Multi-layer approach

Target measure decomposition

IDEA

Decompose the target measure $\nu = \sum_{s\in S} \nu_s \delta_s$ at different scales using the K-means algorithm (Lloyd's algorithm) $\longrightarrow$ Generate a finite sequence of discrete probability measures $\{\nu_l\}_{l=0,\dots,L}$ with decreasing support and such that ν_{l+1} should be a *similar* to ν_l.

MORE PRECISELY

Using a clustering algorithm, generate a finite sequence of finite sets $\{S = S^0, \dots, S^L\}$ (to use as support for the discrete probabilty distributions) such that $|S^l| < |S^{l+1}|$ for all $l \in \{0, \dots, L-1\}$ (and $|S^L| = 1$). Then, define $\nu^0 := \nu$ and use successively transport maps to define the distributions on the supports S^l, i.e. define measurable maps

$$\pi_l : S^l \to S^{l+1} \quad \text{and set} \quad \nu^{l+1} := \pi_{l\#}\nu^l.$$

Thus, we get get successibly probabilty measures $\nu^{l+1} = \sum_{s\in S^{l+1}} \nu_s^{l+1} \delta_s$ supported on S^{l+1} satisfying

$$\nu_s^{l+1} = \nu^{l+1}(s) = \nu^l(\pi_l^{-1}(s)) = \sum_{p\in\pi_l^{-1}(s)} \nu_p^l.$$

Idea: Using Lloyd's algorithm we get: $\pi_l : x \mapsto \arg\min_{s\in S^{l+1}} \|x - s\|^2$.

Multi-layer Transport Map - With 2 Layers (as in Paper)

Reminder from last chapter (is written different):

$$\begin{aligned} h^\nu(x, W) &: \mathbb{R}^d \times \mathbb{R}^{|S|} \to \mathbb{R}, \\ &(x, W) \mapsto \min_{s\in S} \mathrm{pow}_W(x, s) + \langle W, \nu\rangle = (\min_{s\in S} \|x - s\|^2 - W(s)) + \langle W, \nu\rangle \end{aligned}$$

Then we have

$$\nabla_W h^\nu = \nu - \mathbb{1}^S_{T_W(x)=s}\,.$$

Where

$$\mathbb{1}^S_{T_W(x)=s} : S \to \mathbb{R}, \quad x \mapsto \begin{cases} 1 & \text{if } x = T_W(x) \\ 0 & \text{else} \end{cases}.$$

Sketch of the algorithm:

Given: μ (Target distribution), ν (Source distribution), $L = 2$ (number of layers).

- Decompose target measure: $\{\nu^l\}_{l=0,1}$, $\{S^l\}_{l=0,1}$ as above.
- Set $W^l = 0$, for $l = 0, 1$. (Weights to be computed)
- Set $n^l : S^l \to 0$, for $l = 0, 1$. (Number of visits of points in S^l)

Apply ASGD. At each iteration: sample $x \sim \mu$ and then:

1. (L=1: first layer) Compute

$$\tilde{s} = \arg\min_{s \in S^1} \|x - s\|^2 - W^1(s).$$

   I.o.w. compute $T_{W^1}(x)$. If $W^1 = 0$ (as in the first iterations), this is equivalent to computing a Least-squares.

2. Compute gradient

$$g = \nabla_{W^1} h^{\nu^1} = \nu^1 - \mathbb{1}^{S^1}_{s=\tilde{s}}$$

   Where

$$\mathbb{1}^{S^1}_{s=\tilde{s}} : S^1 \to \mathbb{R}, \quad x \mapsto \begin{cases} 1 & \text{if } x = \tilde{s} \\ 0 & \text{else} \end{cases}.$$

3. Update W^1 as in Algorithm 1:

$$W^1 \longleftarrow \text{Use gradient, gradient-step, iteration } (g, C, iter)$$

4. Update number of visits

$$n^1(\tilde{s}) = n^1(\tilde{s}) + 1$$

5. (L=0: second layer) Compute

$$\tilde{\tilde{s}} = \arg\min_{s \in \pi_0^{-1}(\tilde{s})} \|x - s\|^2 - W^0(s)$$

   I.o.w. compute $T_{W^0|_{\pi_0^{-1}(\tilde{s})}}|^{\pi_0^{-1}(\tilde{s})}(x) = T_{W^0}|^{\pi_0^{-1}(\tilde{s})}(x)$, where $T_{W^0}|^{\pi_0^{-1}(\tilde{s})}$ denotes the map T_{W^0} with restricted codomain $\pi_0^{-1}(\tilde{s})$.
   Observations:

   - This computations are faster than computing

$$T_{W^0}(x) = \arg\min_{s \in S^0} \|x - s\|^2 - W^0(s),$$

     as $|\pi_0^{-1}(\tilde{s})| < |S^0|$.
   - It may happend (yes? when?) that

$$T_{W^0}(x) \in S^0 \setminus \pi_0^{-1}(\tilde{s}).$$

     Consequences?

6. Compute gradient

$$\tilde{g} = \nabla_{W^0|_{\pi^{-1}(\tilde{s})}} h^{\nu^0|_{\pi^{-1}(\tilde{s})}} = \nu^0|_{\pi^{-1}(\tilde{s})} - \mathbb{1}^{\pi^{-1}(\tilde{s})}_{s=\tilde{\tilde{s}}}$$

   Where

$$\mathbb{1}^{\pi^{-1}(\tilde{s})}_{s=\tilde{\tilde{s}}} : \pi^{-1}(\tilde{s}) \to \mathbb{R}, \quad x \mapsto \begin{cases} 1 & \text{if } x = \tilde{\tilde{s}} \\ 0 & \text{else} \end{cases}$$

7. Update W^0 as in Algorithm 2:

$$W^0 \longleftarrow \text{Use gradient, gradient-step, number of visits}(\tilde{g}, C, n^1)$$

Actually, only update entries on $\pi^{-1}(\tilde{s})$, i.e

$$W^0|_{\pi^{-1}(s)} \longleftarrow \text{Use gradient, gradient-step, number of visits}(\tilde{g}, C, n^1)$$