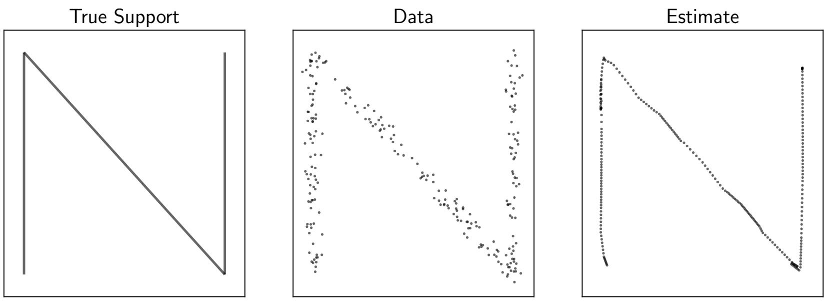
On low-dimensional piecewise linear support estimation

Kentaro Minami (The University of Tokyo)

1 Background and problem setting

A widely believed conjecture in machine learning is that generative distributions of real-world high-dimensional data, such as natural images and audio signals, are concentrated on low-dimensional sub-manifolds. Hence, a major challenge in the study of generative models is to develop stable learning algorithms for probability distributions with low-dimensional supports.

We consider the problem of estimating distributions with 1-dimensional piecewise linear supports. To be precise, suppose that $x_1, \dots, x_n \in \mathbb{R}^d$ are drawn according to the model $x_i = \psi^*(z_i) + \xi_i$ $(i = 1, \dots, n)$, where $\{z_i\}_{i=1}^n$ are unobserved latent variable that are uniformly distributed on $[0, 1]$, $\psi^* : [0, 1] \to \mathbb{R}^d$ is a "decoding" map, and $\{\xi_i\}_{i=1}^n$ are additive noise variables. We assume that ψ^* is piecewise linear, which means that the support of noiseless variable $\tilde{x}_i = \psi^*(z_i)$ consists of connected line segments. Observing $\{x_i\}_{i=1}^n$, our problem is to estimate the true decoding map ψ^*. The following figure shows an illustrative example of the problem.



2 Method

Our proposed method attempts to minimize the following objective with respect to "embeddings" $z_1, \dots, z_n$ and a decoding map ψ:

$$\underbrace{\sum_{i=1}^{n} \|x_i - \psi(z_i)\|_2^2}_{(A)} + \underbrace{\lambda_1 \mathcal{W}^2(z)}_{(B)} + \underbrace{\lambda_2 \mathcal{R}(\psi, z)}_{(C)} . \tag{1}$$

- (A) is the reconstruction error of the encoder-decoder output $\psi(z_i)$.
- (B) is the (squared) Wasserstein distance between the empirical distribution of $\{z_i\}$ and the uniform distribution over $[0, 1]$.
- (C) is the regularization term that encourages ψ to be a less complex piecewise linear function.

The objective (1) is closely related to some existing methods; Minimizing $(A) + (B)$ is a special case of Wasserstein Auto-Encoder [1], while minimizing $(A) + (C)$ for fixed $\{z_i\}$ can be regarded as piecewise linear regression (see e.g. [2, 3]). We will discuss the algorithms and their statistical performances through numerical experiments.

References

[1] I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schölkopf. Wasserstein Auto-Encoders. In ICLR, 2018.

[2] J. H. Friedman. Multivariate Adaptive Regression Splines. *Ann. Statist.*, 19(1):1–67, 1991.

[3] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky. ℓ_1 Trend Filtering. *SIAM Review*, 51(2):339–360, 2009.