Learning from complementary label

Traditional risk minimization:

$$R(f) = \mathbb{E}_{p(x,y)}\left[\ell(f(x), y)\right]$$

where $f(x) = \underset{y \in \{1, \dots, k\}}{\arg\max} \; g_y(x)$

$$f(x) : \mathbb{R}^d \to \{1, \dots, k\}$$

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

$g_y(x) : \mathbb{R}^d \to \mathbb{R}$ binary OVA classifier.

$\ell(z) : \mathbb{R} \to \mathbb{R}$ loss

should be high

loss 1

$$\mathcal{L}_{OVA}(f(x), y) = \ell(g_y(x)) + \frac{1}{k-1} \sum_{y' \neq y} \ell(-g_{y'}(x))$$

should be low

loss 2

$$\mathcal{L}_{PC}(f(x), y) = \sum_{y' \neq y} \ell(g_y(x) - g_{y'}(x))$$

should have high difference.

Complementary Classifier:

Sample from $\{(x_i, \bar{y}_i)\}_{i=1}^{n}$

from Distribution

one of them is same class

$$\bar{p}(x, \bar{y}) = \frac{1}{K-1} \sum_{y \neq \bar{y}} p(x, y)$$

what & why ??

Complementary loss: $\bar{\mathcal{L}}(f(x), \bar{y})$

constant

Condition $\underbrace{\bar{\mathcal{L}}(f(x), y) + \mathcal{L}(f(x), y)}_{\text{one high other is low}} = M_2$

formulation

$$\bar{\mathcal{L}}(f(x), \bar{y}) = \frac{1}{\ldots} \sum \ell(g_y(x)) + \ell(-g_{\bar{y}}(x))$$

$$\mathcal{L}_{OVA} \quad \cdots \quad K-1 \quad \cdots$$
$$y \neq \bar{y}$$

$$\mathcal{L}_{PC}\left(f(x), \bar{y}\right) = \sum_{y \neq \bar{y}} \ell\left(g_y(x) - g_{\bar{y}}(x)\right)$$

$\Rightarrow$ SEE Difference with previous formulation