Low-rank geometric mean metric learning

Mukul Bhutani[1] **Pratik Jawanpuria**[2] **Hiroyuki Kasai**[3] **Bamdev Mishra**[2]

Introduction. We propose a scalable solution for the *Mahalanobis* metric learning problem (Kulis, 2012). The Mahalanobis distance is defined as $d_{\mathbf{A}}(\boldsymbol{x}, \boldsymbol{x'}) = (\boldsymbol{x} - \boldsymbol{x'})^\top \mathbf{A}(\boldsymbol{x} - \boldsymbol{x'})$, where $\boldsymbol{x}, \boldsymbol{x'} \in \mathbb{R}^d$ are input vectors and $\mathbf{A}$ is a $d \times d$ symmetric positive definite (SPD) matrix. The objective is to learn a suitable SPD matrix $\mathbf{A}$ from the given data. Since $\mathbf{A}$ is a $d \times d$ SPD matrix, most state-of-the-art metric learning algorithms scale poorly with the number of features d (Harandi et al., 2017). To mitigate this, a pre-processing step of dimensionality reduction (e.g., by PCA) is generally applied before using popular algorithms like LMNN and ITML (Weinberger & Saul, 2009; Davis et al., 2007).

Recently, (Zadeh et al., 2016) proposed the geometric mean metric learning (GMML) formulation, which enjoys a closed-form solution. However, it requires matrix $\mathbf{A}$ to be positive definite, which makes it unscalable in a high dimensional setting. To alleviate this concern, we propose a low-rank decomposition of $\mathbf{A}$ in the GMML setting. Low-rank constraint also has a natural interpretation in the metric learning setting, since the group of similar points in the given dataset reside in a low-dimensional subspace. We jointly learn the low-dimensional subspace along with the metric. We show that the optimization is on the Grassmann manifold and propose a computationally efficient algorithm. On real-world datasets, we achieve competitive results comparable with the GMML algorithm, even though we work on a smaller dimensional space.

Problem formulation. We follow a weekly supervised approach in which we are provided two sets $\mathcal{S}$ and $\mathcal{D}$ containing pairs of input points belonging to same and different classes respectively. Taking inspiration from GMML, we formulate the objective function as:

$$\begin{array}{ll} \min\limits_{\mathbf{A} \succeq 0} & \mathrm{Tr}(\mathbf{A}\mathbf{S}) + \mathrm{Tr}(\mathbf{A}^\dagger \mathbf{D}) \\ \text{subject to} & \mathrm{rank}(\mathbf{A}) = r, \end{array} \tag{1}$$

where $\mathbf{S} := \sum_{(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{S}} (\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^\top$, $\mathbf{D} := \sum_{(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{D}} (\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^\top$, and $\mathbf{A}^\dagger$ is the pseudoinverse of $\mathbf{A}$.

Exploiting a particular fixed-rank factorization (Meyer et al., 2011), we factorize rank-r matrix $\mathbf{A}$ as $\mathbf{A} = \mathbf{U}\mathbf{B}\mathbf{U}^\top$, where $\mathbf{U}$ is an orthonormal matrix of size $d \times r$ and $\mathbf{B} \succ 0$ is of size $r \times r$. Consequently, we rewrite (1) as:

$$\min_{\mathbf{U}^\top \mathbf{U} = \mathbf{I}} \min_{\mathbf{B} \succ 0} \quad \mathrm{Tr}(\mathbf{U}\mathbf{B}\mathbf{U}^\top \mathbf{S}) + \mathrm{Tr}(\mathbf{U}\mathbf{B}^{-1}\mathbf{U}^\top \mathbf{D}). \tag{2}$$

If we define $\widetilde{\mathbf{S}} = \mathbf{U}^\top \mathbf{S}\mathbf{U}$ and $\widetilde{\mathbf{D}} = \mathbf{U}^\top \mathbf{D}\mathbf{U}$, then the inner minimization problem has a closed-form solution as the geometric mean of $\widetilde{\mathbf{S}}^{-1}$ and $\widetilde{\mathbf{D}}$ (Zadeh et al., 2016). Using this fact, the outer optimization problem is readily checked to be only on the column space of $\mathbf{U}$. The set of column spaces is the abstract Grassmann manifold, which is defined as the set of r-dimensional subspaces in $\mathbb{R}^d$. Equivalently, (2) is an optimization problem on the Grassmann manifold.

