Bayesian manifold estimation

Busovikov Vladimir
Skoltech, MIPT

1. Introduction

In the last twenty years, there are several novel nonlinear dimension reduction procedures appeared, such as Isomap (Tenenbaum, de Silva and Langford, 2000), LLE (Roweis and Saul, 2000) and its modification (Zhang and Wang, 2007), Laplacian eigenmaps (Belkin and Niyogi, 2003), and t-SNE (van der Maaten and Hinton, 2008).

Besides its worth mention some recent works using geometric multi retsolution analysis (Maggioni, Minsker and Strawn, 2016), local polynomial estimators (Aamari and Levrard, 2019) and numerical solution of PDE (Shi and Sun,2017).

There are some methods which allows to handle with large noise, based on an optimization problem, such as mean-shift (Cheng, 1995) and its variants (Ozertem and Erdogmus, 2011), (Genovese et al., 2014).

We now introduce new method based on tangent space estimation, which is able to handle with large-amplitude noise. All theoretical results are now in process.

2. Background and problem statement

Suppose we have data $\{X_n\}_{n=1}^N \subset \mathbb{R}^D$ sampled from uniform distribution on compact smooth manifold without boundaty $\mathcal{M}$. Let $\mathcal{M}$ have dimension $d < D$. Now let ε be Gaussian unbiased noise, s.t. $\varepsilon \mid X$ has normal distribution and $\mathbb{E}\left(\varepsilon \mid X\right) = 0$. We observe noised sample

$$y_i = X_i + \varepsilon_i$$

The problem of denoising observed sample can be formulated as problem of constructing estimation $\hat{X}$ which is close to X in some way.

3. Purpose and definitions

A presentation of a **topological manifold** is a second countable Hausdorff space that is locally homeomorphic to a vector space, by a collection (called an atlas) of homeomorphisms called charts. The composition of one chart with the inverse of another chart is a function called a transition map, and defines a homeomorphism of an open subset of the linear space onto another open subset of the linear space.

A **differentiable manifold** is a topological manifold equipped with an equivalence class of atlases whose transition maps are all differentiable. More generally, a C^k-manifold is a topological manifold with an atlas whose transition maps are all k-times continuously differentiable.

© B. Vladimir.

4. Algorithm

Main idea of our approach is to estimate not only recovered point $\hat{X}$ bat also tangent space to out manifold in point $\hat{X}_i$, which can be represented as projection matrix $\hat{P}_i$ onto tangent subspace.

Algorithm 1 Bayesian manifold estimator

---

[1] The training sample $_n = (Y_1, \ldots, Y_n)$, the number of iterations K, a sequence of bandwidths $\{h_k : 1 \le k \le K\}$ and of regularizers $\{\beta_{k,i} : 1 \le k \le K, 1 \le i \le n\}$ are given. Initialize $\Sigma_i 0 = I_D$, $1 \le i \le n$. **for** k *from* 0 *to* $K-1$ **do**

end

Compute the weights $w_{ij}k$ according to the formula

$$w_{ij}k = \mathrm{K}\left(\frac{(Y_j - Y_i)^T (\Sigma_i k)^{-1} (Y_j - Y_i)}{h_k^2}\right), \quad 1 \le i, j \le n,$$

where $\mathrm{K}(t)$ is a localizing kernel. Compute

$$N_i = \sum_{j=1}^{n} w_{ij}k,$$

$$\mu_i = \frac{1}{N_i} \sum_{j=1}^{n} w_{ij}k Y_j,$$

$$\Sigma_i = \frac{1}{N_i} \sum_{j=1}^{n} w_{ij}k (Y_j - \mu_i)(Y_j - \mu_i)^T$$

Sample $\Sigma_i^b \sim IW_p(\beta_{k,i} I_D + N_i \Sigma_i, N_i + D)$. Put $\Sigma_i k + 1 = \Sigma_i, \mu_i k + 1 = \mu_i$. **return** the estimates $\widehat{X}_1 = \mu_1 K, \ldots, \widehat{X}_n = \mu_n K$.

---

5. Numerical experiments

Here are some experiments on deleting noise from artificial data. Figure 1 and 2 illustrate some simple examples. Parameters of algorithm were chosen to be approximately equal to noise amplitude. Also we can see importance of bayesian step for numerical stability. On figure 3 we can see results of two variations of algorithm: with bayesian step and without one.

