

Lecture 8 – Dimension Reduction

Instructor: *Alex Andoni*Scribe: *Kevin Shi*

1 Dimension Reduction

Exact algorithms typically have linear or worse complexity in the dimension. As an example, consider the nearest neighbor search problem. The naive solution of computing the distance from the query point to every training point takes time $O(n \cdot d)$ and space $O(n \cdot d)$. The query time can be improved up to $O(\log n \cdot d)$ by constructing a Voronoi diagram around each point.

Approximate distance preserving dimension reduction cannot be for all points. To see this, suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a function which preserves distance approximately, that is, $(1 - \alpha)\|x - y\| \leq \|f(x) - f(y)\| \leq (1 + \alpha)\|x - y\|$. Then f must be a continuous function; given any $\epsilon > 0$, just choose $\delta = \frac{\epsilon}{1+\alpha}$. But from standard results in point-set topology, there does not exist a continuous mapping from \mathbb{R}^n to any lower dimensional Euclidean space.

2 Johnson-Lindenstrauss

If a subset of points from \mathbb{R}^n is fixed, then dimension reduction can be done while preserving the l_2 norm of each point with high probability.

Claim 1. *There exists a randomized linear map $F : l_2^d \rightarrow l_2^k$, with $k \ll d$, such that for all $x, y \in l_2^d$,*

$$\|x - y\| \leq \|F(x) - F(y)\| \leq (1 + \epsilon)\|x - y\|$$

with probability at least $1 - \exp(-C\epsilon^2k)$.

Note that in the statement of the theorem, the query points are fixed first, and then the random linear map is drawn. Johnson-Lindenstrauss does not hold if the query points are allowed to be chosen adaptively. To prove this statement, we need the following stability property of the Gaussian distribution.

Lemma 2. *Fix $x \in \mathbb{R}^d$. Given d draws $g_i \stackrel{iid}{\sim} \mathcal{N}(0, 1)$, then $\sum_i g_i x_i$ is distributed as $\|x\|_2 g$ where $g \sim \mathcal{N}(0, 1)$.*

In other words, the weighted sum of Gaussian random variables is itself distributed as a Gaussian. The Gaussian distribution is the only 2-parameter distribution which is stable.

Proof of Lemma. The claim is immediate if $x = e_i$ for some coordinate vector e_i . Note that (g_1, \dots, g_d) is a spherical Gaussian and thus the projection onto any vector x only depends on the l_2 -norm of x . Thus

the projection is just $\|x\| \langle \frac{x}{\|x\|}, g \rangle$. Now we can change the basis so that $\frac{x}{\|x\|}$ is a coordinate vector without changing the norm $\|x\|$, so it is sufficient to just consider $x = e_i$. \square

Proof of Claim. It is sufficient to prove that the Johnson Lindenstrauss transform preserves the norm of a vector z with high probability (let $z = x - y$). First we consider the case of a random linear function $f : l_2^d \rightarrow l_2$. The expectation for this is

$$\begin{aligned} \mathbb{E}[\|f(z)\|^2] &= \mathbb{E}[\|z\|^2 g^2] \\ &= \|z\|^2 \mathbb{E}[g^2] \\ &= \|z\|^2 \end{aligned}$$

The general embedding is obtained by repeat this one-dimensional embedding k times, so that

$$F(z) = \frac{1}{\sqrt{k}}(G_1 z, G_2 z, \dots, G_k z) = \frac{1}{\sqrt{k}} G z$$

where G is a $k \times n$ Gaussian random matrix where each entry is drawn from $\mathcal{N}(0, 1)$. By the stability property,

$$F(z) \stackrel{d}{=} \frac{1}{\sqrt{k}}(\|z\| a_1, \dots, \|z\| a_k)$$

where each $a_i \sim \mathcal{N}(0, 1)$. The squared norm of this embedding is distributed as

$$\begin{aligned} \left\| \frac{1}{\sqrt{k}} G z \right\|^2 &= \sum_i \frac{1}{k} (G_i z)^2 \\ &= \frac{1}{k} \|z\|^2 \sum_{i=1}^k G_i^2 \end{aligned}$$

Recall that each G_i is a Gaussian drawn from $\mathcal{N}(0, 1)$, so this is an unbiased estimator. The random variable $\sum_{i=1}^k G_i^2$ is a chi-squared distribution with k degrees of freedom. For this distribution we have the following two concentration results:

$$\begin{aligned} \Pr[\mathcal{X}_k^2 \geq (1 + \epsilon)k] &\leq \exp\left(-\frac{k}{4}(\epsilon^2 - \epsilon^3)\right) \\ \Pr[\mathcal{X}_k^2 \leq (1 - \epsilon)k] &\leq \exp\left(-\frac{k}{4}(\epsilon^2 - \epsilon^3)\right) \end{aligned}$$

which show that $F(z)$ is within $(1 \pm \epsilon)$ of $\|z\|$.

□

The Johnson Lindenstrauss transform can be thought of as projecting x into an almost-orthogonal coordinate system. Given two independent Gaussian vectors g_1, g_2 , their inner product $\frac{g_1}{\|g_1\|} \cdot \frac{g_2}{\|g_2\|}$ has the same distribution as if $g_1 = e_1$ since g_2 is a spherical Gaussian. Thus this distribution is approximately equal to $\frac{g_2}{\sqrt{d}}$ since $\mathbb{E}[\|g\|] = \sqrt{d}$. Note that the variance of this normalized dot product decreases with the dimension d , so for large d if we take $k \ll d$ independent unit Gaussian random vectors, then with high probability they are nearly all orthogonal.

3 Dimension reduction for the l_1 norm

There is an analogous distribution which has the stability property for the l_1 norm. The Cauchy distribution is defined as

$$p(s) = \frac{1}{\pi(s^2 + 1)}$$

and has the property

$$\sum c_i z_i \sim c \cdot \|z\|_1$$

However this distribution is heavy tailed - $\mathbb{E}[s] = \infty$. Thus by a similar argument to a homework problem from last time, taking the mean here over several independent copies of the estimate does not give a good estimator. However, we can still obtain a good estimator with high probability by taking the median. An estimator can be arbitrarily bad (e.g. infinite expectation) but still give the correct answer with high probability. More details next time...