Minimax-optimal semi-supervised regression on unknown manifolds

by Moscovich, Jaffe, and Nadler

Daniel J. McDonald

14 Sept. 2017

Basic structure

- Introduction — Brief sketch of applications, overview of previous work, no definitions or notation, paragraph stating the main goal
- Problem statement and notation — general setup for semi-supervised learning and definition of kNN regression on the geodesic
- Statistical analysis — what is minimaxity, previous necessary results, main result and proofs
- Computational considerations
- Simulated and real data

Definitions and concepts

- We have Labeled data $\mathcal{L} = \{(x_i, y_i)\}_{i=1}^{n}$ and Unlabeled data $\mathcal{U} = \{(x_j)\}_{j=1}^{m}$.
- We assume that all that all $m + n$ x's are iid from the same marginal distribution supported on a manifold $\mathcal{M}$ of intrinsic dimension d embedded in D-dimensional space.
- Our estimator is in three stages:
    1. For all $x_i \in \mathcal{L} \cup \mathcal{U}$, construct a graph with each point as a vertex. Construct edge weights as $d(x_i, x_j)$.
    2. For all $x_i \in \mathcal{L}$ and $x_j \in \mathcal{U}$, calculate the shortest path graph distance $d_g(x_i, x_j)$.
    3. Use some metric-based supervised learner (they choose kNN)

Geodesic kNN

Let $kNN(x)$ denote the k nearest labeled neighbors of x based on d_g

We can predict at 2 types of points

- If $x_j \in \mathcal{L} \cup \mathcal{U}$,

$$\widehat{f}(x_j) = \frac{1}{|kNN(x_j)|} \sum_{(x_i, y_i) \in kNN(x_j)} y_j.$$

- If $x \notin \mathcal{L} \cup \mathcal{U}$, find the nearest $x^* \in \mathcal{L} \cup \mathcal{U}$, and set

$$\widehat{f}(x) = \widehat{f}(x^*).$$

- Assume $y = f(x) + N(0, \sigma_2)$.
- They are interested in the MSE of their estimator at some point x
- Known lower bound (best estimator) if f is

$$MSE(x) \geq cn^{-\frac{2}{2+d}}$$

- They show that their estimator has

$$MSE(x) \leq c_1 n^{-\frac{2}{2+d}} + O\left(\exp\left\{-c_2(n+m)\right\}\right).$$

The technique

- Previous work shows that d_G actually approximates the geodesic on the manifold for $n + m$ large with high probability.
- Their proof is in 2 steps:
  1. Show that $\mathbb{E}\left[(f(x) - f(x^*))^2\right]$ is small. This relies on an assumption on the marginal distribution of x such that points are reasonably close with exponential probability.
  2. Show that $\mathbb{E}\left[(\widehat{f}(x^*) - f(x^*))^2\right]$ is small. This is via the bias-variance decomposition. But they condition on on the event that d_G is a good approximator.

Computations

- The whole thing seems like a slow operation.
- They present a new algorithm which is faster than the standard but resembles some that others have proposed.

Sims and data

- Their simulation is a bit weak. It's just something closely related to their first dataset
- They examine 2 data sets.
- They compare their method to naive kNN using only the labeled samples.
- They also try Laplacian eigenbasis regression. Essentially it constructs a basis for the estimated manifold using $\mathcal{L} \cup \mathcal{U}$ (see 675 Laplacian Eigenmaps) and then does linear regression with the basis as features.
- They don't give many details. I bet the tuning parameters are garbage.

Conclusions

- I don’t like this paper as much as the first.
- Nonetheless, it has a reasonable theoretical component.
- It also contains the usual requirements for a conference like this: (theory, algorithmic discussion, simulations, real data).
- The structure and formatting are much less nice than the other paper, but fine.
- There is no “Conclusion” or “Future work”. Boo!