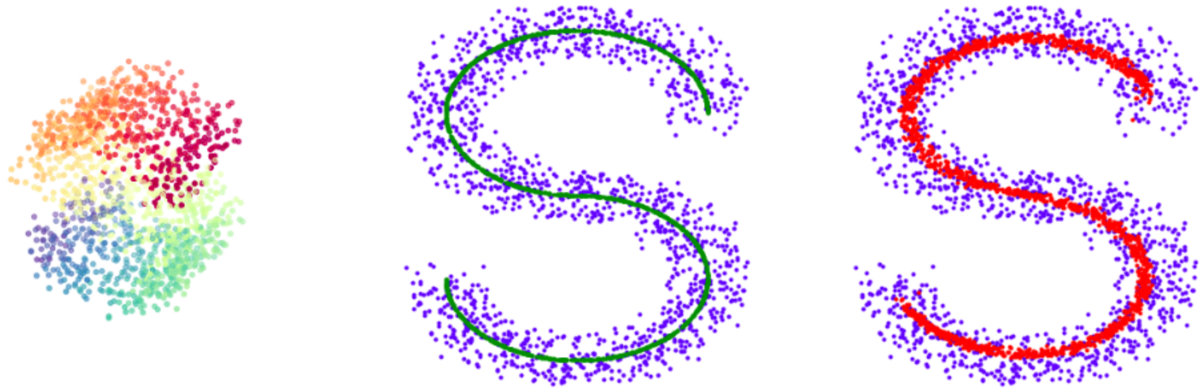

Figure 1: Green points - real data, blue points - noised observations, red points - result of algorithm

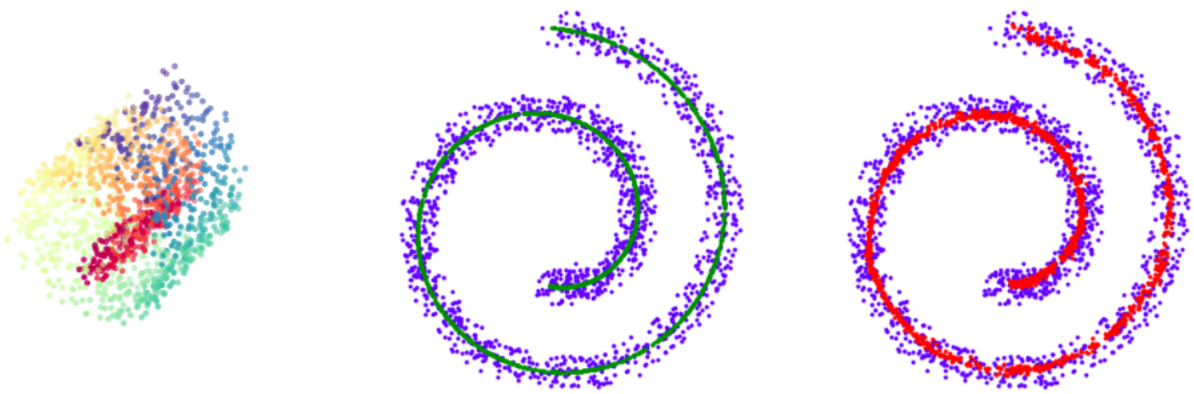

Figure 2: Green points - real data, blue points - noised observations, red points - result of algorithm

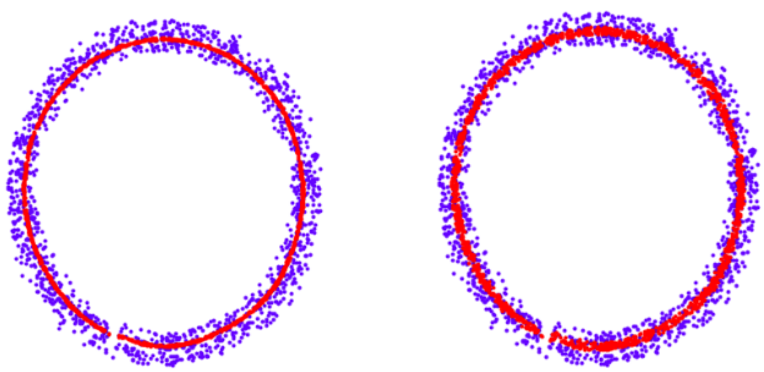

Figure 3: Result of algirithm with bayesian step on the left and without it on the right

6. Discussion of results and plans

New algorithm seems to work well on simple artificial data, so it is time to try it on more complicated examples.

References

[1] Tenenbaum, J. B., de Silva, V. and Langford, J. C. (2000). A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290 2319

[2] Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. SCIENCE 290 2323–2326.

[3] Zhang, Z. and Wang, J. (2007). MLLE: Modified Locally Linear Embedding Using Mul- tiple Weights. In Advances in Neural Information Processing Systems 19 (B. Sch ¨olkopf, J. C. Platt and T. Hoffman, eds.) 1593–1600. MIT Press.

[4] Belkin, M. and Niyogi, P. (2003). Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Neural Comput. 15 1373–1396

[5] van der Maaten, L. and Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research 9 2579–2605.

[6] Maggioni, M. , Minsker, S. and Strawn, N. (2016). Multiscale dictionary learning: non-asymptotic bounds and robustness. J. Mach. Learn. Res. 17 Paper No. 2, 51.

[7] Aamari, E. and Levrard, C. (2019). Nonasymptotic rates for manifold, tangent space and curvature estimation. Ann. Statist. 47 177–204.

[8] Shi, Z. and Sun, J. (2017). Convergence of the point integral method for Laplace–Beltrami equation on point cloud. Research in the Mathematical Sciences 4 22

[9] Cheng, Y. (1995). Mean Shift, Mode Seeking, and Clustering. IEEE Trans. Pattern Anal. Mach. Intell. 17 790–799

[10] Ozertem, U. and Erdogmus, D. (2011). Locally defined principal curves and surfaces. J. Mach. Learn. Res. 12 1249–1286