Johnson Lindenstrauss

ECE 4200

Goal

Perform a Linear projection to **preserve interpoint distances.**

A **linear projection** from $\mathrm{R}^d \rightarrow \mathrm{R}^k$ is specified by a $k \times d$ matrix W.

A point $\bar{X} \in \mathrm{R}^d$ is mapped to $\hat{\bar{X}} = W\bar{X}$

Why **preserve interpoint distances?**

If all pairwise distances are preserved, the overall geometric structure is preserved

Faster computations in lower dimensions

ε-distance preservation

Suppose $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n$ are n points in d (large) dimensions. A matrix $W: R^d \to R^k$ is ε-**distance** preserving if:

$$(1-\varepsilon)\left\|\bar{X}_i - \bar{X}_j\right\|_2^2 \leq \left\|\hat{\bar{X}}_i - \hat{\bar{X}}_j\right\|_2^2 \leq (1+\varepsilon)\left\|\bar{X}_i - \bar{X}_j\right\|_2^2$$

Interpoint distances are preserved up to multiplicative factor

Q1: how small can k be made?

Q2: how to find such a W?

JL theorem

Q1: Let $0.5 > \varepsilon > 0$ and $k > \frac{30 \log n}{\varepsilon^2}$.

There exists an ε-**distance** preserving map $W: R^d \to R^k$.

Q2: **Random** projection works

A random linear transform

Let W be a $k \times d$ matrix such that

$$W_{ij} \sim N\left(0, \frac{1}{k}\right).$$

Suppose we map each $\bar{X}_i$ to $W\bar{X}_i \in R^k$.

We will show that there is a ***W*** *that satisfies the theorem.*

Step 1: Mean preservation

Let W be the $k \times d$ matrix such with each

$$W_{ij} \sim N\left(0, \frac{1}{k}\right).$$

Claim: Given a fixed $\bar{X}$, and W generated as above, then

$$\mathbb{E}[\|W\bar{X}\|_2^2] = \|\bar{X}\|_2^2$$

Step 1: Mean preservation

Claim: Given a fixed $\bar{X}$, and W generated as above, then

$$\mathbb{E}[\|W\bar{X}\|_2^2] = \|\bar{X}\|_2^2$$

Using this, for any i, j

$$\mathbb{E}\left[\left\|\widehat{\bar{X}}_i - \widehat{\bar{X}}_j\right\|_2^2\right] = \mathbb{E}\left[\|W(\bar{X}_i - \bar{X}_j)\|_2^2\right] = \left\|\widehat{\bar{X}}_i - \widehat{\bar{X}}_j\right\|_2^2$$

Step2: Concentration

Claim: Suppose $k > \frac{30 \log n}{\varepsilon^2}$, when W is chosen at random

$$\Pr\left((1-\varepsilon)\left\|\bar{X}_i - \bar{X}_j\right\|_2^2 \le \left\|\widehat{\bar{X}}_i - \widehat{\bar{X}}_j\right\|_2^2 \le (1+\varepsilon)\left\|\bar{X}_i - \bar{X}_j\right\|_2^2 \right) > 1 - \frac{1}{2n^2}$$

Once we have this we can finish the proof by the union bound.

Step2: Concentration

For i, j, let

$$A_{ij}^{c} := \left\{(1-\varepsilon)\|\bar{X}_i - \bar{X}_j\|_2^2 \le \left\|\hat{\bar{X}}_i - \hat{\bar{X}}_j\right\|_2^2 \le (1+\varepsilon)\|\bar{X}_i - \bar{X}_j\|_2^2\right\}.$$

A_{ij}: event that for the distance between $\bar{X}_i$, and $\bar{X}_j$ is not preserved.

Then, we will show that

$$\Pr(A_{ij}) < \frac{1}{2n^2}.$$

UNION Bound

For any events $E_1, \ldots$

$$\Pr(E_1 \cup E_2 \cup \cdots) \leq \Pr(E_1) + \Pr(E_2) + \cdots$$

Applying this,

$$\Pr(A_{12} \cup A_{13} \cup \cdots) < \Pr(A_{12}) + \Pr(A_{13}) + \cdots$$
$$< \binom{n}{2} \frac{1}{2n^2} < \frac{1}{4}.$$

This shows that

$$\Pr\big((A_{12} \cup A_{13} \cup \cdots)^c\big) > 0.$$

Showing that there is some W for which this event happens! (what is this event?)

How to prove concentration?

A general statement reads as:

A random variable X is pretty close to its mean, with **high probability.**

Markov Inequality

For a non-negative random variable X, and $\alpha > 1$:

$$\Pr(X > \alpha \mathbb{E}[X]) < \frac{1}{\alpha}.$$

Eg,

$$\Pr(X > 2\mathbb{E}[X]) < \frac{1}{2}.$$

PROOF: HZ

Chebychev's Inequality

VARIANCE implies CONCENTRATION.

For a random variable X with variance σ^2, and $\alpha > 1$:

$$\Pr(|X - \mathbb{E}[X]| > \alpha \cdot \sigma) < \frac{1}{\alpha^2}.$$

Eg, X: # Heads in 1024 coin tosses. $\mathrm{Var}(X) = 256.$

$$\Pr(|X - \mathbb{E}[X]| > 32) < \frac{1}{4}.$$

PROOF: HZ

To prove JL

Markov and Chebychev will be insufficient.

Chernoff Bounds!

Chebychev's Inequality

Exponential bounds!