CSE5401 Lec.15

Dimensionality Reduction

1. The problem
2. Random projections suffice ("Johnson-Lindenstrauss")
3. Application: fast approximate matrix multiplication

Sometimes we need to solve problems on very high-dimensional inputs. We'd like to find smaller-dimension summaries of the inputs we can work with. Problems may depend on the entire input (e.g., is the # of 1-bits even or odd?). Here, we'll focus on geometric problems. Given m points $\vec{x}^{(1)}, \ldots, \vec{x}^{(m)}$ in n dimensions, we would like to find d-dimensional points $\vec{y}^{(1)}, \ldots, \vec{y}^{(m)}$ that preserve the lengths and distances. Recalling that the (Euclidean) length of $\|\vec{x}^{(i)}\|_2 = \sqrt{\langle \vec{x}^{(i)}, \vec{x}^{(i)} \rangle}$ and the distance

$$\|\vec{x}^{(i)} - \vec{x}^{(j)}\|_2 = \sqrt{\langle \vec{x}^{(i)} - \vec{x}^{(j)}, \vec{x}^{(i)} - \vec{x}^{(j)} \rangle}$$
$$= \sqrt{\langle \vec{x}^{(i)}, \vec{x}^{(i)} \rangle - 2\langle \vec{x}^{(i)}, \vec{x}^{(j)} \rangle + \langle \vec{x}^{(j)}, \vec{x}^{(j)} \rangle}$$

It is enough to find $\vec{y}^{(1)}, \ldots, \vec{y}^{(m)}$ such that $\langle \vec{x}^{(i)}, \vec{x}^{(j)} \rangle = \langle \vec{y}^{(i)}, \vec{y}^{(j)} \rangle$ for all pairs $i \neq j$ (possibly $i = j$), with d small. Unfortunately this is generally not possible for $d < m-1$: suppose $\vec{x}^{(i)} = \vec{e}^{(i)}$ i.e., the i^{th} standard basis vector in $\mathbb{R}^{m-1}$, and $\vec{x}^{(m)} = \vec{0}$. Then $\|\vec{x}^{(i)}\|_2 = 1$ for $i = 1, \ldots, m-1$, $\|\vec{x}^{(m)}\|_2 = 0$, and $\langle \vec{x}^{(i)}, \vec{x}^{(j)} \rangle = 0$ for all $i \neq j$. So, we must put $\vec{y}^{(m)} = \vec{0}$ and put $\vec{y}^{(1)}, \ldots, \vec{y}^{(m-1)}$ equal to mutually orthogonal vectors in $\mathbb{R}^d$. But this means that $\vec{y}^{(1)}, \ldots, \vec{y}^{(m-1)}$ are linearly independent, so we know $d \geq m-1$.

If we allow approximation, this turns out to be possible with surprisingly small dimension, $d \sim O\left(\frac{\log m}{\varepsilon^2}\right)$.

for $\|\vec{y}^{(i)} - \vec{y}^{(j)}\|_2 = (1\pm\varepsilon)\|\vec{x}^{(i)} - \vec{x}^{(j)}\|_2$. Moreover, using randomization, we can obtain $\vec{y}^{(i)}$ from $\vec{x}^{(i)}$ via a function that does not depend on the data.

Defn: A Rademacher random variable takes value 1 w.p. $1/2$ and -1 w.p. $1/2$.

Let $R = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & & \vdots \\ \sigma_{d1} & \cdots & \sigma_{dn} \end{bmatrix}$, $S = \sqrt{\frac{1+\varepsilon}{d}} R$ where $\{\sigma_{ij}\}$ is an ensemble of independent Rademacher random variables.

