How Classification Baseline Works for Deep Metric Learning: A Perspective of Metric Space

1. Complete proof

***Lemma* 1** Set $d(0,1) = M$. If $d(x, y)$ is a semi-metric on R that $L_m(\boldsymbol{x}, \boldsymbol{y})$ is also semi-metric on R^c and by the way:

$$L_m(\boldsymbol{l_a}, \boldsymbol{l_b}) = F(2M)$$

.

***Proof* :** By the definition of semi-metric in (1,2):

$$\begin{aligned} L_m(\boldsymbol{l_a}, \boldsymbol{l_b}) &= F\left[d(0,1) + d(1,0) + (c-2)d(0,0)\right] \\ &= F[d(0,1) + d(0,1) + 0] = F(2M) \end{aligned}$$

***Lemma* 2** If F is a non-convex function, that d is a weak-metric on R contains that $L_m(\boldsymbol{x}, \boldsymbol{y})$ is a weak-metric on R^c, d is a metric on R contains that $L_m(\boldsymbol{x}, \boldsymbol{y})$ is a metric on R^c.

***Proof* :** For any $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in R^c$, firstly shows that d is semi-metric contains that F is semi-metric:

Non-negative:

$$L_m(\boldsymbol{x}, \boldsymbol{y}) = F(\sum_{i=1}^{c} d(x_i, y_i)) \geq F(0) = 0$$

$$L_m(\boldsymbol{x}, \boldsymbol{y}) = F(\sum_{i=1}^{c} d(x_i, y_i)) = 0$$

$$\Rightarrow \sum_{i=1}^{c} d(x_i, y_i) = 0 \Rightarrow d(x_i, y_i) = 0 \quad i = 1, 2, ..., c$$

$$\Rightarrow x_i = y_i \quad i = 1, 2, ..., c \Rightarrow \boldsymbol{x} = \boldsymbol{y}$$

Symmetry:

$$L_m(\boldsymbol{x}, \boldsymbol{y}) = F(\sum_{i=1}^{c} d(x_i, y_i)) = F(\sum_{i=1}^{c} d(y_i, x_i)) = L_m(\boldsymbol{y}, \boldsymbol{x})$$

Then in unit function, non-convex is equivalent to non-negative second derivative, that is for any $a, b \in [0, +\infty)$:

$$F''(a) \leq 0$$

By differential median theorem in unit function, set $\xi_1 \in (0,a), \xi_2 \in (a, a+b)$:

$$F(a) - F(0) = aF^{'}(\xi_1)$$
$$F(a+b) - F(b) = aF^{'}(\xi_2)$$
$$F(a+b) - F(b) - F(a) + F(0) = a(F^{'}(\xi_2) - F^{'}(\xi_1)) \leq 0$$
$$\Rightarrow F(a+b) \leq F(a) + F(b)$$

Thus when d satisfies triangle inequality:

$$\begin{aligned}
&L_m(\boldsymbol{x}, \boldsymbol{y}) = \\
F(\sum_{i=1}^{c} d(x_i, y_i)) &\leq F(\sum_{i=1}^{c} d(x_i, z_i) + \sum_{i=1}^{c} d(y_i, z_i)) \\
&\leq F(\sum_{i=1}^{c} d(x_i, z_i)) + F(\sum_{i=1}^{c} d(y_i, z_i)) \\
&\leq L_m(\boldsymbol{x}, \boldsymbol{z}) + L_m(\boldsymbol{y}, \boldsymbol{z})
\end{aligned}$$

Lemma 3 If $L_m(\boldsymbol{x}, \boldsymbol{y})$ is a weak-metric with uniform point l or a metric, $\boldsymbol{a}, \boldsymbol{b} \in R^c$ are two samples with same label $l \in R^c$, that:

$$L_m(\boldsymbol{a}, \boldsymbol{b}) \leq L_m(\boldsymbol{a}, \boldsymbol{l}) + L_m(\boldsymbol{l}, \boldsymbol{b}) \leq 2\epsilon$$

Proof : By the definition of weak-metric in (1,2):

$$L_m(\boldsymbol{a}, \boldsymbol{b}) \leq L_m(\boldsymbol{a}, \boldsymbol{l}) + L_m(\boldsymbol{l}, \boldsymbol{b})$$
$$\leq 2sup_{\boldsymbol{a} \in f(P)} L_m(\boldsymbol{a}, \mathscr{L}(\boldsymbol{a})) = 2\epsilon$$

