GEOMETRIC DATA ANALYSIS: PROBLEM SET 4

1. Problems

(1) Prove that a metric space represented as a metric tree has nonpositive curvature. (Here by a metric tree I mean a metric space such that any pair of points x and y is joined by a unique path homeomorphic to $[0,1]$.)
(2) Prove that the Hausdorff and Gromov-Hausdorff metrics are actually metrics.
(3) For a finite metric space X, let $\text{diam}(X) = \max_{x,y\in X} d_X(x,y)$. For finite metric spaces X and Y, establish a bound on $d_{GH}(X,Y)$ in terms of $\text{diam}(X)$ and $\text{diam}(Y)$.
(4) Let X be a finite metric space and $Y \subset X$ be an ϵ-net. Prove that $d_{GH}(X,Y) < \epsilon$.
(5) Write code for dimensionality estimation either using local estimates of the dimension of the tangent plane or a "growth of balls" volume estimate. Test on cubes and Gaussians embedded in ambient Euclidean spaces.
(6) Perform a numerical exploration of the Johnson-Lindenstrauss lemma on solid spheres of dimension k embedded in Euclidean spaces of dimension $N > k$.
(7) Write code that uses the technique of finding the centroid in Wasserstein space to "average" distributions in $\mathbb{R}^2$. (You may use a library to compute the Wasserstein barycenter if you wish.) Explore the performance of this by applying it to families of images (e.g., all of the number "1" pictures from MNIST, or a collection of images of cats, or whatever you like). Also explore the behavior of the barycenter on distributions given by sums of Gaussians.
(8) Cluster the tooth or bone data available at `http://www.wisdom.weizmann.ac.il/~ylipman/CPsurfcomp/code_and_data/` using an approximation of d_{GH}. (The paper describing these data sets is available at `https://arxiv.org/abs/1110.3649`.)