Theorem ("Johnson-Lindenstrauss") For any set of m vectors $\vec{x}^{(1)}, \ldots, \vec{x}^{(m)} \in \mathbb{R}^n$, $d = O(\frac{1}{\varepsilon^2} \log \frac{m}{\delta})$ gives S such that for $\vec{y}^{(i)} = S\vec{x}^{(i)}$ for $i = 1, \ldots, m$, with probability $1-\delta$ all pairs i, j satisfy $\|\vec{y}^{(i)}\|_2 = (1\pm\varepsilon)\|\vec{x}^{(i)}\|_2$ and $\|\vec{y}^{(i)} - \vec{y}^{(j)}\|_2 = (1\pm\varepsilon)\|\vec{x}^{(i)} - \vec{x}^{(j)}\|_2$

For the analysis, we'll need the following variant of a "Chernoff bound" for Rademacher random variables

Theorem (Khintchine's Inequality) For n independent Rademacher random variables $\sigma_1, \ldots, \sigma_n$ any $\vec{x} \in \mathbb{R}^n$, $\varepsilon \in [0, 1]$

$\Pr[|\langle \vec{\sigma}, \vec{x} \rangle| > \varepsilon \|x\|_2] \leq 2e^{-\varepsilon^2/2}$.

Proof of Khintchine's Inequality: For any $t \in \mathbb{R}$, we can consider the value $\mathbb{E}[e^{t\langle \vec{\sigma}, \vec{x} \rangle}]$ (the "Moment Generating Function" of $\langle \vec{\sigma}, \vec{x} \rangle$).

Since the σ_i's are independent, so are $e^{t\sigma_i x_i}$, and we find

$\mathbb{E}[e^{t\langle \vec{\sigma}, \vec{x} \rangle}] = \prod_{i=1}^{n} \mathbb{E}[e^{t\sigma_i x_i}] = \prod_{i=1}^{n} \frac{1}{2}(e^{tx_i} + e^{-tx_i}) = \prod_{i=1}^{n} \left(1 + \frac{(tx_i)^2}{2!} + \frac{(tx_i)^4}{4!} + \cdots\right)$

(for $|tx_i| < 1$ by the Taylor expansion - the odd terms cancel)

Now observe $(2k)! \geq 2^k \cdot k!$ (even terms of $(2k)!$) so this is

$\leq \prod_{i=1}^{n} \left(1 + \frac{(tx_i)^2}{2} + \frac{1}{2!}\left(\frac{(tx_i)^2}{2}\right)^2 + \cdots + \frac{1}{k!}\left(\frac{(tx_i)^2}{2}\right)^k + \cdots\right) = \prod_{i=1}^{n} e^{t^2 x_i^2/2} = e^{t^2 \|\vec{x}\|_2^2/2}$

To obtain the final bound? First consider $\Pr[\langle \vec{\sigma}, \frac{\vec{x}}{\|\vec{x}\|_2} \rangle > \varepsilon]$.

Since $z \mapsto e^{tz}$ is strictly monotone for $t > 0$, this is equal to

$\Pr[e^{t\langle\vec{\sigma},\vec{u}\rangle} > e^{t\varepsilon}]$. Since $e^{tz}>0$, we can use Markov's Inequality: $\mathbb{E}[e^{t\langle\vec{\sigma},\vec{u}\rangle}] = e^{t^2/2}$, so this is at most $e^{t^2/2 - t\varepsilon} = e^{\varepsilon^2/2}$ for $t=\varepsilon$. The same holds for $-\vec{x}$,

Giving $\Pr[-\langle\vec{\sigma},\vec{x}\rangle > \varepsilon\|\vec{x}\|_2] < e^{-\varepsilon^2/2}$ also. ☒

Corollary $\Pr[\|S\vec{x}\|_2^2 > (1+\varepsilon)^2\cdot\|\vec{x}\|_2^2] < e^{-d\varepsilon^2/4}$ and

$\Pr[\|S\vec{x}\|_2^2 < (1-\varepsilon^2)\|\vec{x}\|_2^2] < e^{-d\varepsilon^2/4}$.

Proof: (Notice $e^{t\|R\vec{u}\|_1} = \prod_{j=1}^{d} e^{t|\langle\vec{\sigma}_j,\vec{u}\rangle|}$. where the $\vec{\sigma}_i$'s are independent, so $\mathbb{E}[e^{t\|R\vec{u}\|_1}] = e^{t^2 d/2}$ and $\Pr[e^{t\|R\vec{u}\|_1} > e^{td\varepsilon}] < e^{-td\varepsilon + t^2 d/2}$; for $t=\varepsilon$ we get $\Pr[\|R\vec{u}\|_1 > \varepsilon d] < e^{-d\varepsilon^2/2}$.) Now, $\mathbb{E}[\|R\vec{x}\|_2^2] = \mathbb{E}\left[\sum_{j=1}^{d}\sum_{i\neq i'} x_i x_{i'}\sigma_{ij}\sigma_{i'j}\right] = d\|\vec{x}\|_2^2$ and $\|R\vec{x}\|_2^2 - \mathbb{E}[\|R\vec{x}\|_2^2] = \sum_{j=1}^{d}\sum_{i\neq i'} x_i x_{i'}\sigma_{ij}\sigma_{i'j}$