Lemma 4 If $L_m(\boldsymbol{x}, \boldsymbol{y})$ is a metric, $\boldsymbol{a}, \boldsymbol{b} \in R^c$ are two samples with different labels $l_a, l_b \in R^c$ respectively, that:

$$L_m(\boldsymbol{a}, \boldsymbol{b}) \geq F(2M) - 2\epsilon$$

Proof : By the definition of metric in (3):

$$\begin{aligned}
&L_m(\boldsymbol{a}, \boldsymbol{b}) + L_m(\boldsymbol{a}, \boldsymbol{l}_a) \geq L_m(\boldsymbol{b}, \boldsymbol{l}_a) \\
&L_m(\boldsymbol{a}, \boldsymbol{b}) \geq L_m(\boldsymbol{b}, \boldsymbol{l}_a) + L_m(\boldsymbol{b}, \boldsymbol{l}_b) - L_m(\boldsymbol{a}, \boldsymbol{l}_a) - L_m(\boldsymbol{b}, \boldsymbol{l}_b) \\
&\geq L_m(\boldsymbol{l}_a, \boldsymbol{l}_b) - 2sup_{\boldsymbol{a} \in f(P)} L_m(\boldsymbol{a}, \mathscr{L}(\boldsymbol{a})) \\
&= F(2M) - 2\epsilon.
\end{aligned}$$

Example $L^1_{CE'}$ is a weak-metric. $L^2_{CE'}$ is a metric iff.:

$$2^{p+1} - 2^{2p} \geq \frac{b}{a+1}$$

Proof : $L^1_{CE'}$ is weak-metric with label as its uniform point, that is: for any labels and any $\boldsymbol{x}, \boldsymbol{z} \in [0,1]^c$:

$$\begin{aligned}
&\sum_{i=1}^{c} log(1 - |x_i - z_i|) \geq \sum_{i=1}^{c} log(1 - |x_i|) + \sum_{i=1}^{c} log(1 - |z_i|) \\
&\Leftarrow |x_i - z_i| \leq |x_i| + |z_i| - |x_i z_i| \\
&\Leftarrow |x_i - z_i| \leq |x_i| + |z_i| - min\{|z_i|, |x_i|\} = max\{|x_i|, |z_i|\}
\end{aligned}$$

which is obvious for x_i, z_i are both positive. And:

$$
\begin{aligned}
&-\sum_{i=1}^{c} log(1-|x_i-z_i|) \leq \\
&-\sum_{i=1}^{c} log(1-|x_i-1|) - \sum_{i=1}^{c} log(1-|1-z_i|) \\
&\Leftarrow |x_i-z_i| \leq |x_i-1| + |z_i-1| - |x_i-1||1-z_i| \\
&\Leftarrow |1-x_i-(1-z_i)| \leq |1-x_i| + |1-z_i| - |1-x_i||1-z_i|
\end{aligned}
$$

which is similar to above for $1-x_i, 1-z_i$ are both positive. entry in label is only 0 or 1, thus it's down. When y is arbitrary in $[0,1]^c$, $L^1_{CE'}$ is not a metric for when x, y, z are respectively $0, 1/2, 1$ triangle inequality is broken, thus when we need a metric loss, $L^1_{CE'}$ is need, all we need prove is (set $a' = log(1+a)$):

$$
\begin{aligned}
&- log(1+a-b|x_i-z_i|^p) + a' \leq \\
&- log(1+a-b|x_i-y_i|^p) - log(1+a-b|y_i-z_i|^p) + 2a' \\
&1+a-b|x_i-z_i|^p \geq \frac{(1+a-b|x_i-y_i|^p)(1+a-b|y_i-z_i|^p)}{1+a}
\end{aligned}
$$

set $k = \frac{b}{a+1}$, it's:

$$
\Leftarrow |x_i-z_i|^p \leq |x_i-y_i|^p + |y_i-z_i|^p - k|x_i-y_i|^p|y_i-z_i|^p
$$

i. when y_i is between x_i, z_i(i.e. $x_i \leq y_i \leq z_i, z_i \leq y_i \leq x_i$ is similar), set $u = y_i - x_i, v = z_i - y_i$, that:

$$
|u+v|^p \leq |u|^p + |v|^p - k|u|^p|v|^p \quad s.t.\ 0 \leq u+v \leq 1
$$

It's easy to derive that there's an only extremum for $u = v = 1/2$:

$$
\begin{aligned}
&1 \leq (1/2)^p + (1/2)^p - k(1/2)^p(1/2)^p \\
&\Rightarrow 2^{p+1} - 2^{2p} \geq k
\end{aligned}
$$

ii. when y is out of range between x and z(i.e., $x \leq z \leq y$, or similarly $y \leq x \leq z$), set $u = z - x, v = y - z$, that:

$$
\begin{aligned}
|u|^p &\leq |u+v|^p + |v|^p - k|u+v|^p|v|^p \quad s.t.\ 0 \leq u+v \leq 1 \\
&\leq |u+v|^p + |v|^p - |v|
\end{aligned}
$$

for both k and $|u+v|$ is less than 1.