1. Probabilistic LME formulation

The probabilistic interpretation of the similarity measure \(d_w(x, u_c)\) between a sample \(x\) and a prototype \(u_c\) in embedding space is given by the equation:

\[
p(y = c|x, d = 1) = \frac{e^{d_w(x, u_c)}}{\sum_{i=1}^{C} e^{d_w(x, u_i)}} / 2\sigma^2. \tag{1}
\]

and the large margin constraint is

\[
d_w(x_i, u_{yi}) + \xi_i = 1, \quad i \in \{1 \ldots N\}, c = \{1 \ldots C\}, c \neq yi, \tag{2}
\]

The ratio between \(p(y = yi|x_i, d = 1)\) and \(p(y = c|x_i, d = 1)\), \(c \neq yi\) can be then rewritten as:

\[
\frac{p(y = yi|x_i, d = 1)}{p(y = c|x_i, d = 1)} = \frac{e^{d_w(x, u_{yi})}}{e^{d_w(x, u_i)}} / 2\sigma^2 \tag{3}
\]

while the large margin constraint is equivalent to

\[
d_w(x_i, u_{yi}) + \xi_i \geq d_w(x_i, u_c) + 1 \iff e^{d_w(x_i, u_{yi})-d_w(x_i, u_c)} \geq e^{-\xi_i} \approx e^{-\xi_{ic}} \tag{5}
\]

in the neighbourhood of 0 for \(\xi_{ic}\).

Therefore, large margin constraint for similarity measure is equivalent to the large margin constraint on the probability measure, given in the following form:

\[
\frac{p(y = yi|x_i, d = 1)}{p(y = c|x_i, d = 1)} \geq e - \xi_{ic} \tag{8}
\]

Equivalently, given the probabilistic interpretation

\[
p(d = 1|x) = \frac{1}{1 + e^{a d_w(x) + b}} \tag{9}
\]

\[
d_w^m(x) = \max_c d_w(x, u_c), \tag{10}
\]

where \(a < 0, b\) parameters, that have the following interpretation: \(-\frac{\sigma}{a}\) is the standard deviation and \(\frac{\xi}{a}\) is the mean of the logistic distribution.

Detection constraints

\[
d_w(x_i^0, u_c) \leq 1 + \xi_i^0, c = \{1, \ldots, C\} \iff \tag{11}
\]

\[
d_w^m(x_i^0) \leq 1 + \xi_i^0, \tag{12}
\]

can be reformulated as:

\[
ad_w^m(x) + b \geq a + a \xi_j^0 + b \iff \tag{13}
\]

\[
e^{a d_w^m(x) + b} \geq e^{a + a \xi_j^0 + b} \iff \tag{14}
\]

\[
p(d = 1|x) = \frac{1}{1 + e^{a d_w^m(x) + b}} \leq \frac{1}{1 + e^{a + a \xi_j^0 + b}} \approx \tag{15}
\]

\[
\approx \frac{1}{1 + e^{a + b} + \xi_j^0} \tag{16}
\]

in the neighbourhood of 0 for \(\xi_j^0\) which states that for non-object samples its probability of being an object \(p(d = 1|x) \geq \frac{1}{1 + e^{a + b}}\) is pushed towards \(\frac{1}{1 + e^{a + b}}\) as \(\xi_j^0 \to 0\). In practice, we set \(a = -8\) and \(b = 12\) in all experiments, so \(\frac{1}{1 + e^{a + b}} \ll \frac{1}{2}\), which can be seen as a margin in object - non-object case.

The minimization functional is then formulated as following:

\[
\sum_{i, c, i \neq y_i} \max(\xi_{ic}, 0) + \sum_j \max(\xi_j^0, 0) + \frac{1}{2} \lambda \|W\|_F^2 + \frac{1}{2} \gamma \|U\|_F^2 \tag{17}
\]

\[
+ \frac{1}{2} \lambda \|W\|_F^2 + \frac{1}{2} \gamma \|U\|_F^2, \tag{18}
\]
2. Full multi-prototype LME formulation

The optimization problem for learning multiple prototypes at once can be formulated as:

\[
\begin{align*}
\text{minimize:} & \quad \sum_{i,c,c\neq y_i} \max(\xi_{ic}, 0) + \sum_j \max(\xi_j^0, 0) + \\
& \quad + \frac{1}{2} \lambda \|W\|_F^2 + \frac{1}{2} \gamma \|U\|_F^2 \\
\text{subject to:} & \quad S_\alpha(W(x_i, U_y) + \xi_{ic}) \geq S_\alpha(W(x_i, U_c)) + 1, \\
& \quad i = \{1, \ldots, N\}, \quad c = \{1, \ldots, C\}, \quad c \neq y_i, \\
& \quad S_\alpha(x_j^0, U_c) \leq 1 + \xi_j^0, \\
& \quad c = \{1, \ldots, C\}, \quad j = \{1, \ldots, N_j\},
\end{align*}
\]

where the classification can be done for example:

\[
y^* = \begin{cases} 
\arg\max_c S_\alpha(x^*, U_c), & S_\alpha(x^*, U_c) \geq \tau, \\
\epsilon_0, & \forall c = 1, \ldots, C : S_\alpha(x^*, U_c) < \tau,
\end{cases}
\]