1 intriguing Properties of CL Loss

Intriguing properties of CL loss

Generalized loss function.

$$\mathcal{L}_{gen\ CL} = \mathcal{L}_{align} + \lambda \mathcal{L}_{distin}$$

$$\cancel{\tau \mathcal{L}^{NT\ xent} = -\frac{1}{N} \sum_{i,j} \text{sim}(z_i, z_j) + \frac{\tau}{n} \sum_i \log \sum_{k=1}^{2n} \mathbb{1}_{[k \neq i]} \exp\{\text{sim}(z_i, z_k)/\tau\}}$$

$$\mathcal{L}^{NT\ xent} = -\frac{1}{n} \sum_{i,j \in MB} \log \frac{\exp[\text{sim}(z_i, z_j/\tau)]}{\sum_{k=1}^{2n} \mathbb{1}_{[k \neq i]} \exp[\text{sim}(z_i, z_k)/\tau]}$$

By some calculation, log breaking λ $\boxed{\text{log sum exp}}$ normalized.

$$\tau \mathcal{L}^{NT\ xent} = \underbrace{-\frac{1}{n} \sum_{i,j} \text{sim}[\cancel{\exp}(z_i, z_j \cancel{/\tau})]}_{\text{Alignment}} + \underbrace{\frac{\tau}{n} \sum_i \log \sum_{k=1}^{2n} \mathbb{1}_{k \neq i} \exp(\text{sim}(z_i, z_k)/\tau)}_{\text{Distribution}}$$

remove it

[match hidden dist uniform in hypersphere]

[Pairwise potential of gaussian kernel.]

[minimized by perfect uniform encoder]

(B) intriguing CL Loss. (u)

Sliced Wasserstein Distance (SWD) loss

Activation vectors. $H \in \mathbb{R}^{b \times d}$, a prior distribution S

Draw prior vector $P \in \mathbb{R}^{b \times d}$ using S

Generate random orthogonal vector matrix $W \in \mathbb{R}^{d \times d'}$ $\boxed{d > d'}$??

Projection $H^{\perp} = HW$; $P^{\perp} = PW$ // rotation.

Initialize $L = 0$ // SWD

for $j \in \{1, 2, \dots d'\}$ do projection for all.

$$L = L + \| \text{sort}(H^{\perp}_{:,j}) - \text{sort}(P^{\perp}_{:,j}) \|^2$$

end for column

return $L/(d\, d')$

where it fits ??