CSCE 658: Randomized Algorithms

Lecture 6

Samson Zhou

Class Logistics

- MUST e-mail me to set up research meeting by next week if you’re interested in doing final project

- Everyone else will be opting into the final exam

Recall: Concentration Inequalities

- Concentration inequalities bound the probability that a random variable is "far away" from its expectation

- Looking at the k^{th} moment for sufficiently high k gives a number of very strong (and useful!) concentration inequalities with exponential tail bounds

- Chernoff bounds, Bernstein's inequality, Hoeffding's inequality, etc.

Recall: Concentration Inequalities

- Suppose we flip a fair coin $n = 100$ times and let H be the total number of heads

- Markov's inequality: $\Pr[H \geq 60] \leq 0.833$
- Chebyshev's inequality: $\Pr[H \geq 60] \leq 0.25$
- 4^{th} moment: $\Pr[H \geq 60] \leq 0.186$
- Bernstein's inequality: $\Pr[H \geq 60] \leq 0.15$
- Truth: $\Pr[H \geq 60] \approx 0.0284$

Last Time: Chernoff Bounds

- Useful variant of Bernstein's inequality when the random variables are binary

- Chernoff bounds: Let $X_1, \ldots, X_n \in \{0, 1\}$ be independent random variables and let $X = X_1 + \cdots + X_n$ have mean μ. Then for any $\delta \geq 0$:

$$\Pr[|X - \mu| \geq \delta\mu] \leq 2\exp\left(-\frac{\delta^2 \mu}{2 + \delta}\right)$$

Last Time: Median-of-Means Framework

- Suppose we design a randomized algorithm A to estimate a hidden statistic Θ of a dataset and we know $0 < \Theta \leq 1000$.
- Suppose each time we use the algorithm A, it outputs a number X such that $\mathrm{E}[X] = \Theta$ and $\mathrm{Var}[X] = 100\Theta^2$
- Suppose we want to estimate Θ to accuracy ε, with probability $1 - \delta$

- Accuracy boosting: Repeat A a total of $\frac{10^{12}}{\varepsilon^2}$ times and take the mean
- Success boosting: Find the mean a total of $O\left(\log\frac{1}{\delta}\right)$ times and take the median, to be correct with probability $1 - \delta$

Trivia Question #3 (Max Load)

- Suppose we have a fair n-sided die that we roll n times. “On average”, what is the largest number of times any outcome is rolled? Example: $1, 5, 2, 4, 1, 3, 1$ for $n = 7$

- $\Theta(1)$
- $\tilde{\Theta}(\log n)$
- $\tilde{\Theta}(\sqrt{n})$
- $\tilde{\Theta}(n)$

Last Time: Max Load

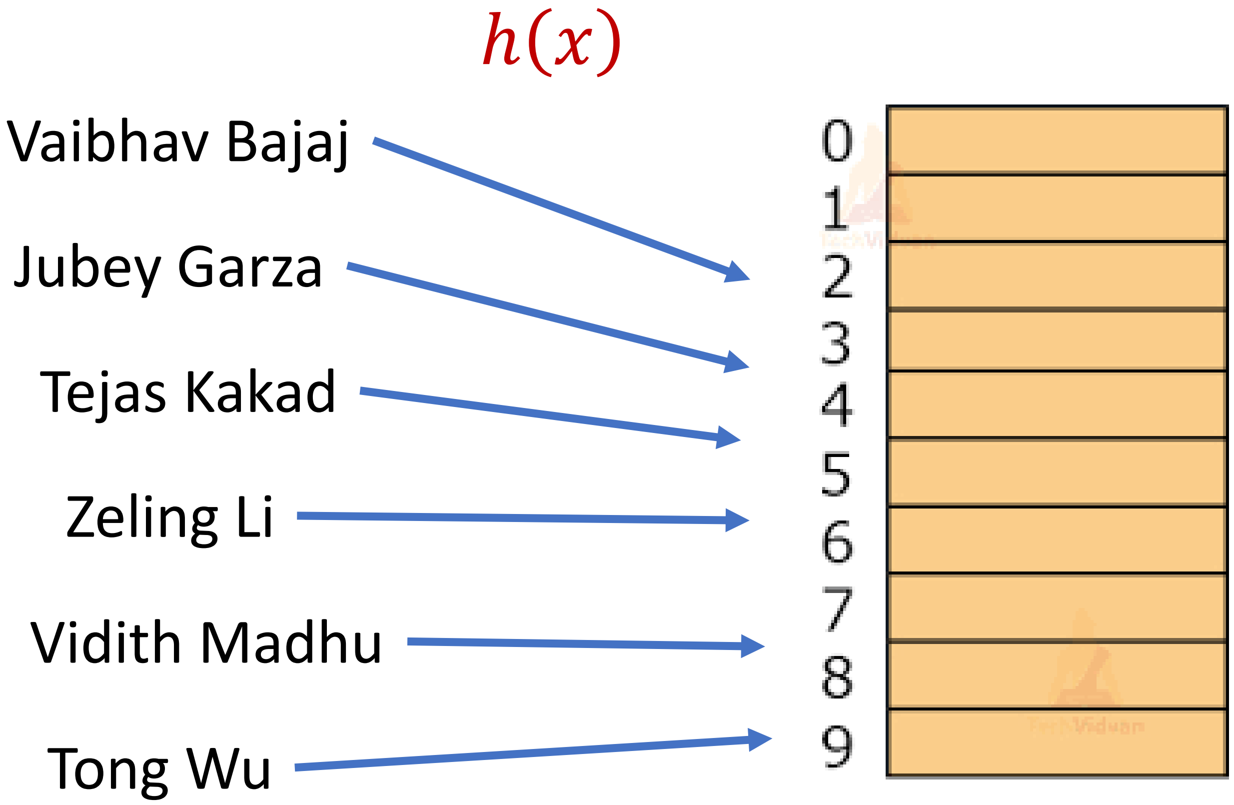
- Recall we fixed a value $k \in [n]$
- $\Pr[X \geq 3\log n] \leq \frac{1}{n^2}$ means that with probability at least $1 - \frac{1}{n^2}$, we will get fewer than $3\log n$ rolls with value k

- Union bound: With probability at least $1 - \frac{1}{n}$, no outcome will be rolled more than $3\log n$ times

Hashing

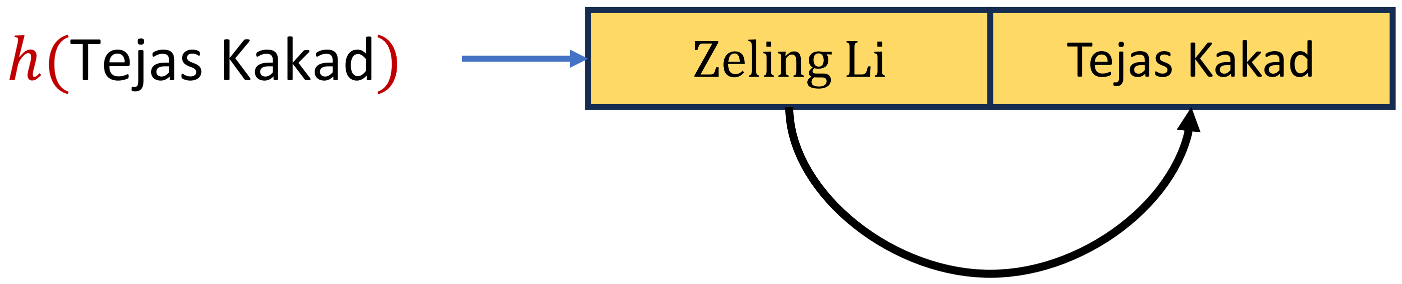
- Suppose we have a number of files, how do we consistently store them in memory?

