Johnson –Lindenstrauss Cont^d

\[(1 - \varepsilon) \|f(u) - f(v)\| \leq \|u - v\| \leq (1 + \varepsilon) \|f(u) - f(v)\|\]

\[k \sim \frac{1}{\varepsilon^2} \log(n)\]

\((*)\) holds with probability if we choose

\[f(u) = \frac{1}{\sqrt{k}} Au\]

where A is a \(k \times d\) Gaussian

1. Fast – JL Transforms (FJLT)

When the map \(f\) is realized via a Gaussian random matrix, the cost to evaluate \(u \mapsto f(u)\) is \(O(k \times d)\)

Ailon & Chazelle proposed the projection

\[f(u) = PHDu\]

where

- \(D_{d \times d}\) is Diagonal
- \(H_{d \times d}\) is a Hessenberg Transform
- \(P_{k \times d}\) is a ”sort of” subsampling

\(f(u)\) can be evaluated in \(O(d \log(d))\) time.

Ailon & Chazelle proved that the map can approximately preserve distance with high probability.

Later, the subsampled FFT was proposed

\[f((u) = SFD\]

where

- \(F\) is the discrete Fourier Transform
- \(D\) is Diagonal and \(D(i, j) = e^{i\Theta j}\) with \(\Theta j \in [0, 2\Pi]\)

The cost to evaluate \(f\) is \(O(mn \log k)\) \(\rightarrow\) reduced

Recall: We previously used SRFT to accelerate the randomized SVD from \(O(mnk)\) to \(O(mn \log k)\) for rank-\(k\) approximation of an \(m \times n\) matrix

Recent work includes looking for sparse maps or matrices with integer entries.

Question: Can we generalize the J–L Theorem to metric spaces?

2. Bourgain Embedding

The idea of embedding a set \(Q\) of \(n\) points in Euclidean space \(\mathbb{R}^d\) to \(\mathbb{R}^k\) (for ”small” \(k\)) while approximately preserving distances can be generalized to METRIC spaces.

**Theorem (Bourgain Embedding):** Let \(Q\) be a set of \(n\) points and let \(d \times Q \to [0, \infty)\) be a metric. Then, \(\exists\) a map \(f : Q \to \mathbb{R}^k\) for some \(k = O((\log n)^2)\) such that

\[\|f(u) - f(v)\|_{l_1} < d(u, v) \leq \|f(u) - f(v)\|_{l_1} \cdot c \cdot \log(n)\]
where $c$ is a universal constant.

A map $f$ that realizes the bound can "with high probability" be built in polynomial time $n$.

**NOTE:** The "sparsest–cut" problem for a graph $V, E$ is to find a partition $V = S1 \cup S2$ such that $\frac{|E(S1, S2)|}{|S1||S2|}$ is minimized.

3. **Connection to Central Limit Theorem**

**Theorem (Central Limit):** Let $\{\bar{X}_i\}_{i=1}^n$ be a set of i.i.d random variables with mean $\mu$ and variance $\sigma^2$

$$S_k = \frac{1}{k} \sum_{j=1}^{n} \bar{X}_j \text{ (average)}$$

As $K$ increases, the distribution of $S_k$ will approach a normal distribution with mean $\mu$ and variance $\frac{1}{k} \sigma^2$

Let $A$ be a $k \times d$ Gaussian matrix and $y = \frac{1}{\sqrt{k}} Ax$

Then

$$||y||^2 = \frac{1}{k} ||g||^2 ||x||^2$$

where $||g||^2$ has a $\chi^2_k$ distribution.

So, the distribution of $||y||^2$ looks like.
As $k$ grows, the variance in $\|y\|^2$ shrinks, but pretty slowly. This result is similar to classical MONTE–CARLO, where the expected errors shrink as $\frac{1}{\sqrt{k}}$ where $k = \# \text{ samples}$.