Convolutional prototypical network:

feature extractor $f(x; \theta)$

prototypes m_{ij} ; $i \in \{1, 2, \cdots C\}$ // class

$j \in \{1, \cdots K\}$

// prototype of each class.

Prototype table $C \times K$ matrix.

<u>feed forward for prediction</u> :

$x \in$ class $\quad \underset{i=1}{\overset{C}{\arg\max}}\ g_i(x)$ $\quad \leq 0$

$$g_i(x) = -\min_{j=1}^{K} \| f(x; \theta) - m_{ij} \|_2^2$$

{ Euclidean distance

[Avoid (-) would seem minimization problem.

Trainable parameters: θ, m_{ij}

(jointly training them)

Prospective loss functions

① Minimum Classification Error: MCE (modified)

$$\mu_y = -g_y(x) + \left[\frac{1}{C-1}\sum_{j \neq y} g_j(x)^\eta\right]^{1/\eta}$$

$\eta \to \infty$ then: $\mu_y = -g_y(x) + g_m(x)$

$$\mu_y = \|f(x) - m_{yi}\|_2^2 - \|f(x) - m_{ji}\|$$

$\|f(x) - m_{yi}\|_2^2$: min from original class

$\|f(x) - m_{ji}\|$: min from other classes.

→ should be negative ideally

Loss function: $l((x, y); \theta, m)$

$$= \frac{1}{1 + e^{-\varepsilon \mu_y}}$$

(ii) margin Based loss function:

$$\ell\left((x, y) ; \theta, m\right) = \left[d\left(f(x), m_{y,i}\right) - d\left(f(x), m_{rj}\right) \right]_+$$

Adding margin m (hyper param)

$$\ell = \left[d\left(f(x), m_{y_i}\right) - d\left(f(x), m_{rj}\right) + m \right]_+$$

(m is proto)

Generalized MB loss function:

$$\ell = \left[\frac{d\left(f(x), m_{y_i}\right) - d\left(f(x), m_{rj}\right)}{d\left(f(x), m_{y_i}\right) + d\left(f(x), m_{rj}\right)} + m \right]$$

(iii) Distance based cross Entropy loss:

$$p\left(x \in m_{ij} \mid x\right) = \frac{e^{-\gamma d\left(f(x), m_{ij}\right)}}{\sum_{k=1}^{C} \sum_{l=1}^{K} e^{-\gamma d\left(f(x), m_{kl}\right)}}$$

$$p(y|x) = \sum_{j=1}^{k} p(x \in m_{yj} | x)$$

loss: $l((x,y);\theta,m) = -\log p(y|x)$

(iv) generalize cPL with prototypical loss.

Decision boundary for cPL

$$\|f - m_{ij}\|_2^2 = \|f - m_{kl}\|_2^2$$

(linear boundary)

Extra Constraint than the min distance:

$$pl((x,y);\theta,m) = \|f(x) - m_{yj}\|_2^2$$ // circle boundary

$$loss((x,y);\theta,m) = l((x,y);f,\theta) + \lambda pl((x,y);\theta,m)$$