- If we hash n items, we require $\Theta(n^2)$ slots to avoid collisions



Dealing with Collisions

- Suppose we store multiple items in the same location as a linked list



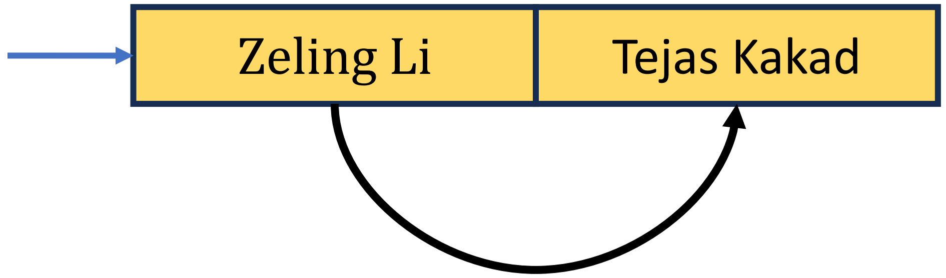
- If the maximum number of collisions in a location is c, then could traverse a linked list of size c for a query

- Query runtime: $O(c)$

Collisions and Max Load

- With probability at least $1 - \frac{1}{n}$, no outcome will be rolled more than $3 \log n$ times

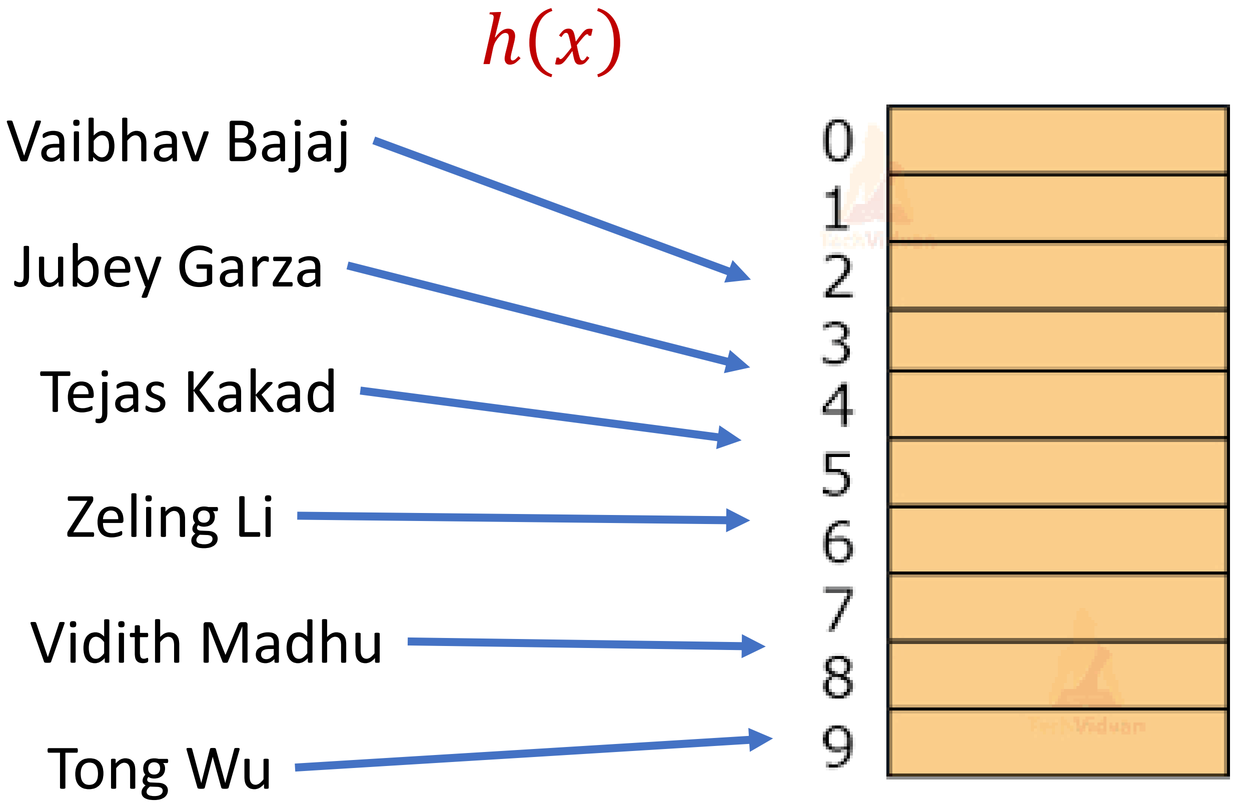
- Worst case query time: $O(\log n)$



Hashing

- For $O(1)$ query time, use $\Theta(n^2)$ slots to avoid collisions

- For $O(\log n)$ query time, use $\Theta(n)$ slots with linked lists



End of Probability Unit

Dimensionality Reduction

Many images from:

Cameron Musco's

COMPSCI 514: Algorithms for Data Science

Big Data

- Not only many data points, but also many measurements per data point, i.e., very high dimensional data

Big Data

- Not only many data points, but also many measurements per data point, i.e., very high dimensional data

- Twitter has 450 million active monthly users (as of 2022), records (tens of) thousands of measurements per user: who they follow, who follows them, when they last visited the site, timestamps for specific interactions, how many tweets they have sent, the text of those tweets, etc...

Big Data

- Not only many data points, but also many measurements per data point, i.e., very high dimensional data

- A 3 minute Youtube clip with a resolution of 500 x 500 pixels at 15 frames/second with 3 color channels is a recording of 2 billion pixel values. Even a 500 x 500 pixel color image has 750,000 pixel values

Big Data

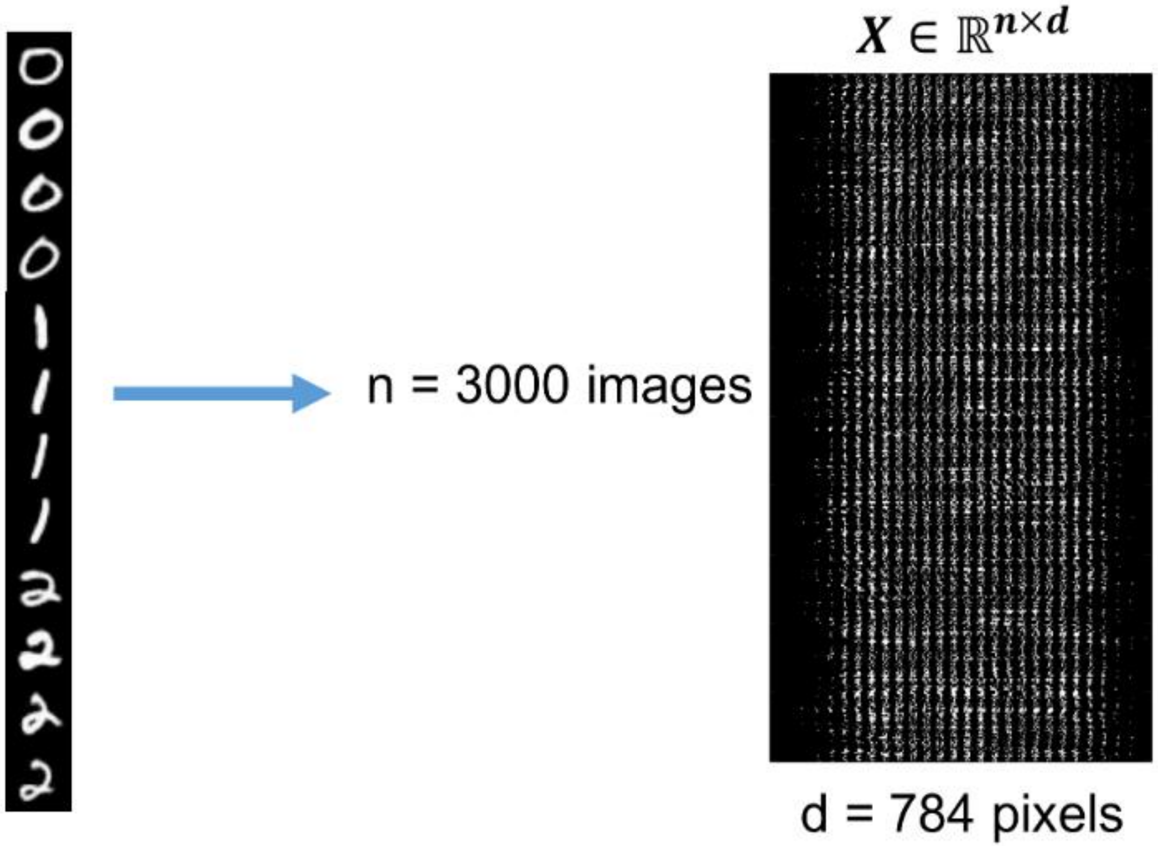
- Not only many data points, but also many measurements per data point, i.e., very high dimensional data

