CS 369M: Metric Embeddings and Algorithmic Applications Autumn 2018-2019

Homework 2

Due: Monday 11/12 at 11:59 PM

1. In class (lecture 7), we proved the following theorem [Regev 2013].

   Theorem. *Let $k \geq 2, n \geq 1$. Assume $f : [2k]^n \to \mathbb{R}^d$ satisfies that for all $x_1, \ldots, x_n \in [2k]$, $\|f(x_1, \ldots, x_n)\|_1 \leq 1$ and, moreover, for some $\varepsilon < 1/(k-1)$ and for all $\ell \in [n]$, $x_1, \ldots, x_{\ell-1} \in [2k]$, and $r \in [k-1]$,*

   $$\frac{1}{2k} \left\| \sum_{b=1}^{r} (f(x_1, \ldots, x_{\ell-1}, b) + f(x_1, \ldots, x_{\ell-1}, b+k)) - \sum_{b=r+1}^{k} (f(x_1, \ldots, x_{\ell-1}, b) + f(x_1, \ldots, x_{\ell-1}, b+k)) \right\|_1 \geq 1 - \varepsilon$$

   where $f(x_1, \ldots, x_\ell)$ denotes the average of $f(x_1, \ldots, x_n)$ over $x_{\ell+1}, \ldots, x_n$ chosen uniformly in $[2k]$. Then,

   $$d \geq 2^{(\log_2 k - \delta \log_2 (k-1) - H(\delta))n - 1} - 1/2$$

   where $\delta := (k-1)\varepsilon/2 < 1/2$.

   In the proof, we made the assumptions that for all $x_1, \ldots, x_n \in [2k]$, each coordinate of $f(x_1, \ldots, x_n)$ is non-negative and $\|f(x_1, \ldots, x_n)\|_1 = 1$. Extend the proof to the general case where these assumptions may not hold.

2. Let (X, d) be an n-point metric and $\mathcal{D}$ be a distribution over tree metrics that is a probabilistic embedding of (X, d) into trees with expected distortion α. That is, for all $u, v \in X$, for every tree T in the support of $\mathcal{D}$, $d_T(u, v) \geq d(u, v)$ and $E_{T \sim \mathcal{D}}[d_T(u, v)] \leq \alpha d(u, v)$.

   Show that there is a distribution $\mathcal{D}'$ that is a probabilistic embedding of (X, d) into trees with expected distortion α and has support size $O(n^2)$.

   Hint: Use Carathéodory's theorem.
   Note: This is relevant to derandomizing applications of the tree approximation machinery to approximation algorithms. Typically, such applications involve sampling a tree metric from the distribution and running an algorithm on the tree metric. If the only use of randomness is in the initial choice of tree metric, then this can be removed (at the cost of increased running time) if we have a distribution over tree metrics with polynomial support. In this case, the algorithm simply runs over each tree in the support of the distribution and outputs the best solution produced.

3. Consider a tree metric that is given by a σ-HST T with n leaves (which correspond to points in the metric). That is, we are given a rooted tree such that all the leaves have the same depth, and the weight of an edge connecting a node at level $i-1$ with its parent at level i is σ^i (the leaves are at level 0). Denote the set of leaves by L.

Show a procedure that given the tree T outputs another tree T' such that: (i) T has the same set of leaves L, (ii) T has depth $O(\log n)$, (iii) the distances between each two leaves are distorted by a factor of at most $2\sigma/(\sigma-1)$ (when comparing the metrics defined using T and T'), and (iv) the weights of successive edges on any root-to-leaf path in T' decrease at least by a multiplicative factor of σ.

Hint: Contract edges in T in order to get a new tree T' that is *balanced* according to the following recursive definition. We say that a tree T with n leaves is balanced if for every child p of the root of T, the subtree of T rooted at p is balanced and has at most $\lceil n/2 \rceil$ leaves.

4. In this question, you will show that a certain metric called the *earth mover distance* embeds into ℓ_1 with distortion $O(\log n)$. Let (X, d_X) be a metric space. The earth mover distance (EMD) defined over (X, d_X) is a distance between distributions on points in X (in fact, this is a slightly restricted version of EMD). It is commonly used to compute distances between sets of features in computer vision (where the name comes from). It also has a long history of study in mathematics, where such metrics are called *transportation metrics*.

   For any two distributions P, Q on the points in X (that is, $P, Q \in [0,1]^{|X|}$ such that $\sum_{x \in X} P(x) = \sum_{x \in X} Q(x) = 1$), the earth mover distance between P and Q is the minimum cost of a matching between P and Q: The weight $P(x)$ for each $x \in X$ is assigned to one or more points $y \in X$, such that the total weight assigned to each y is $Q(y)$. If we denote by w_{xy} the amount of weight from $P(x)$ that was assigned to $Q(y)$, the cost of the matching is $\sum_{x,y \in X} d_X(x,y) w_{xy}$. We define $\mathrm{EMD}(P, Q)$ to be the minimum cost over all feasible matchings. More precisely, $\mathrm{EMD}(P, Q)$ is the optimal value of the following linear program:

$$\min \sum_{x,y \in X} d_X(x,y) w_{xy}$$

   subject to

$$\begin{aligned} \sum_{y \in X} w_{xy} &= P(x) \quad \forall x \in X \\ \sum_{x \in X} w_{xy} &= Q(y) \quad \forall y \in X \\ w_{xy} &\geq 0 \qquad\quad \forall x, y \in X \end{aligned}$$

   (a) Show that the earth mover distance defined over any *tree metric* is isometrically embeddable into ℓ_1.

   (b) Consider a general metric (X, d) and a distribution $\mathcal{D}$ on tree metrics that is a probabilistic embedding of (X, d) into trees with expected distortion $O(\log n)$. Using the embedding you defined in part (a) and the distribution $\mathcal{D}$, define an embedding f from probability distributions on X into ℓ_1 such that for every P, Q,

$$\mathrm{EMD}(P, Q) \leq \|f(P) - f(Q)\|_1 \leq O(\log n) \cdot \mathrm{EMD}(P, Q).$$