Claim $\sum_{i\neq i'} x_i x_{i'}\sigma_{ij}\sigma_{i'j} \le 4\left(\sum_{i\in H} x_i\sigma_{ij}\right)\left(\sum_{i'\notin H} x_{i'}\sigma'_{i'j}\right)$ where the σ'_{ij} and σ_{ij} are independent Rademacher ensembles for some $H\subseteq\{1,\dots,n\}$

Proof: Introducing independent Bernoulli random variables $\eta_1,\dots,\eta_n$ we find $\mathbb{E}_{\vec{\eta}}\left[\sum_{i\neq i'} x_i x_{i'}\sigma_{ij}\sigma_{i'j}\eta_i(1-\eta_{i'})\right] = \frac{1}{4}\sum_{i\neq i'} x_i x_{i'}\sigma_{ij}\sigma_{i'j}$ So there is a fixing of $\eta_1,\dots,\eta_n$ such that $\sum_{i\neq i'} x_i x_{i'}\sigma_{ij}\sigma_{i'j}\eta_i(1-\eta_{i'}) \ge \frac{1}{4}\sum_{i\neq i'} x_i x_{i'}\sigma_{ij}\sigma_{i'j}$ that is, putting $H=\{i : \eta_i = 1\}$, $\sum_{i\in H}\sum_{i'\notin H} x_i x_{i'}\sigma_{ij}\sigma_{i'j} \ge \frac{1}{4}\sum_{i\neq i'} x_i x_{i'}\sigma_{ij}\sigma_{i'j}$ where now we can switch to $\sum_{i\in H}\sum_{i'\notin H} x_i x_{i'}\sigma_{ij}\sigma'_{i'j}$ with the same distn. ◊

Now, we again have independence so $e^{t\sum_{j=1}^{d}\sum_{i\in H_j}\sum_{i'\notin H_j} x_i x_{i'}\sigma_{ij}\sigma'_{i'j}} = \prod_{j=1}^{d}\prod_{i\in H_j}\prod_{i'\notin H_j} e^{t x_i x_{i'}\sigma_{ij}\sigma'_{i'j}}$

Let's again focus on $\vec{u} = \frac{\vec{x}}{\|\vec{x}\|_2}$. $\mathbb{E}[e^{t(\|R\vec{u}\|_2^2 - \mathbb{E}[\|R\vec{u}\|_2^2])}] \le \prod_{j=1}^{d}\prod_{i\in H_j}\prod_{i'\notin H_j} e^{2(t u_i u_{i'})^2}$

$\le e^{2dt^2}$. So, Markov's Inequality gives $\Pr[e^{t(\|R\vec{u}\|_2^2 - \mathbb{E}[\|R\vec{u}\|_2^2])} > e^{td\varepsilon}] < e^{-td\varepsilon + 2dt^2}$. For $t = \frac{\varepsilon}{3}$ this gives $\Pr[\|S\vec{u}\|_2^2 > (1+\varepsilon)^2] < e^{-d\varepsilon^2/4}$.

Similarly, $\Pr[e^{t(\mathbb{E}[\|R\vec{u}\|_2^2] - \|R\vec{u}\|_2^2)} > e^{td\varepsilon}] < e^{-td\varepsilon + 2dt^2}$ so $t=\frac{\varepsilon}{3}$ also gives $\Pr[\|S\vec{u}\|_2^2 < (1-\varepsilon^2)] < e^{-d\varepsilon^2/4}$ ☒