- The human genome contains 3 billion+ base pairs. Genetic datasets often contain information on 100s of thousands+ mutations and genetic markers

Visualizing Big Data

- Data points are interpreted as high dimensional vectors, with real valued entries: $x_1, \dots, x_n \in R^d$

- Dataset is interpreted as a matrix: $X \in R^{n \times d}$ with k-th row x_k



Dimensionality Reduction

- Dimensionality Reduction: Transform the data points so that they have much smaller dimension

$$x_1, \ldots, x_n \in R^d \longrightarrow y_1, \ldots, y_n \in R^m \quad \text{for} \quad m \ll d$$

$$\longrightarrow x_i = (0,1,0,0,1,0,1,1) \longrightarrow y_i = (-1,2,1)$$

- Transformation should still capture the key aspects of $x_1, \ldots, x_n$

Low Distortion Embedding

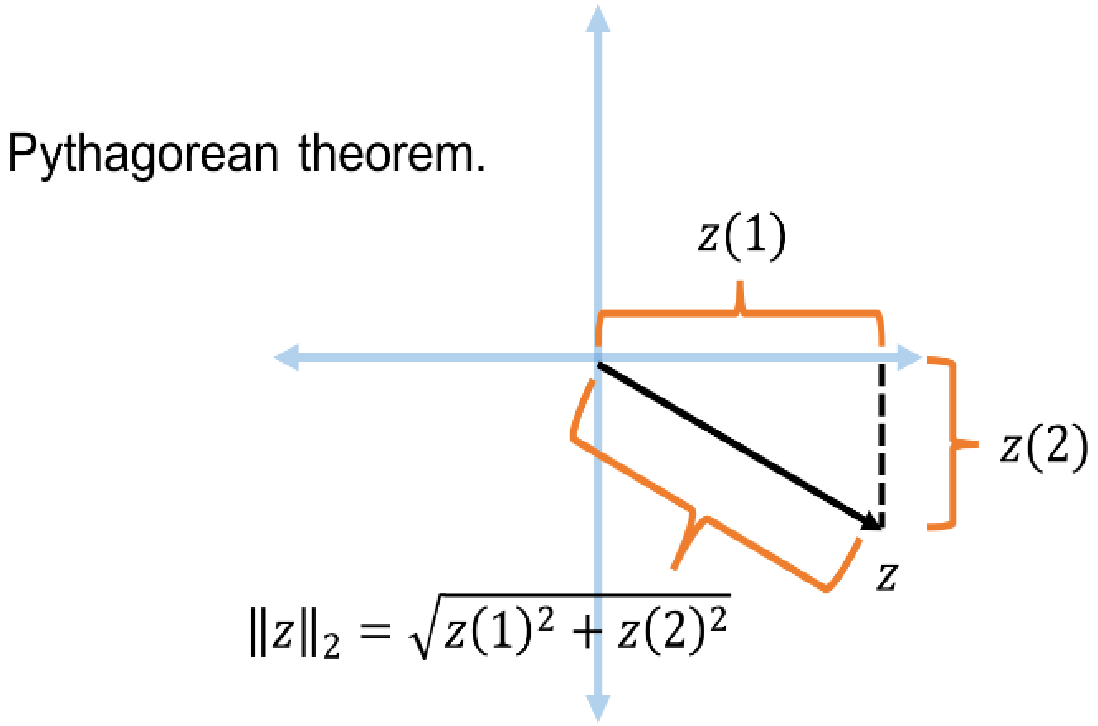
- Given $x_1, \dots, x_n \in R^d$, a distance function D, and an accuracy parameter $\varepsilon \in (0,1)$, a low-distortion embedding of $x_1, \dots, x_n$ is a set of points $y_1, \dots, y_n$, and a distance function D' such that for all $i, j \in [n]$

$$(1-\varepsilon)D(x_i, x_j) \le D'(y_i, y_j) \le (1+\varepsilon)D(x_i, x_j)$$

Euclidean Space

- For $z \in R^d$, the ℓ_2 norm of z is denoted by $\|z\|_2$ and defined as:

$$\|z\|_2 = \sqrt{z_1^2 + z_2^2 + \cdots + z_d^2}$$

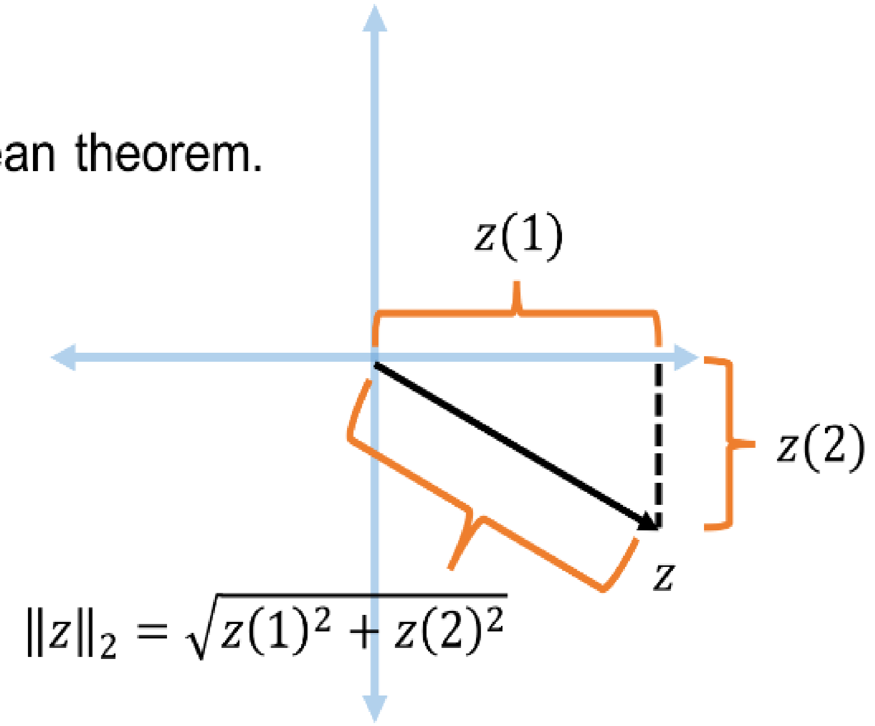


Euclidean Space

- For $z \in R^d$, the ℓ_2 norm of z is denoted by $\|z\|_2$ and defined as:

$$\|z\|_2 = \sqrt{z_1^2 + z_2^2 + \cdots + z_d^2}$$

Pythagorean theorem.

