Shrinkage Expansion Adaptive Metric Learning

Qilong Wang[1,3], Wangmeng Zuo[2], Lei Zhang[3], and Peihua Li[1]

[1] School of Information and Communications Engineering, Dalian University of Technology, China
[2] School of Computer Science and Technology, Harbin Institute of Technology, China
[3] Department of Computing, Hong Kong Polytechnic University, Hong Kong
csqlwang@gmail.com, cswmzuo@gmail.com, cslzhang@comp.polyu.edu.hk, peihuali@dlut.edu.cn

Appendix

Table A.1. Summary of main symbols and abbreviations

SEAML	Shrinkage Expansion Adaptive Metric Learning
F-Norm	Frobenius norm
KKT	Karush-Kuhn-Tucker
PSD	Positive Semidefinite (matrix)
$\mathcal{S}$	The set of pairs of samples which belong to the same class
$\mathcal{D}$	The set of pairs of samples which belong to different classes
ξ_{ij}	The soft penalty on the pairwise inequality constraint

1 The Lagrange Dual of SEAML with Squared F-Norm Regularizer

We adopt the squared F-Norm regularizer and hinge loss function in the proposed SEAML framework, resulting in the following metric learning model:

$$\begin{aligned} &\min_{\mathbf{M},\boldsymbol{\xi}} \tfrac{1}{2}\|\mathbf{M}\|_F^2 + C\sum\nolimits_{ij}\xi_{ij} \\ &\text{s.t. } \; D_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j) \le f_s(D_{ij}) + \xi_{ij} \;\; (\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}, \\ &\qquad D_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j) \ge f_d(D_{ij}) - \xi_{ij} \;\; (\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}, \\ &\qquad \xi_{ij}\ge 0, \forall(i,j),\; \mathbf{M}\succcurlyeq 0. \end{aligned} \tag{1}$$

The original problem of SEAML can be rewritten as follows:

$$\begin{aligned} &\min_{\mathbf{M},\boldsymbol{\xi}} \tfrac{1}{2}\|\mathbf{M}\|_F^2 + C\sum\nolimits_{ij}\xi_{ij} \\ &\text{s.t. } \; l_{ij}D_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j) \ge f(D_{ij},l_{ij}) - \xi_{ij}, \\ &\qquad \xi_{ij}\ge 0, \forall(i,j),\; \mathbf{M}\succcurlyeq 0. \end{aligned} \tag{2}$$

where $l_{i,j} = -1$, $f(D_{ij}, l_{i,j}) = -f_s(D_{ij})$ if $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}$ and $l_{i,j} = 1$, $f(D_{ij}, l_{i,j}) = f_d(D_{ij})$ if $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}$. For simplicity, hereafter $f(D_{ij}, l_{i,j})$ is abbreviated as f_{ij}.

Let $\mathbf{Z}_{ij} = (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T$. The Mahalanobis distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ is: $D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{Z}_{ij}, \mathbf{M} \rangle = tr(\mathbf{Z}_{ij}\mathbf{M})$. Then, the Lagrangian of Eq. (2) can be expressed as follows:

$$
\begin{aligned}
&L(\mathbf{M}, \mathbf{Y}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\xi}) = \tfrac{1}{2}\|\mathbf{M}\|_F^2 + C\sum\nolimits_{ij} \xi_{ij} - \langle \mathbf{Y}, \mathbf{M} \rangle \\
&- \sum\nolimits_{ij} \beta_{ij}[l_{ij}\langle \mathbf{Z}_{ij}, \mathbf{M}\rangle - f_{ij} + \xi_{ij}] - \sum\nolimits_{ij} \gamma_{ij}\xi_{ij} \\
&\text{s.t.}\ \ \beta_{ij} \geq 0,\ \gamma_{ij} \geq 0, \forall(i,j),\ \mathbf{Y} \succcurlyeq 0.
\end{aligned} \tag{3}
$$

where $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\mathbf{Y}$ are the Lagrange multipliers. In order to obtain the dual of the original problem in Eq. (2), the following KKT conditions should be satisfied:

$$
\begin{aligned}
&\frac{L(\mathbf{M}, \mathbf{Y}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\xi})}{\xi_{ij}} = 0 \Rightarrow C - \beta_{ij} - \gamma_{ij} = 0, \forall(i,j) \\
&\Rightarrow 0 \leq \beta_{ij} \leq C, \forall(i,j).
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
&\frac{L(\mathbf{M}, \mathbf{Y}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\xi})}{\mathbf{M}} = 0 \Rightarrow \mathbf{M} - \sum\nolimits_{ij} \beta_{ij} l_{ij} \mathbf{Z}_{ij} - \mathbf{Y} = 0 \\
&\Rightarrow \mathbf{M} = \sum\nolimits_{ij} \beta_{ij} l_{ij} \mathbf{Z}_{ij} + \mathbf{Y}.
\end{aligned} \tag{5}
$$

Substituting Eq. (4) into the Lagrangian in Eq. (3), we get the following formula:

$$
\begin{aligned}
L(\mathbf{M}, \mathbf{Y}, \boldsymbol{\beta}) = \tfrac{1}{2}\|\mathbf{M}\|_F^2 - \langle \mathbf{Y}, \mathbf{M} \rangle - \sum\nolimits_{ij} \beta_{ij} l_{ij} \langle \mathbf{Z}_{ij}, \mathbf{M} \rangle + \sum\nolimits_{ij} \beta_{ij} f_{ij} \\
\text{s.t.}\ \ \beta_{ij} \geq 0, \forall(i,j),\ \mathbf{Y} \succcurlyeq 0.
\end{aligned} \tag{6}
$$

Finally, we substitute Eq. (5) into Eq. (6), obtaining the Lagrange dual problem of Eq. (2):

$$
\begin{aligned}
\max\nolimits_{\mathbf{Y}, \boldsymbol{\beta}} -\tfrac{1}{2}\|\sum\nolimits_{ij} \beta_{ij} l_{ij} \mathbf{Z}_{ij} + \mathbf{Y}\|_F^2 + \sum\nolimits_{ij} f_{ij}\beta_{ij} \\
\text{s.t.}\ \ 0 \leq \beta_{ij} \leq C, \forall(i,j),\ \mathbf{Y} \succcurlyeq 0.
\end{aligned} \tag{7}
$$

The problem in Eq. (7) involves the joint optimization of PSD matrix $\mathbf{Y}$ and vector $\boldsymbol{\beta}$, which can be solved by using the alternative optimization algorithm summarized in Algorithm 1 of the main paper.