With this bound, we're ready to give our guarantee

Proof of Johnson-Lindenstrauss: since $(1-\varepsilon)^2 < (1-\varepsilon)(1+\varepsilon) = 1-\varepsilon^2$, we immediately have $\Pr[\|S\vec{x}^{(i)}\|_2 > (1+\varepsilon)\|\vec{x}^{(i)}\|_2] < e^{-d\varepsilon^2/4}$ and $\Pr[\|S\vec{x}^{(i)}\|_2 < (1-\varepsilon)\|\vec{x}^{(i)}\|_2] < e^{-d\varepsilon^2/4}$. Now, consider $\vec{z}^{(i,j)} = \vec{x}^{(i)} - \vec{x}^{(j)}$ for each pair $i<j$. We similarly find $\Pr[\|S\vec{z}^{(i,j)}\|_2 > (1+\varepsilon)\|\vec{z}^{(i,j)}\|_2]$ and $\Pr[\|S\vec{z}^{(i,j)}\|_2 < (1-\varepsilon)\|\vec{z}^{(i,j)}\|_2]$ are less than $e^{-d\varepsilon^2/4}$. For $d = \frac{18}{\varepsilon^2}\ln\frac{2m}{\delta}$, observe that $e^{-d\varepsilon^2/4} = e^{-2\ln\frac{2m}{\delta}} = \left(\frac{\delta}{2m}\right)^2$. To conclude? We take a union bound over the events $\|S\vec{x}^{(i)}\|_2 > (1+\varepsilon)\|\vec{x}^{(i)}\|_2$ and $\|S\vec{x}^{(i)}\|_2 < (1-\varepsilon)\|\vec{x}^{(i)}\|_2$ for $i=1,\dots,m$ and the events $\|S\vec{z}^{(i,j)}\|_2 > (1+\varepsilon)\|\vec{z}^{(i,j)}\|_2$ and $\|S\vec{z}^{(i,j)}\|_2 < (1-\varepsilon)\|\vec{z}^{(i,j)}\|_2$ for $i<j$. This is $2m + 2\binom{m}{2} = m(m-1)+2m < 2m^2$ events, each of probability $< \frac{\delta}{2m^2}$, so none occur with probability $\geq \delta$. Then, $\|\vec{y}^{(i)}\| = (1\pm\varepsilon)\|\vec{x}^{(i)}\|$ and $\|\vec{y}^{(i)} - \vec{y}^{(j)}\| = \|S\vec{z}^{(i,j)}\|_2 = (1\pm\varepsilon)\|\vec{x}^{(i)} - \vec{x}^{(j)}\|_2$ for all $i \neq j$. ☒

We note that our choice of S is not unique. It can be taken to be a matrix of Gaussian random variables, though sampling is harder in that case. We can also use $O(\log n)$-wise independent random variables to obtain a much smaller representation that is easier to store. Another variant sets all but a fixed number of (randomly chosen) entries per column to 0. This is faster to compute. Other "fast" variants use a structured dense part that is fast to compute with sparse random transformations.

Approximate matrix multiplication

Multiplying large matrices is a bottleneck in many applications. Recall the product of two $n \times n$ matrices $A = \begin{bmatrix} - & \vec{a}^{(1)} & - \\ & \vdots & \\ - & \vec{a}^{(n)} & - \end{bmatrix}$, $B = \begin{bmatrix} | & & | \\ \vec{b}^{(1)} & \cdots & \vec{b}^{(n)} \\ | & & | \end{bmatrix}$ is the matrix C s.t. $C_{ij} = \langle \vec{a}^{(i)}, \vec{b}^{(j)} \rangle$. Naively computing C

using this formula takes time $\Theta(n^3)$. Strassen's divide-and-conquer algorithm runs in time $O(n^{2.81})$. Very sophisticated methods have brought the time down to $O(n^{2.37})$ — but we'd still like to get this closer to $O(n^2)$. If we're willing to take an approximate solution, we can first compute $\widetilde{A} = AS^T$ and $\widetilde{B} = SB$, which are respectively $n \times d$ and $d \times n$ matrices, in time $O(dn^2)$. We can then compute $\widetilde{C} = \widetilde{A}\widetilde{B}$ in time $O(dn^2)$ as well.

Def^n The Frobenius norm of a $n \times m$ matrix M, $\|M\|_F$ is $\sqrt{\sum_{i,j} M_{ij}^2}$.