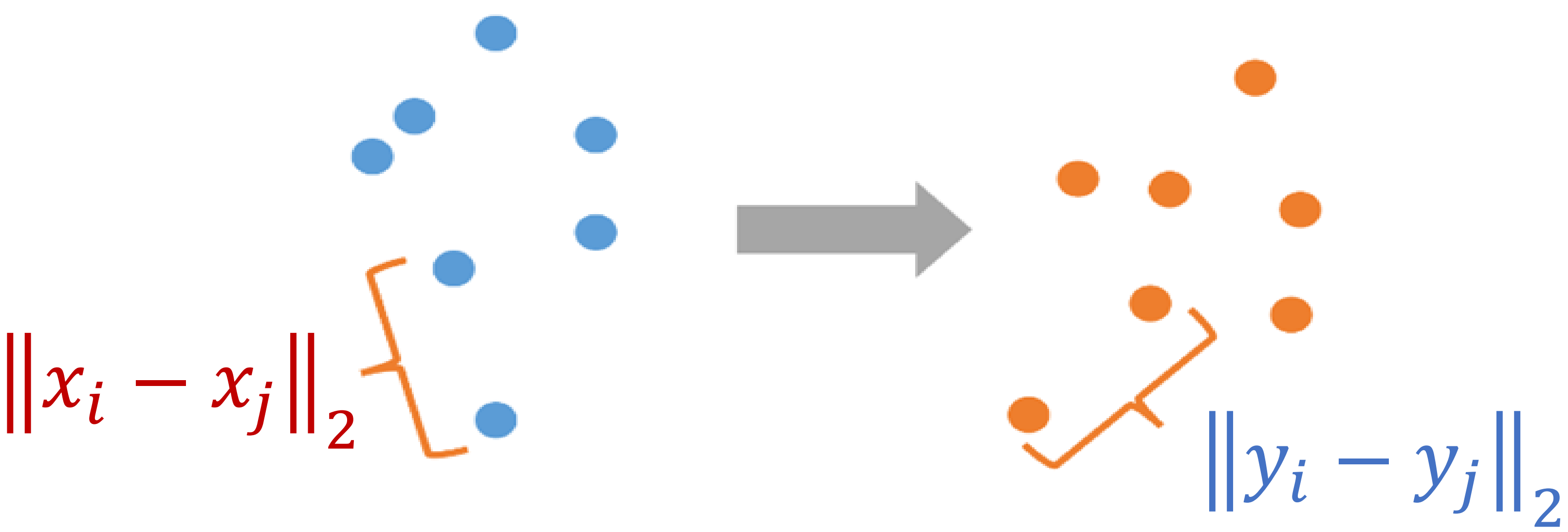


- For $x, y \in R^d$, the distance function D is denoted by $\|\cdot\|_2$ and defined as $\|x - y\|_2$

Low Distortion Embedding for Euclidean Space

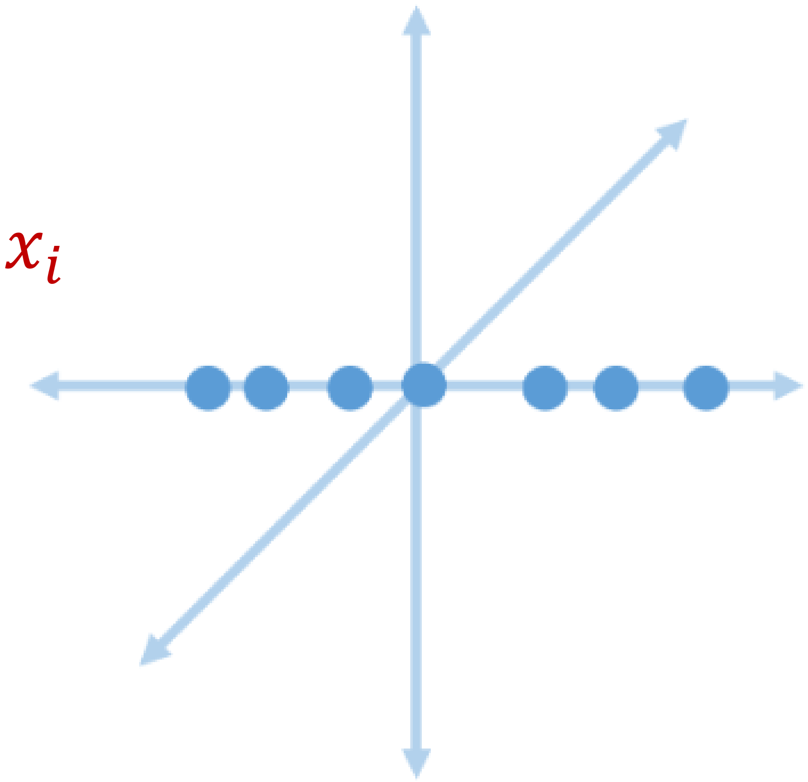
- Given $x_1, \ldots, x_n \in R^d$ and an accuracy parameter $\varepsilon \in (0,1)$, a low-distortion embedding of $x_1, \ldots, x_n$ is a set of points $y_1, \ldots, y_n$ such that for all $i, j \in [n]$

$$(1-\varepsilon)\|x_i - x_j\|_2 \le \|y_i - y_j\|_2 \le (1+\varepsilon)\|x_i - x_j\|_2$$



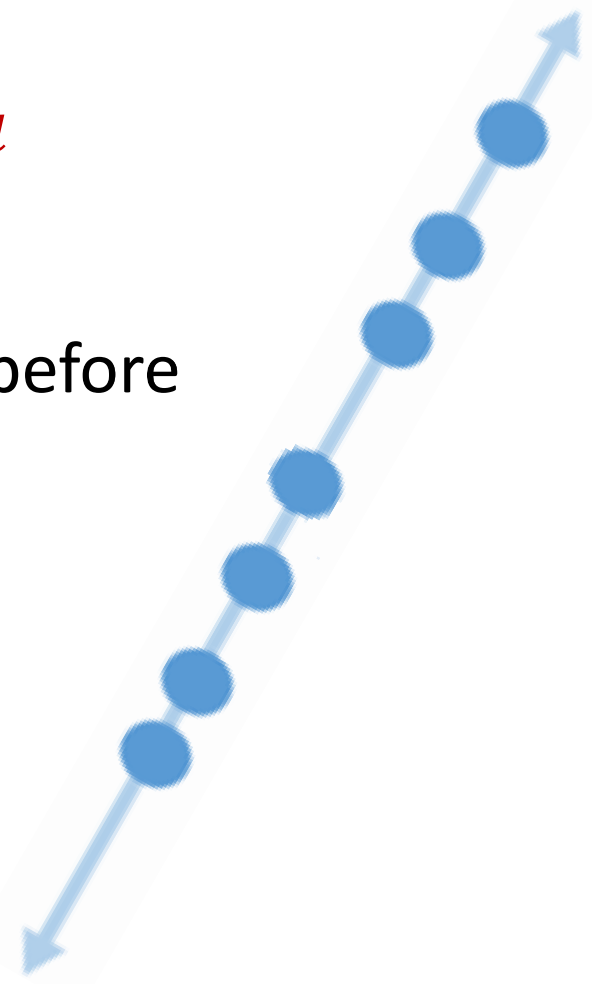
Examples: Embeddings for Euclidean Space

