Supplementary Material for "A General Regularization Framework for Domain Adaptation"

Wei Lu[1] and **Hai Leong Chieu**[2] and **Jonathan Löfgren**[3]
[1]Singapore University of Technology and Design
[2]DSO National Laboratories
[3]Uppsala University
luwei@sutd.edu.sg, chaileon@dso.org.sg, lofgren021@gmail.com

Abstract

This is the supplementary material for (Lu et al., 2016). Here we present a detailed proof of Lemma 2.2 and give details to show how we arrive at Equation 7 in the main paper.

1 Proof of Lemma 2.2

Lemma 2.2 *For any vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N \in \mathbf{R}^m$*, any scalars* $\lambda_0, \lambda_1, \ldots, \lambda_N \in \mathbf{R}^+$*, let* $\mathbf{v}_0 = (\sum_{i=1}^N \lambda_i \mathbf{v}_i)/\lambda_0$*, then the following always holds:*

$$\lambda_0 ||\mathbf{v}_0||^2 + \sum_{i=1}^N \lambda_i ||\mathbf{v}_i||^2$$
$$= \sum_{i=1}^N \eta_{0,i} ||\mathbf{v}_i + \mathbf{v}_0||^2 + \sum_{1 \le j < k \le N} \eta_{j,k} ||\mathbf{v}_j - \mathbf{v}_k||^2$$

where

$$\eta_{i,j} = \frac{\lambda_i \lambda_j}{\sum_{l=0}^N \lambda_l}$$

for all $0 \le i < j \le N$*.*

Proof First note that $\sum_{j=0}^N \eta_{j,k} = \lambda_k$ and since $\lambda_0 \mathbf{v}_0 = \sum_{j=1}^N \lambda_i \mathbf{v}_j$ we also have

$$\sum_{j=1}^N \sum_{k=1}^N \eta_{j,k} \mathbf{v}_j \cdot \mathbf{v}_k = \sum_{i=1}^N \eta_{0,i} \mathbf{v}_j \cdot \mathbf{v}_0 = \eta_{0,0} \left\| \mathbf{v}_0 \right\|^2.$$

Although we formally specify $i < j$ for the parameters $\eta_{j,k}$ we wil relax this for now to simplify the notation. We denote the terms on the right hand in original equation by A and B respectively and expand them as follows:

$$\begin{aligned}
A &= \sum_{i=1}^N \eta_{0,i} \left\| \mathbf{v}_i + \mathbf{v}_0 \right\|^2 \\
&= \sum_{i=1}^N \eta_{0,i} \left\| \mathbf{v}_i \right\|^2 + \sum_{i=1}^N \eta_{0,i} \left\| \mathbf{v}_0 \right\|^2 \\
&+ 2 \sum_{i=1}^N \eta_{0,i} \mathbf{v}_i \cdot \mathbf{v}_0 \\
&= \sum_{i=1}^N \eta_{0,i} \left\| \mathbf{v}_i \right\|^2 + (\lambda_0 - \eta_{0,0}) \left\| \mathbf{v}_0 \right\|^2 \\
&+ 2\eta_{0,0} \left\| \mathbf{v}_0 \right\|^2 \\
&= \sum_{i=1}^N \eta_{0,i} \left\| \mathbf{v}_i \right\|^2 + \lambda_0 \left\| \mathbf{v}_0 \right\|^2 + \eta_{0,0} \left\| \mathbf{v}_0 \right\|^2 \\
B &= \sum_{1 \le j < k \le N} \eta_{j,k} \left\| \mathbf{v}_j - \mathbf{v}_k \right\|^2 \\
&= \frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \eta_{j,k} \left\| \mathbf{v}_j - \mathbf{v}_k \right\|^2 \\
&= \sum_{j=1}^N \sum_{k=1}^N \eta_{j,k} \left\| \mathbf{v}_j \right\|^2 - \sum_{j=1}^N \sum_{k=1}^N \eta_{j,k} \mathbf{v}_j \cdot \mathbf{v}_k \\
&= \sum_{j=1}^N \left(\sum_{k=0}^N \eta_{j,k} \left\| \mathbf{v}_j \right\|^2 - \eta_{j,0} \left\| \mathbf{v} j \right\|^2 \right) \\
&- \eta_{0,0} \left\| \mathbf{v}_0 \right\|^2 \\
&= \sum_{j=1}^N \lambda_j \left\| \mathbf{v}_j \right\|^2 - \sum_{j=1}^N \eta_{0,j} \left\| \mathbf{v}_j \right\|^2 - \eta_{0,0} \left\| \mathbf{v}_0 \right\|^2
\end{aligned}$$

