Self labelling via simultaneous clustering

methods:

Network

$$x = \phi(I)$$

Image

$\in \mathbb{R}^D$ [representation]

N data points $I_1, I_2, \dots I_N$

[k labels $y_1 \dots y_N \in \{1, \dots k\}$]

classification head $h : \mathbb{R}^D \to \mathbb{R}^k$

$$\Rightarrow \quad p(y = \cdot \mid x_i) = \text{softmax}(h \circ \phi(x_i))$$

$$\Rightarrow \quad E(p \mid y_1 \dots y_N) = -\frac{1}{N} \sum_{i=1}^{N} \log p(y_i \mid x_i)$$

[where is the label??]

$\boxed{\text{Self-labeling}}$

Deterministic posterior

N K

$$E(p,q) = -\frac{1}{N}\sum_{i=1}\sum_{y=1} q(y|x_i) \log p(y|x_i)$$

Objective

Delta label

$$\min_{p,q} E(p,q) \quad \text{Subj. to} \quad \underbrace{\forall y: q(y|x_i) \in \{0,1\}}_{} \;\&\; \underbrace{\sum_{i=1}^{N} q(y|x_i) = \frac{N}{K}}_{\text{equipartition}}$$

[optimal transport problem]

Let, $P_{yi} = p(y|x)\frac{1}{N}$; $K \times N$ // joint PDF

$Q_{yi} = q(y|x)\frac{1}{N}$; $K \times N$) // Assigned joint PDF

↳ relax the Q to be transport prototype.

$\frac{1}{K}\cdot\mathbb{1}$ $\qquad\qquad$ $\frac{1}{N}\cdot\mathbb{1}$

$$U(r,c) := \left\{ Q \in \mathbb{R}_+^{K\times N} \;\middle|\; Q\mathbb{1} = r,\; Q^T\mathbb{1} = c \right\}$$

$\frac{\mathbb{1}^{N\times 1}}{K}$ $\qquad\qquad$ $\frac{\mathbb{1}^{K\times 1}}{N}$

Rewriting

Dot product

$$\underbrace{E(p, q) + \log N}_{\text{minimize}} = \langle Q, -\log P \rangle$$

$\hookrightarrow$ MI maximization

Require to Solve.

$$\min_{Q \in U(r, c)} \langle Q, -\log P \rangle$$

Sinkhorn-Knopp:

$$\min_{Q \in U(r, c)} \langle Q, -\log P \rangle + \frac{1}{\lambda} KL(Q \,\|\, rc^T)$$

Element wise

$\Rightarrow$ then $\boxed{Q = \operatorname{diag}(\alpha)\, P^{\lambda}\, \operatorname{diag}(\beta)}$

Learn: $\boxed{h \circ \phi}$ $\to$ Determines $P_{y_i} = \operatorname{softmax}(h \circ \phi(x_i))$

& a label assignment matrix $\boxed{Q}$

Two alternating approach

① Representation learning

$h \circ \phi$: cross entropy

(ii) Self-labeling — given current $h \circ \phi$, compute P

Then find $Q = \text{diag}(\alpha)\, P^{\lambda}\, \text{diag}\, \beta$

by iterative approach

$$\forall y : \alpha_y \leftarrow \left[P^{\lambda}\beta\right]_y^{-1} \; ; \; \forall i : \beta_i \leftarrow \left[\alpha^T P^{\lambda}\right]_i^{-1}$$