- Suppose $x_1, \dots, x_n \in R^d$ all lie on the 1^{st} - axis
- Take $m = 1$ and y_i to be the first coordinate of x_i
- Then $\left\|y_i - y_j\right\|_2 = \left\|x_i - x_j\right\|_2$ for all $i, j \in [n]$
- Embedding has no distortion



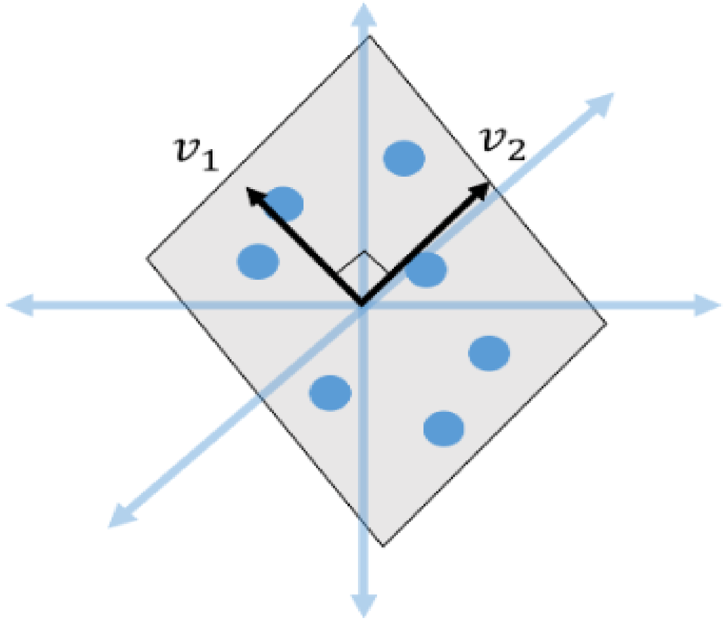
Examples: Embeddings for Euclidean Space

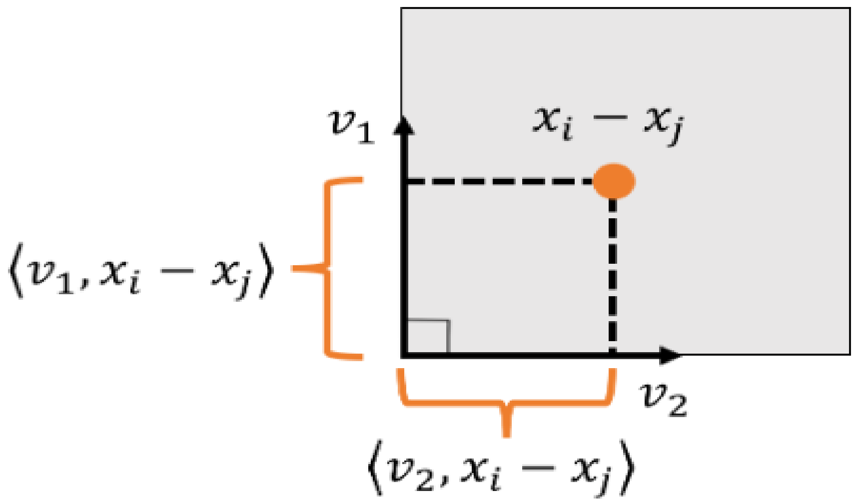
- Suppose $x_1, \ldots, x_n \in R^d$ all lie on some line in R^d
- Rotate to line to be the 1^{st} - axis and proceed as before
- Require $m = 1$ for embedding with no distortion



Examples: Embeddings for Euclidean Space

- Suppose $x_1, \dots, x_n \in R^d$ lie in some k-dimensional subspace V of R^d





- Rotate V to coincide with the k - axes of R^d and set $m = k$

Embeddings for Euclidean Space

- Given $x_1, \ldots, x_n \in R^d$ that lie in *general position*, does there exist an embedding with no distortion?

Embeddings for Euclidean Space

- Given $x_1, \dots, x_n \in R^d$ that lie in *general position*, does there exist an embedding with no distortion? NO!

- Given $x_1, \dots, x_n \in R^d$ that lie in *general position*, does there exist an embedding with ε distortion?

Embeddings for Euclidean Space

- Given $x_1, \dots, x_n \in R^d$ that lie in *general position*, does there exist an embedding with no distortion? NO!

- Given $x_1, \dots, x_n \in R^d$ that lie in *general position*, does there exist an embedding with ε distortion? YES!

- Johnson-Lindenstrauss Lemma

Johnson-Lindenstrauss Lemma

- Johnson-Lindenstrauss Lemma: Given $x_1, \ldots, x_n \in R^d$ and an accuracy parameter $\varepsilon \in (0,1)$, there exists a linear map $\Pi: R^d \to R^m$ with $m = O\left(\frac{\log n}{\varepsilon^2}\right)$ so that if $y_i = \Pi x_i$, then for all $i, j \in [n]$:

$$(1-\varepsilon)\|x_i - x_j\|_2 \leq \|y_i - y_j\|_2 \leq (1+\varepsilon)\|x_i - x_j\|_2$$

Johnson-Lindenstrauss Lemma

- Johnson-Lindenstrauss Lemma: Given $x_1, \ldots, x_n \in R^d$ and an accuracy parameter $\varepsilon \in (0,1)$, there exists a linear map $\Pi: R^d \to R^m$ with $m = O\left(\frac{\log n}{\varepsilon^2}\right)$ so that if $y_i = \Pi x_i$, then for all $i, j \in [n]$:

$$(1-\varepsilon)\|x_i - x_j\|_2 \leq \|y_i - y_j\|_2 \leq (1+\varepsilon)\|x_i - x_j\|_2$$

- For $d = 10^{12}$, $n = 10^5$, and $\varepsilon = 0.5$, only requires $m \approx 6600$