The full right hand side is $A + B$ and cancelling the terms gives us

$$\mathbf{RHS} = A + B = \lambda_0 \left\|\mathbf{v}_0\right\|^2 + \sum_{i=1}^{N} \lambda_i \left\|\mathbf{v}_i\right\|^2 = \mathbf{LHS}. \quad \blacksquare$$

2 Proof of Equation 7

In the main paper, we defined

$$\mathbf{w}_0' = \frac{1}{\sum_{l=0}^{N} \lambda_l} \sum_{i=1}^{N} \lambda_i \mathbf{w}_i^*, \tag{1}$$

and we also defined

$$\mathbf{w}_i' = \mathbf{w}_i^* - \mathbf{w}_0' \tag{2}$$

.

We need to prove the following equation (Equation 7 in the main paper):

$$\mathbf{w}_0' = \left(\sum_{i=1}^{N} \lambda_i \mathbf{w}_i'\right) / \lambda_0 \tag{3}$$

Proof Since

$$\mathbf{w}_0' = \frac{1}{\sum_{l=0}^{N} \lambda_l} \sum_{i=1}^{N} \lambda_i \mathbf{w}_i^*, \tag{4}$$

we have:

$$\sum_{i=1}^{N} \lambda_i \mathbf{w}_i^* = \left(\sum_{l=0}^{N} \lambda_l\right) \mathbf{w}_0' \tag{5}$$

Since

$$\mathbf{w}_i' = \mathbf{w}_i^* - \mathbf{w}_0', \tag{6}$$

we have:

$$\begin{aligned} \sum_{i=1}^{N} \lambda_i \mathbf{w}_i' &= \sum_{i=1}^{N} \lambda_i \left(\mathbf{w}_i^* - \mathbf{w}_0'\right) \\ &= \sum_{i=1}^{N} \left(\lambda_i \mathbf{w}_i^* - \lambda_i \mathbf{w}_0'\right) \\ &= \sum_{i=1}^{N} \lambda_i \mathbf{w}_i^* - \sum_{i=1}^{N} \lambda_i \mathbf{w}_0' \end{aligned} \tag{7}$$

Based on Equation 5 above, we have:

$$\begin{aligned} \sum_{i=1}^{N} \lambda_i \mathbf{w}_i' &= \sum_{i=1}^{N} \lambda_i \mathbf{w}_i^* - \sum_{i=1}^{N} \lambda_i \mathbf{w}_0' \\ &= \left(\sum_{l=0}^{N} \lambda_l\right) \mathbf{w}_0' - \left(\sum_{i=1}^{N} \lambda_i\right) \mathbf{w}_0' \\ &= \lambda_0 \mathbf{w}_0' \end{aligned} \tag{8}$$

Therefore, we have:

$$\mathbf{w}_0' = \left(\sum_{i=1}^{N} \lambda_i \mathbf{w}_i'\right) / \lambda_0 \quad \blacksquare \tag{9}$$

References

Wei Lu, Hai Leong Chieu, and Jonathan Löfgren. 2016. A general regularization framework for domain adaptation. In *In Proceedings of EMNLP*.