Extending the idea to a setting which weighs the sets $\mathcal{S}$ and $\mathcal{D}$ unequally, we obtain the formulation

$$\begin{aligned} \min_{\mathbf{U}^\top \mathbf{U} = \mathbf{I}} \min_{\mathbf{B} \succ 0} \quad & (1-t)\delta_R^2(\mathbf{B}, (\mathbf{U}^\top \mathbf{S}\mathbf{U})^{-1}) \\ & + t\delta_R^2(\mathbf{B}, \mathbf{U}^\top \mathbf{D}\mathbf{U}), \end{aligned} \tag{3}$$

where δ_R denotes the Riemannian distance on the SPD manifold and $t \in [0, 1]$ is a hyperparameter. Similarly to (2), the problem (3) is also on the Grassmann manifold as the inner problem has a closed-form solution as the weighted geometric mean between $\widetilde{\mathbf{S}}^{-1}$ and $\widetilde{\mathbf{D}}$.

Results. Our proposed algorithm LR-GMML is implemented using the off-the-shelf conjugate gradients solver of Manopt (Boumal et al., 2014). The codes are available at `https://github.com/muk343/LR-GMML`. We compare LR-GMML with GMML on publicly available UCI datasets by measuring the classification error for a k-NN classifier following the procedure in (Zadeh et al., 2016). Parameter t is optimized for both the algorithms and average errors over five random runs are reported in Figure 1.

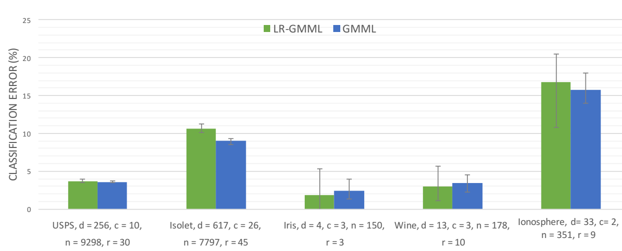

Figure 1. Classification error rates of k-NN classifier comparing LR-GMML with GMML. We obtain comparable performance in lower ranks.

[1]Indian Institute of Science. [2]Microsoft, India. [3]The University of Electro-Communications, Japan. Correspondence to: Mukul Bhutani <mukul.bhutani93@gmail.com>.

References

Boumal, N., Mishra, B., Absil, P.-A., and Sepulchre, R. Manopt, a matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15:1455–1459, 2014.

Davis, J. V., Kulis, B., Jain, P., Sra, S., and Dhillon, I. S. Information-theoretic metric learning. In *International Conference on Machine Learning (ICML)*, pp. 209–216, 2007.

Harandi, M., Salzmann, M., and Hartley, R. Joint dimensionality reduction and metric learning: A geometric take. In *International Conference on Machine Learning (ICML)*, pp. 1404–1413, 2017.

Kulis, B. Metric learning: A survey. *Foundations and Trends in Machine Learning*, 5, 01 2012.

Meyer, G., Bonnabel, S., and Sepulchre, R. Regression on fixed-rank positive semidefinite matrices: a Riemannian approach. *Journal of Machine Learning Research*, 12: 593–625, 2011.

Weinberger, K. Q. and Saul, L. K. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10:207–244, 2009.

Zadeh, P. H., Hosseini, R., and Sra, S. Geometric mean metric learning. In *International Conference on Machine Learning (ICML)*, pp. 2464–2471, 2016.