Johnson-Lindenstrauss Lemma

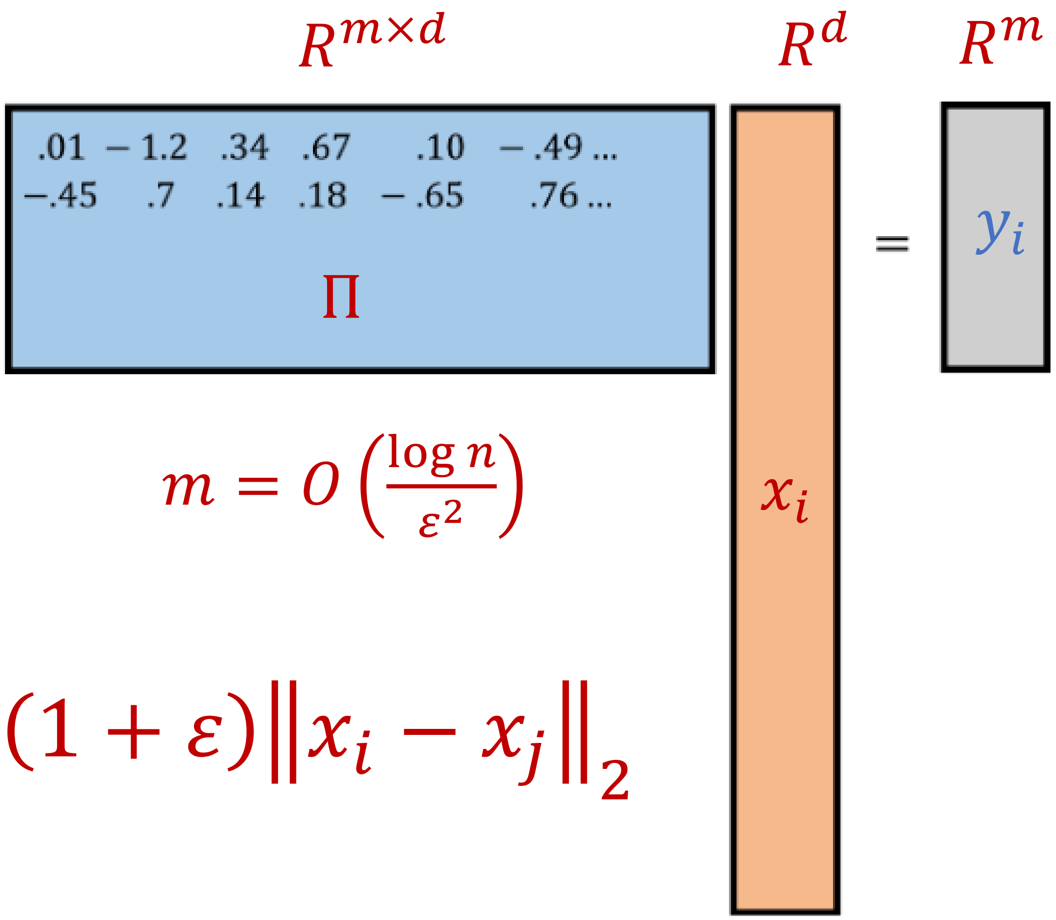
- Johnson-Lindenstrauss Lemma: Given $x_1, \ldots, x_n \in R^d$ and an accuracy parameter $\varepsilon \in (0,1)$, there exists a linear map $\Pi: R^d \to R^m$ with $m = O\left(\frac{\log n}{\varepsilon^2}\right)$ so that if $y_i = \Pi x_i$, then for all $i, j \in [n]$:

$$(1-\varepsilon)\left\|x_i - x_j\right\|_2 \leq \left\|y_i - y_j\right\|_2 \leq (1+\varepsilon)\left\|x_i - x_j\right\|_2$$

- Moreover, if each entry of Π is drawn from $\frac{1}{\sqrt{m}} N(0,1)$, then Π satisfies the guarantee with high probability

Johnson-Lindenstrauss Lemma

- Given $x_1, \dots, x_n \in R^d$ and $\Pi \in R^{m \times d}$ with $m = O\left(\frac{\log n}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}} N(0,1)$ and setting $y_i = \Pi x_i$, then with high probability, for all $i, j \in [n]$:



$$(1 - \varepsilon)\left\|x_i - x_j\right\|_2 \le \left\|y_i - y_j\right\|_2 \le (1 + \varepsilon)\left\|x_i - x_j\right\|_2$$

- Π is called a random projection

Johnson-Lindenstrauss Lemma

- Johnson-Lindenstrauss Lemma: Given $x_1, \dots, x_n \in R^d$ and $\Pi \in R^{m \times d}$ with $m = O\left(\frac{\log n}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}} N(0,1)$ and setting $y_i = \Pi x_i$, then with high probability, for all $i, j \in [n]$:

$$(1-\varepsilon)\|x_i - x_j\|_2 \le \|y_i - y_j\|_2 \le (1+\varepsilon)\|x_i - x_j\|_2$$

- "Applying a simple random linear transformation to a set of points approximately preserves all pairwise distances"

Johnson-Lindenstrauss Lemma

- Distributional Johnson-Lindenstrauss Lemma: Given $\Pi \in R^{m\times d}$ with $m = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}}N(0,1)$, then for any $x \in R^d$ and setting $y = \Pi x$, then with probability at least $1-\delta$

$$(1-\varepsilon)\|x\|_2 \le \|y\|_2 \le (1+\varepsilon)\|x\|_2$$

Johnson-Lindenstrauss Lemma

- Johnson-Lindenstrauss Lemma: Given $x_1, \ldots, x_n \in R^d$ and $\Pi \in R^{m \times d}$ with $m = O\left(\frac{\log n}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}} N(0,1)$ and setting $y_i = \Pi x_i$, then with high probability, for all $i, j \in [n]$:

$$(1-\varepsilon)\left\|x_i - x_j\right\|_2 \leq \left\|y_i - y_j\right\|_2 \leq (1+\varepsilon)\left\|x_i - x_j\right\|_2$$

- Distributional Johnson-Lindenstrauss Lemma: Given $\Pi \in R^{m \times d}$ with $m = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}} N(0,1)$, then for any $x \in R^d$ and setting $y = \Pi x$, then with probability at least $1 - \delta$

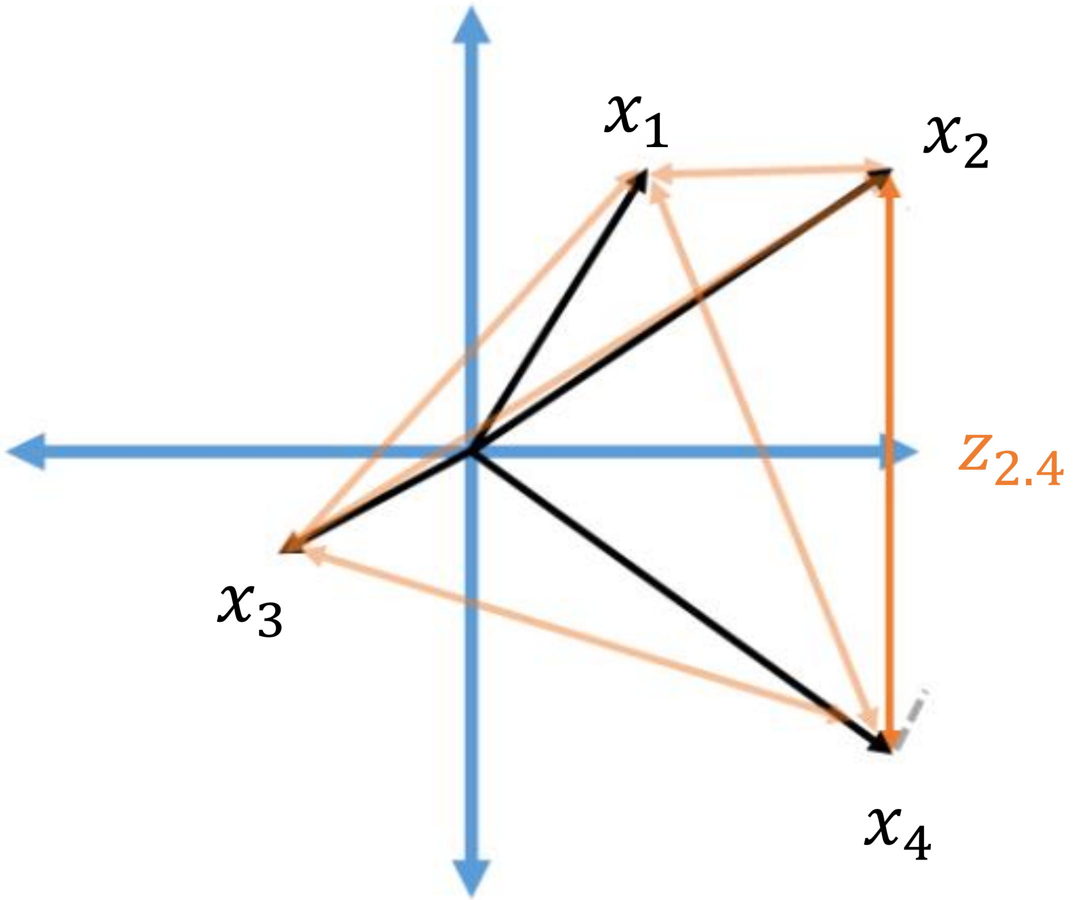
$$(1-\varepsilon)\|x\|_2 \leq \|y\|_2 \leq (1+\varepsilon)\|x\|_2$$

