Sinkhorn label allocation

Partially labeled dataset:

$$L := \{(x_i, y_i) \mid i = 1 \dots n_L\}$$

$$U := \{ x_i \mid i = n_L + 1 \dots n \}$$

$$n \gg n_L$$

$$x \in \mathcal{X}, \quad y \in \mathcal{Y}$$

$$\text{Goal} : f : \mathcal{X} \to \mathcal{Y}$$

given, $\mathcal{Y} = \{1, \dots K\}$ — n examples, K classes.

we seek assignment $Q \in \mathbb{R}^{n \times K}$

minimize assignment cost: $\sum_{ij} Q_{ij} \, c_{ij}(\theta)$

$c_{ij}(\theta)$ = cost of ith example to class j

$$C_{ij}(\theta) = -\log P_\theta(j \mid x_i)$$

Notation:

$$\Delta_d := \{ x \in \mathbb{R}_+^d \mid x^\top 1_d = 1 \}$$ // d simplex.

PDF property.

$$x_+ := \max(0, x)$$

$$x_- := \min(0, x)$$

For pdf, $p, q \in \Delta_d$

Entropy $H(p) = -\sum_i p_i \log p_i$

Cross entropy $h(p, q) = -\sum_i p_i \log q_i$

KL-Divergence: $D_{KL}(p \| q) := \sum_{i=1}^d p_i \log(p_i / q_i)$

$$\langle X, Y \rangle = \sum_{ij} X_{ij} Y_{ij}$$ // Frobenius inner product.

Sinkhorn-Knopp label Assignment:

① Label Assignment:

Target find soft label $q \in \mathbb{R}^K$ corresponding x.

soft → as $\{ q \in \mathbb{R}_+^K \mid q^\top 1_K \leq 1 \}$

can be less than 1

Alternative thinking: $q = \eta \tilde{q}$

$$\tilde{q} \in \Delta_K$$

weight $\eta \in [0, 1]$

Thus: $H(q, p) := -\sum_{i=1}^{K} q_i \log p_i$ → can be used as target dist.

$$= -\eta \sum_{i=1}^{K} \tilde{q}_i \log p_i$$

weighted cross entropy

Optimization problem: Linear program.

$$\underset{Q \in \mathbb{R}^{n \times k}}{\text{minimize}} \quad \langle Q, C \rangle$$

↳ non-negative loss function.

s.t.

$$Q_{ij} \geq 0$$

$$Q \mathbb{1}_k \leq \mathbb{1}_n$$

$$Q^T \mathbb{1}_n \leq \mathbb{1}_k + n b$$

↗ upper bound.

$$\mathbb{1}_n Q \mathbb{1}_k \geq n (\rho - \mu_+) - 1$$

↗ fraction of label to be allocated.

derivation:

$$\min_{Q_i \in \mathbb{1}_k} \sum_{i=1}^{n} D_{KL}(Q_i \| P_i) + f(Q_i)$$

fast approximation Solution:

$$\tilde{Q} = \text{diag}(e^{\alpha}) \, e^{-\gamma \tilde{C}} \, \text{diag}(e^{\beta})$$