Support Vector Guided Dictionary Learning

Sijia Cai[1,3], Wangmeng Zuo[2], Lei Zhang[3⋆], Xiangchu Feng[4], and Ping Wang[1]

[1]School of Science, Tianjin University
[2]School of Computer Science and Technology, Harbin Institute of Technology
[3]Dept. of Computing, The Hong Kong Polytechnic University
[4]Dept. of Applied Mathematics, Xidian University
cssjcai@gmail.com,cslzhang@comp.polyu.edu.hk

Appendix A: Proof of Lemma 1

Lemma 1. *Denote by $\bar{z}_c$ and $\bar{z}$ the mean vectors of Z_c and Z, respectively, where Z_c is the set of coding vectors of samples from class c. Then $\mathcal{L}(Z)$ in FDDL is equivalent to the weighted sum of the squared distances of pairs of coding vectors:*

$$\mathcal{L}(Z) = \sum_{c=1}^{C}\Big(\sum_{y_i=c,y_j=c}\big(\frac{1}{n_c}-\frac{1}{2n}\big)\|z_i-z_j\|_2^2 + \sum_{y_i=c,y_j\neq c}\big(-\frac{1}{2n}\big)\|z_i-z_j\|_2^2\Big).$$

Proof: The Fisher discrimination criterion $\mathcal{L}(Z)$ adopted in FDDL [18] can be formulated into the following equivalent form:

$$\begin{aligned}
\mathcal{L}(Z) &= tr(S_W(Z)) - tr(S_B(Z)) \\
&= \sum_{c=1}^{C}\sum_{y_i=c}\|z_i-\bar{z}_c\|_2^2 - \sum_{c=1}^{C} n_c\|\bar{z}_c-\bar{z}\|_2^2 \\
&= \sum_{c=1}^{C}\sum_{y_i=c}(\|z_i\|_2^2 - 2z_i^T\bar{z}_c + \|\bar{z}_c\|_2^2) - \sum_{c=1}^{C} n_c(\|\bar{z}_c\|_2^2 - 2\bar{z}_c^T\bar{z} + \|\bar{z}\|_2^2) \\
&= \sum_{c=1}^{C}\Big(\sum_{y_i=c}\|z_i\|_2^2 - 2\big(\sum_{y_i=c}z_i\big)^T\bar{z}_c + n_c\|\bar{z}_c\|_2^2\Big) - \Big(\sum_{c=1}^{C}n_c\|\bar{z}_c\|_2^2 - 2\big(\sum_{c=1}^{C}n_c\bar{z}_c\big)^T\bar{z} + n\|\bar{z}\|_2^2\Big) \\
&= \sum_{c=1}^{C}\Big(\sum_{y_i=c}\|z_i\|_2^2 - n_c\|\bar{z}_c\|_2^2\Big) - \Big(\sum_{c=1}^{C}n_c\|\bar{z}_c\|_2^2 - n\|\bar{z}\|_2^2\Big) \\
&= 2\sum_{c=1}^{C}\Big(\sum_{y_i=c}\|z_i\|_2^2 - n_c\|\bar{z}_c\|_2^2\Big) - \Big(\sum_{c=1}^{C}\sum_{y_i=c}\|z_i\|_2^2 - n\|\bar{z}\|_2^2\Big) \\
&= 2\sum_{c=1}^{C}\Big(\sum_{y_i=c}\|z_i\|_2^2 - n_c\|\bar{z}_c\|_2^2\Big) - \Big(\sum_{i}\|z_i\|_2^2 - n\|\bar{z}\|_2^2\Big).
\end{aligned}$$

⋆ Corresponding author.

It is easy to see that:

$$\begin{aligned}\sum_{i,j}\frac{1}{2n}\|z_i-z_j\|_2^2 &= \frac{1}{2n}\sum_{i,j}(\|z_i\|_2^2 - 2z_i^Tz_j + \|z_j\|_2^2)\\ &= \frac{1}{2n}(2n\sum_i\|z_i\|_2^2 - 2(\sum_i z_i)^T(\sum_j z_j))\\ &= \frac{1}{2n}(2n\sum_i\|z_i\|_2^2 - 2n^2\|\bar{z}\|_2^2)\\ &= \sum_i\|z_i\|_2^2 - n\|\bar{z}\|_2^2.\end{aligned}$$

The above equation can also be applied to each class c, i.e.,

$$\sum_{y_i=c,y_j=c}\frac{1}{2n_c}\|z_i-z_j\|_2^2 = \sum_{y_i=c}\|z_i\|_2^2 - n_c\|\bar{z}_c\|_2^2.$$

Finally, the Fisher discrimination criterion $\mathcal{L}(Z)$ can be written as:

$$\begin{aligned}\mathcal{L}(Z) &= \sum_{c=1}^{C}\sum_{y_i=c,y_j=c}\frac{1}{n_c}\|z_i-z_j\|_2^2 - \sum_{i,j}\frac{1}{2n}\|z_i-z_j\|_2^2\\ &= \sum_{c=1}^{C}(\sum_{y_i=c,y_j=c}(\frac{1}{n_c}-\frac{1}{2n})\|z_i-z_j\|_2^2 + \sum_{y_i=c,y_j\neq c}(-\frac{1}{2n})\|z_i-z_j\|_2^2).\end{aligned}$$

The proof is completed.

Appendix B: Proof of Lemma 2

Lemma 2. *Let $w_{ij}(\beta) = y_iy_j\beta_i\beta_j$. If $\sum_{j=1}^n y_j\beta_j = 0$, then the discrimination term $\mathcal{L}(Z)$ can be written as:*

$$\mathcal{L}(Z, w_{ij}(\beta)) = -2\sum_{i,j}y_iy_j\beta_i\beta_jz_i^Tz_j = \beta^TK\beta,$$

where K is the negative semidefinite matrix.

Proof: $w_{ij}(\beta) = y_iy_j\beta_i\beta_j$ is a specific parameterization of w_{ij}. To satisfy the balance property, we let $\sum_{j=1}^n y_j\beta_j = 0$, which leads to:

$$\begin{aligned}\mathcal{L}(Z, w_{ij}(\beta)) &= \sum_{i,j}\|z_i-z_j\|_2^2w_{ij}(\beta)\\ &= \sum_{i,j}y_iy_j\beta_i\beta_j(\|z_i\|_2^2 + \|z_j\|_2^2 - 2z_i^Tz_j)\\ &= -2\sum_{i,j}y_iy_j\beta_i\beta_jz_i^Tz_j\\ &= \beta^TK\beta,\end{aligned}$$

where $K = -2(Z * diag(y))^T(Z * diag(y))$ and $diag(y)$ is the diagonalization of class label vector $y = [y_1, y_2, \ldots, y_n]$. According to the Cholesky decomposition, we can see that K is a negative semidefinite matrix, leading the conclusion of **Lemma 2**.