Johnson-Lindenstrauss Lemma

- JL says that the random projection Π preserves all pairwise distances of n points $x_1, \ldots, x_n \in R^d$
- Distributional JL shows that the random projection Π preserves the norm of any $x \in R^d$
- Take $x_1, \ldots, x_n \in R^d$ and define $z_{i,j} = x_i - x_j \in R^d$ for all $i, j \in [n]$
- $\binom{n}{2}$ total vectors

Johnson-Lindenstrauss Lemma

- Take $x_1, \ldots, x_n \in R^d$ and define $z_{i,j} = x_i - x_j \in R^d$ for all $i, j \in [n]$
- $\binom{n}{2}$ total vectors



Johnson-Lindenstrauss Lemma

- Distributional Johnson-Lindenstrauss Lemma: Given $\Pi \in R^{m \times d}$ with $m = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}} N(0,1)$, then for any $x \in R^d$ and setting $y = \Pi x$, then with probability at least $1 - \delta$

$$(1 - \varepsilon)\|x\|_2 \leq \|y\|_2 \leq (1 + \varepsilon)\|x\|_2$$

- What happens when we set $\delta = \frac{1}{n^3}$?

Johnson-Lindenstrauss Lemma

- Distributional Johnson-Lindenstrauss Lemma: Given $\Pi \in R^{m\times d}$ with $m = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}}N(0,1)$, then for any $x \in R^d$ and setting $y = \Pi x$, then with probability at least $1-\delta$

$$(1-\varepsilon)\|x\|_2 \le \|y\|_2 \le (1+\varepsilon)\|x\|_2$$

- What happens when we set $\delta = \frac{1}{n^3}$?
- Union bound

Johnson-Lindenstrauss Lemma

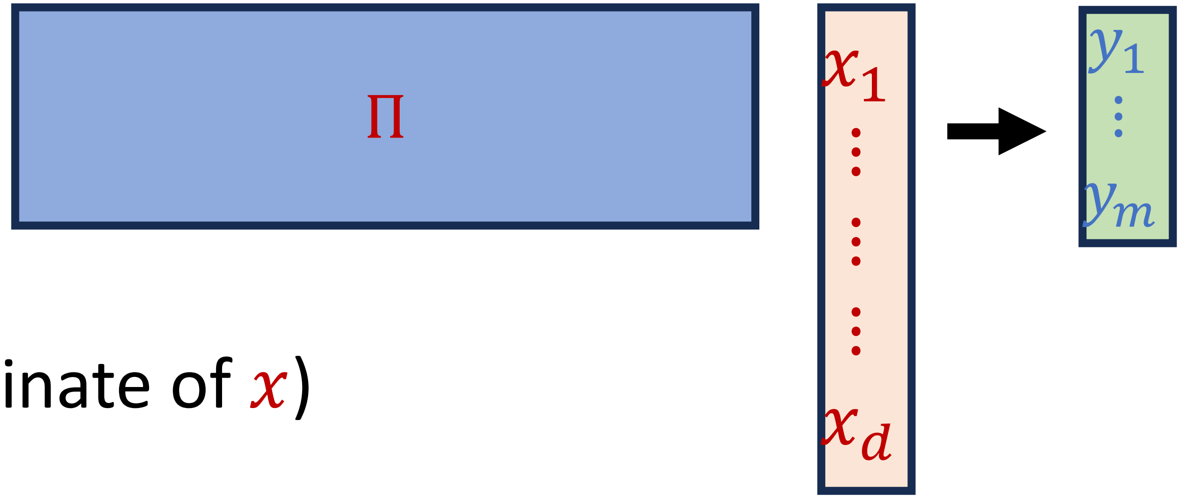
- Distributional Johnson-Lindenstrauss Lemma: Given $\Pi \in R^{m \times d}$ with $m = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}}N(0,1)$, then for any $x \in R^d$ and setting $y = \Pi x$, then with probability at least $1 - \delta$

$$(1 - \varepsilon)\|x\|_2 \le \|y\|_2 \le (1 + \varepsilon)\|x\|_2$$

Johnson-Lindenstrauss Lemma

- Distributional Johnson-Lindenstrauss Lemma: Given $\Pi \in R^{m \times d}$ with $m = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}} N(0,1)$, then for any $x \in R^d$ and setting $y = \Pi x$, then with probability at least $1 - \delta$

$$(1 - \varepsilon)\|x\|_2 \le \|y\|_2 \le (1 + \varepsilon)\|x\|_2$$



(Here x_1 is the first coordinate of x)

Trivia Question #5 (Gaussian Behavior)

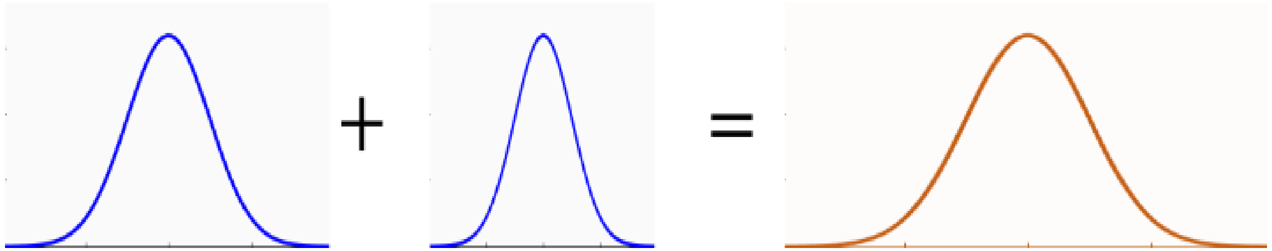
- Let $x \sim N(\mu, \sigma^2)$. What is $\mathrm{E}[x]$ and what is $\mathrm{E}[|x-\mu|^2]$?

- $(0, 1)$
- $(0, \sigma)$
- (μ, σ)
- (μ, σ^2)

PDF of Gaussian $N(\mu, \sigma^2)$ is $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Trivia Question #6 (Gaussian Stability)

- For independent $a \sim N(\mu_1, \sigma_1^2)$ and $b \sim N(\mu_2, \sigma_2^2)$. What is the distribution of $a + b$?



