2020

L1-norm Regularized L1-norm Best-fit line problem

Xiao Ling
Virginia Commonwealth University

Paul Brooks

The L_1-Norm Regularized L_1-Norm Best-Fit Line Problem

Paul Brooks, Xiao Ling
Department of Supply Chain Management and Analytics, Department of Statistical Sciences and Operations Research

Introduction

Subspace estimation can be used for dimension reduction by projecting data in a high dimensional feature space to a low dimensional subspace. It sheds light on a board range of tasks from computer vision to pattern recognition. Conventional Principal Component Analysis (PCA) is a widely used technique to fit the subspace. PCA finds linear combinations of the original features capturing maximal variance of data via Singular Value Decomposition (SVD). However, SVD is sensitive to outliers. Some Robust PCA procedures, subspaces are fitted to the data using the L_1 criterion. In our work, we are using same criterion to fit subspace. Additionally, we apply an L_1-norm regularization in order to achieve a sparser solution.

Methodology

We describe a method to fit a lower-dimensional subspace by approximating a non-linear, non-convex, non-smooth optimization problem called L_1-Norm Regularized L_1-Norm Best-Fit line problem. The procedure can be performed using ratios and sorting. Also we present applications in the area of video analytics.

Problem Formulation

Consider the optimization problem to find L_1-Norm Regularized L_1-Norm Best-Fit Line given data $x_i \in \mathbb{R}^m$, $i = 1,\ldots,n$.

$$\min_{v,\alpha} \sum_{i=1}^{n} \|x_i - v\alpha_i\|_1 + \lambda \|v\|_1 \qquad (1)$$

The vector v determines the line through the origin. The α_i are the scaling factors that locate the projection of each point on the fitted line. The λ is a positive penalty on the L_1 norm of v.

By introducing four sets of goal variables and preserving one of the coordinates $\hat{j}$, the optimization problem in (1) can be recast as the following constrained linear programming:

$$\min_{v,\epsilon,\zeta,v_{\hat{j}}=1} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\epsilon_{ij}^{+} + \epsilon_{ij}^{-}\right) + \lambda \sum_{j=1}^{m} \left(\zeta_j^{+} + \zeta_j^{-}\right) \qquad (2)$$

Subject to:

$$v_j x_{i\hat{j}} + \epsilon_{ij}^{+} - \epsilon_{ij}^{-} = x_{ij}, i = 1,\ldots,n, j = 1,\ldots,m, j \neq \hat{j}$$

$$v_j + \zeta_j^{+} - \zeta_j^{-} = 0, j = 1,\ldots,m$$

$$\epsilon_{ij}^{+}, \epsilon_{ij}^{-}, \zeta_j^{+}, \zeta_j^{-} \geq 0, , i = 1,\ldots,n, j = 1,\ldots,m$$

Procedures and Algorithm

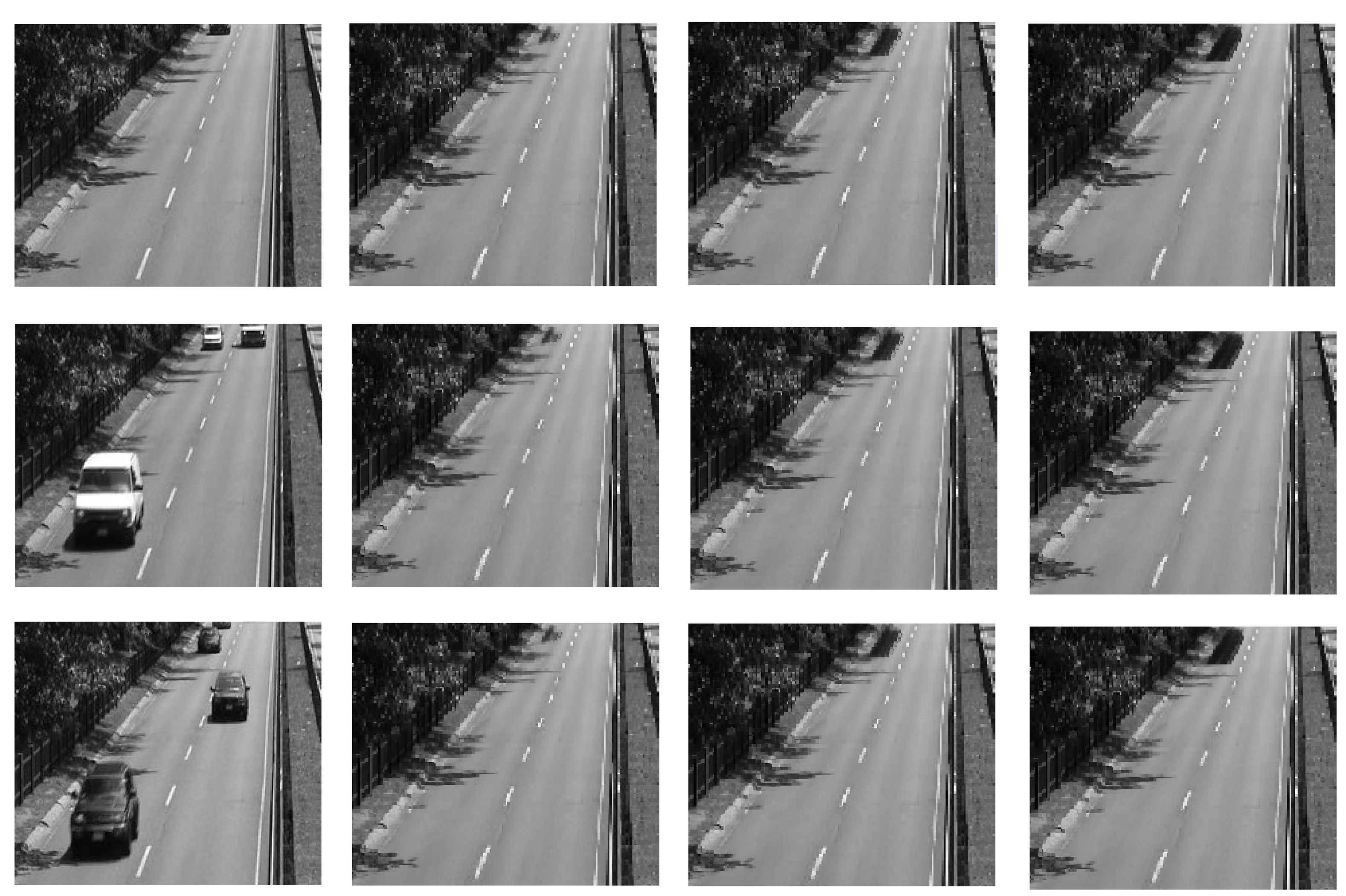
An optimal solution to (2) can be constructed as follows. If $x_{\tilde{i}\hat{j}} = 0$ for all i, then set $v = 0$. Otherwise, for each $j \neq \hat{j}$,

