Metric Embeddings, Dimension Reduction

[These notes may not be distributed outside this class without the permission of Gregory Valiant.]

1 Dimension Reduction

In the previous lecture notes, we saw that any metric \((X, d)\) with \(|X| = n\) can be embedded into \(R^{O(\log^2 n)}\) under any \(\ell_p\) metric, with distortion \(O(\log n)\). Here, we describe an extremely useful approach for reducing the dimensionality of a Euclidean \((\ell_2)\) metric, while incurring very little distortion. Such dimension reduction is useful for a number of reasons: on the practical side, many geometric algorithms have runtimes that scale poorly with the dimension of the space in which they operate. From a theoretical perspective these dimension-reduction procedures have been used numerous times as components within other algorithms (e.g. Locality Sensitive Hashing).

2 Johnson-Lindenstrauss Transformation

The randomized dimension reduction approach is essentially due to Johnson and Lindenstrauss in 1984 [1], and many variants (and de-randomizations) have been explored in the past 30 years.

**Theorem 1.** Given any \(\epsilon \in (0, 1)\), and a set \(X \subset \mathbb{R}^k\) with \(|X| = n\), there exists a randomized linear map \(f : \mathbb{R}^k \to \mathbb{R}^d\) with \(d = O\left(\frac{\log n}{\epsilon^2}\right)\) that embeds \((X, \ell_2)\) into \((\mathbb{R}^d, \ell_2)\) with distortion at most \((1 + \epsilon)\).

**Proof.** Let \(A\) be the \(k \times d\) matrix with entries chosen independently from \(N(0, 1/\sqrt{d})\), and define the map \(f : \mathbb{R}^k \to \mathbb{R}^d\) by \(f(x) = Ax\). Hence each of the \(d\) coordinates of \(f(x)\) are given by the projection of \(x\) onto a \(k\)-dimensional Gaussian. To analyze the distortion, for a given pair of vectors \(x, y\), we could imagine rotating the coordinate system so that \(x - y\) is a basis vector, in which case we see that \(\|f(x) - f(y)\|_2 = \|x - y\|_2 \frac{1}{\sqrt{d}} \sqrt{\sum_{i=1}^d X_i^2}\), where the \(X_i\)'s are independent Gaussians of unit variance. Hence the theorem will follow provided we show that \(\sum X_i^2\) is sufficiently tightly concentrated about its expectation of \(d\), namely within a \((1 + \delta)\) factor of \(d\), for \(\delta = O(\epsilon)\).

To analyze \(\sum X_i^2\), we will prove a Chernoff-style bound, leveraging the easily verified fact that for \(t < 1/2\), \(\mathbb{E}[e^{tX^2}] = 1/\sqrt{1 - 2t}\) for \(X \sim N(0, 1)\). We begin by bounding the probability that \(\sum X_i^2 > (1 + \delta)d\); a similar argument will show an analogous bound for \(\Pr[\sum X_i^2 < (1 - \delta)d]\).

\[
\Pr[\sum X_i^2 > (1 + \delta)d] = \Pr[e^{t \sum X_i^2} > e^{(1+\delta)t}d], \text{ for } t > 0 \\
\leq \frac{1}{(1-2t)^{d/2}} \text{ [by Markov’s inequality, for } t \in (0, 1/2)] \\
= e^{-d(t(1+\delta)+(1/2)\log(1-2t))}.
\]

Plugging in \(t = \delta/2\), and using the fact that \(\log(1-2t) > -2t - 2t^2\) yields that the above probability is at most \(e^{-d\delta^2/2}\). Hence by choosing \(d > \frac{8\log n}{\delta^2}\), this probability is bounded by \(1/n^2\), and hence we may perform a union bound over all \(<n^2/2\) possible pairs of points \(x, y\) to argue that with constant probability, the embedding does not significantly distort any of the \(O(n^2)\) distances. \(\Box\)
References