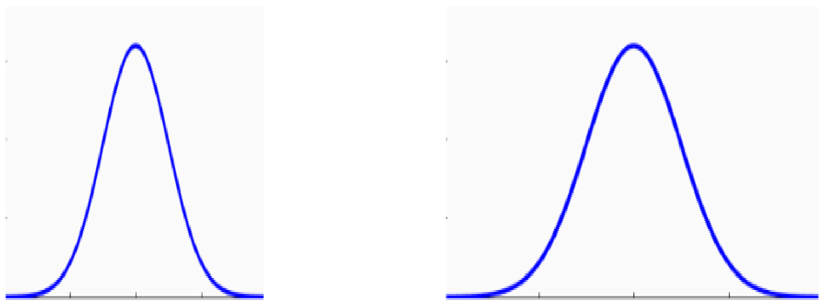
- $N\left(\frac{\mu_1+\mu_2}{2}, \frac{\sigma_1+\sigma_2}{2}\right)$
- $N(\mu_1 + \mu_2, \sigma_1 + \sigma_2)$
- $N\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$
- $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Johnson-Lindenstrauss Lemma

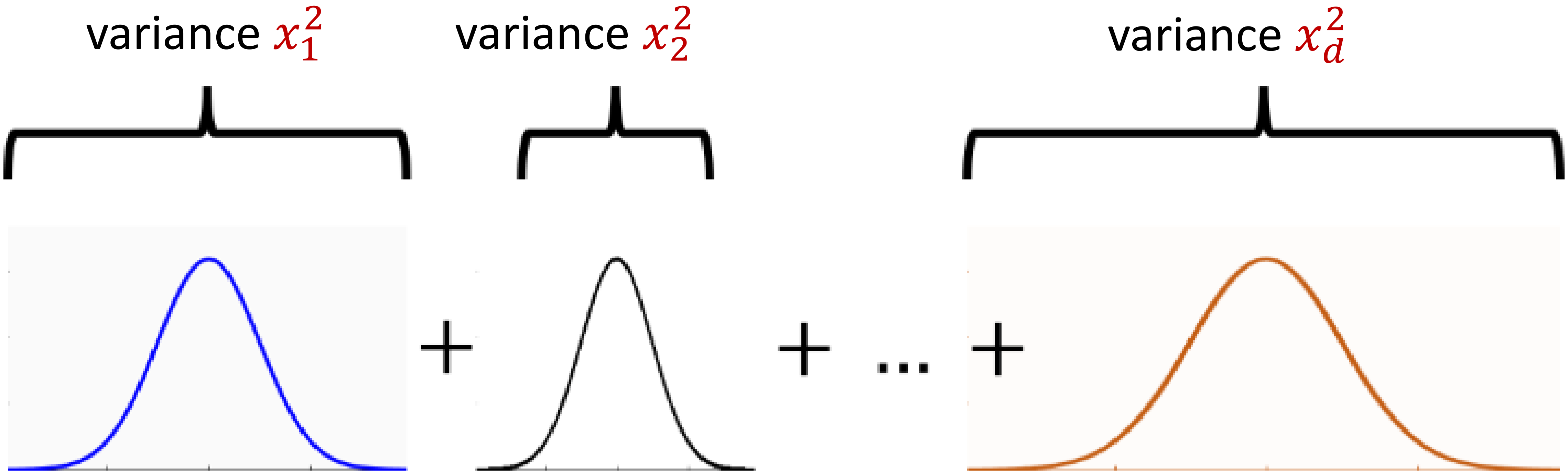
- $y_i = \langle \Pi_i, x \rangle = \frac{1}{\sqrt{m}} \sum_{j=1}^{d} g_j \cdot x_j$ for $g_j \sim N(0,1)$

- $g_j \cdot x_j \sim N\left(0, x_j^2\right)$, normal random variable with variance x_j^2

variance 1 variance x_j^2

g_j  $g_j \cdot x_j$

Gaussian Stability



$$y_i = \langle \Pi_i, x \rangle = \frac{1}{\sqrt{m}}(g_1 \cdot x_1 + g_2 \cdot x_2 + \cdots + g_d \cdot x_d)$$

What is the distribution of y_i?

Gaussian Stability

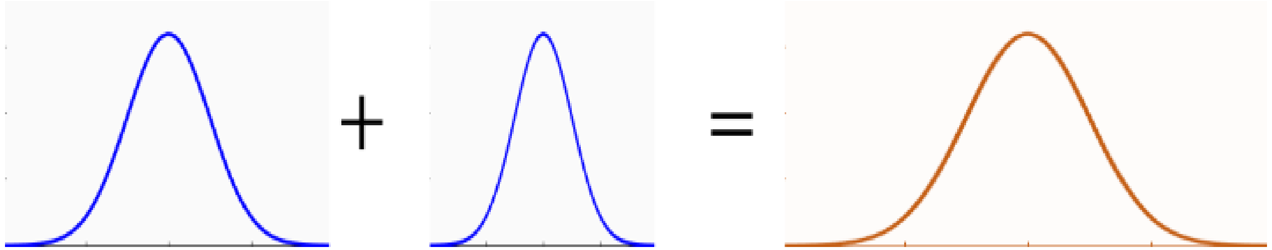
- For independent $a \sim N(\mu_1, \sigma_1^2)$ and $b \sim N(\mu_2, \sigma_2^2)$, we have

$$a + b \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Gaussian Stability

- For independent $a \sim N(\mu_1, \sigma_1^2)$ and $b \sim N(\mu_2, \sigma_2^2)$, we have

$$a + b \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$



$$y_i = \langle \Pi_i, x \rangle = \frac{1}{\sqrt{m}}(g_1 \cdot x_1 + g_2 \cdot x_2 + \cdots + g_d \cdot x_d)$$

$$y_i \sim N\left(0, \frac{1}{m}\|x\|_2^2\right)$$

Gaussian Stability

- For $y_i \sim N\left(0, \frac{1}{m}\|x\|_2^2\right)$, we have $\mathrm{E}\left[y_i^2\right] = \frac{1}{m}\|x\|_2^2$

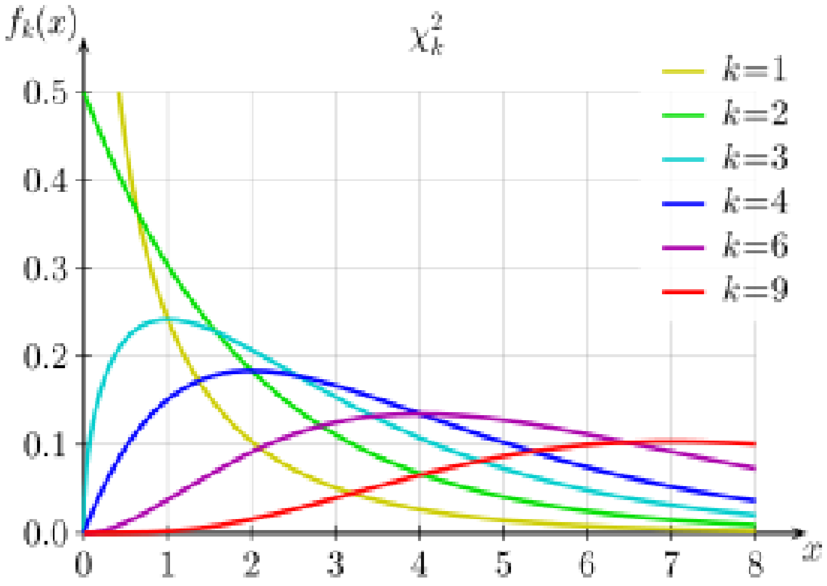
- We have $\mathrm{E}[\|y\|_2^2] = \mathrm{E}[y_1^2 + \cdots + y_m^2] = \mathrm{E}[y_1^2] + \cdots + \mathrm{E}[y_m^2] = \|x\|_2^2$

- Correct expectation!

- How is it distributed?

Johnson-Lindenstrauss Lemma

- $\|y\|_2^2$ is distributed as Chi-Squared random variable with m degrees of freedom (sum of m squared independent Gaussians)



Johnson-Lindenstrauss Lemma

- $\|y\|_2^2$ is distributed as Chi-Squared random variable with m degrees of freedom (sum of m squared independent Gaussians)

- Chi-Squared Concentration: Let Z be a Chi-Squared random variable with m degrees of freedom. Then

$$\Pr[|Z - \mathrm{E}[Z]| \geq \varepsilon \cdot \mathrm{E}[Z]] \leq 2e^{-m\varepsilon^2/8}$$

- Claim follows from setting $m = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$