- Take points x_i, such that $x_{i\hat{j}} \neq 0$ and sort the ratios $\frac{\{x_{ij}\}}{\{x_{i\hat{j}}\}}$ in increasing order,
- If there is an $\tilde{i}$ where $\left|\operatorname{sgn}\left(\frac{x_{ij}}{x_{i\hat{j}}}\right)\lambda + \sum_{i<\tilde{i}}|x_{i\hat{j}}| - \sum_{i>\tilde{i}}|x_{i\hat{j}}|\right| \leq |x_{\tilde{i}\hat{j}}|$, then set $v_j = \frac{x_{ij}}{x_{\tilde{i}\hat{j}}}$,
- If no such $\tilde{i}$ exists, then set $v_j = 0$.

Algorithm

Given points $x_i \in \mathbb{R}^m$ for $i = 1,\ldots,n$.
1: Set $z^* = j^* = \infty$.
2: for ($\hat{j} = 1; \hat{j} = \hat{j}$+1) do
3: Solve LP in 2.
4: if $z_{\hat{j}} < z^*$, then
5: Set $z^* = z_{\hat{j}}; j^* = \hat{j}$.
6: end if
7: end for

Background Subtraction Application



Original λ=0 λ=8 λ=14

Figure 1: We show three frames of a highway video in the first column. The background images for these frames extracted with different λs are shown in the rest of columns. Larger λ tends to produce more sparser solution, in another word causes more black pixel area.

Results

We compared our performance with SVD on synthetic data. The numerical results showed our algorithm successfully found a better principal component from a grossly corrupted data than SVD in terms of discordance. Moreover, our algorithm provided a sparser principal component than SVD. However, we expect it to be faster on multi-node environment.

Data(n×m)		*Outliers*	*Measures*	*PCA*	*Sparse L_1*	
1000	100	0	L_0	100	100	96.8
			discordance	3.2×10^{-5}	3.0×10^{-5}	3.6×10^{-4}
1000	100	100	L_0	100	100	97.2
			discordance	0.93	3.3×10^{-5}	4.9×10^{-4}
10000	100	0	L_0	100	100	97.2
			discordance	2.8×10^{-6}	6.4×10^{-6}	3.3×10^{-4}
10000	100	1000	L_0	100	100	96.2
			discordance	0.83	6.8×10^{-6}	4.6×10^{-4}

Table 1: Mean L_0 is the number of non-zero components in solutions. Mean discordance between the true line v and the fitted line v*, measured as $1-v^Tv^*$.

Conclusions

This work proposes a new algorithm able to estimate a best-fit line as efficiently as several sorting of data. When subspaces are projected from contaminated data, out algorithm can achieve arguably smaller discordance and less number of non-zero components solution than that of traditional PCA. This demonstrates that our solution is robust and sparser than that of traditional PCA.

References

- Brooks, J. P., Dula, J. H., and Boone, E. L. (2013). A pure L1-norm principal component analysis. Computational Statistics and Data Analysis, 61:83-98.
- Brooks, J. P. and Dula, J. H. (2017). Estimating L1-Norm Best-Fit Lines for Data.
- Chierichetti, F., Gollapudi, S., Kumar, R., Lattanzi, S., Panigrahy, R., and Woodru, D. P. (2017). Algorithms for L_p low-rank approximation. In Proceedings of the 34th International Conference on Machine Learning-Volume 70, pages 806-814. JMLR. org.
- Markopoulos, P. P., Dhanaraj, M., and Savakis, A. (2018). Adaptive L1-Norm Principal-Component Analysis With Online Outlier Rejection. IEEE Journal on Selected Topics in Signal Processing, 12(6):1131-1143.