Prototypical Network ①

(P) Prototypical Network

given N labelled examples $S_N = \{(x_1, y_1) \ldots (x_n, y_n)\}$ label $\in \{0, 1, \ldots k\}$, $x_i \in \mathbb{R}^D$

$y_i \in \{1, \ldots k\}$

$S_k \rightarrow$ examples only from k (small) classes out of K (capital) classes.

Prototype $c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\phi(x_i)$ //mean vectors.

(uses later) Network parameters.

Now;

$$p_\phi(y = k \mid x) = \frac{\exp(-d(f_\phi(x), c_k))}{\sum_{k'} \exp(-d(f_\phi(x), c_{k'}))}$$

Distance function. //probabilistic approach

Target to minimize : $J = -\log p_\phi(y = k \mid x_u)$

Prototype as mixture density Estimation :

Bregman Divergence (Definition): $d_f(z, z') = f(z) - \{f(z') + \langle \nabla f(z'), z - z' \rangle\}$

(gradient) Derivative; Dot product; convex function. $\mathbb{R}^n \rightarrow \mathbb{R}$

Exponential family of Distribution

$$p_\psi(z \mid \theta) = \exp(z^T\theta - \psi(\theta) - g_\psi(z)) = \exp\left(-d_\phi(z, \cancel{\theta}\, \mu(\theta)) - g_\phi(z)\right)$$

output; Dot; extra term; cumulant function; Bregman Divergence; Any !!

params. params.

$f_\phi(x) = z$ /network embedding.

Exponential family with mixture model param: $\Gamma = \{\theta_k, \pi_k\}_{k=1}^{k'}$

$$p(z|\Gamma) = \sum_{k=1}^{K} \pi_k \, p_\psi(z|\theta_k)$$

$$= \sum_{k=1}^{K} \pi_k \exp\Big(-d_\phi\big(z, \underbrace{\mu(\theta_k)}_{c_k}\big) - \underbrace{g_\phi(z)}_{\text{constant term}}\Big)$$

prototype (low inductive bias) (just Avg)

constant term gets canceled in prob equation.

$$\Rightarrow p(y=k|z) = \frac{\pi_k \exp\big(-d_\phi(z, \mu(\theta_k))\big)}{\sum_{k'} \pi_{k'} \exp\big(-d_\phi(z, \mu(\theta_{k'}))\big)}$$

<u>Connection linear model:</u> for $d_\phi(z, z') = \|z - z'\|^2$ //Euclidean

$$-\|f_\phi(x) - c_k\|^2 = -f_\phi(x)^T f_\phi(x) \underbrace{\underbrace{+2c_k^T f_\phi(x)}_{w_k^T f_\phi(x)} \underbrace{- c_k^T c_k}_{+\,b_k}}_{\text{linear model/term.}}$$