theta(x)) - \mathbb{1}\, k(f_\theta(x^+), f_\theta(x)) \right]$$