Theorem For $d = O(\frac{1}{\varepsilon^2} \log \frac{n}{\delta})$, $\|C - \widetilde{C}\|_F \le O(\varepsilon \|C\|_F + \varepsilon \|A\|_F \|B\|_F)$ w.p. $1-\delta$

Proof: The i^{th} row of $\widetilde{A}$ is $S\vec{a}^{(i)}$ and the j^{th} column of $\widetilde{B}$ is $S\vec{b}^{(j)}$. By our previous theorem, $(1-\varepsilon^2)\|\vec{a}^{(i)}\|_2^2 \le \|S\vec{a}^{(i)}\|_2^2 \le (1+\varepsilon)^2 \|\vec{a}^{(i)}\|_2^2$, $(1-\varepsilon^2)\|\vec{b}^{(j)}\|_2^2 \le \|S\vec{b}^{(j)}\|_2^2 \le (1+\varepsilon)^2 \|\vec{b}^{(j)}\|_2^2$, and $(1-\varepsilon^2)\|\vec{a}^{(i)} - \vec{b}^{(j)}\|_2^2 \le \|S\vec{a}^{(i)} - S\vec{b}^{(j)}\|_2^2 \le (1+\varepsilon)^2 \|\vec{a}^{(i)} - \vec{b}^{(j)}\|_2^2$ for all $i \neq j$ with probability $1-\delta$. Recall that $\|\vec{a}^{(i)} - \vec{b}^{(j)}\|_2^2 = \|\vec{a}^{(i)}\|_2^2 + \|\vec{b}^{(j)}\|_2^2 - 2\langle \vec{a}^{(i)}, \vec{b}^{(j)} \rangle$. We therefore have

$$2\left\langle \frac{S\vec{a}^{(i)}}{\|S\vec{a}^{(i)}\|_2}, \frac{S\vec{b}^{(j)}}{\|S\vec{b}^{(j)}\|_2} \right\rangle = \left\| \frac{S\vec{a}^{(i)}}{\|S\vec{a}^{(i)}\|_2} \right\|_2^2 + \left\| \frac{S\vec{b}^{(j)}}{\|S\vec{b}^{(j)}\|_2} \right\|_2^2 - \left\| \frac{S\vec{a}^{(i)}}{\|S\vec{a}^{(i)}\|_2} - \frac{S\vec{b}^{(j)}}{\|S\vec{b}^{(j)}\|_2} \right\|_2^2$$

$$\le (1+\varepsilon)\left[(1+\varepsilon) 2 - (1-\varepsilon) \left\| \frac{\vec{a}^{(i)}}{\|\vec{a}^{(i)}\|_2} - \frac{\vec{b}^{(j)}}{\|\vec{b}^{(j)}\|_2} \right\|_2^2 \right] \le 2(1+\varepsilon)^2 \left\langle \frac{\vec{a}^{(i)}}{\|\vec{a}^{(i)}\|_2}, \frac{\vec{b}^{(j)}}{\|\vec{b}^{(j)}\|_2} \right\rangle +$$

and similarly, $\ge (1+\varepsilon)\left[(1-\varepsilon) 2 - (1+\varepsilon) \left\| \frac{\vec{a}^{(i)}}{\|\vec{a}^{(i)}\|_2} - \frac{\vec{b}^{(j)}}{\|\vec{b}^{(j)}\|_2} \right\|_2^2 \right] \ge (1-\varepsilon^2) 2 \left\langle \frac{\vec{a}^{(i)}}{\|\vec{a}^{(i)}\|_2}, \frac{\vec{b}^{(j)}}{\|\vec{b}^{(j)}\|_2} \right\rangle$ $\quad 4\varepsilon(1+\varepsilon)$ $\quad -4\varepsilon(1+\varepsilon)$

Thus, $\|C - \widetilde{C}\|_F^2 \le \varepsilon^2 \sum_{i,j} \left((2+\varepsilon) \langle \vec{a}^{(i)}, \vec{b}^{(j)} \rangle + 2(1+\varepsilon) \|\vec{a}^{(i)}\|_2 \|\vec{b}^{(j)}\|_2 \right)^2$

$$\le \varepsilon^2 \cdot 2(2+\varepsilon)^2 \underbrace{\sum_{i,j} \langle \vec{a}^{(i)}, \vec{b}^{(j)} \rangle^2}_{\|C\|_F^2} + 2 \cdot \varepsilon^2 \cdot 4 (1+\varepsilon)^2 \sum_{i,j} \|\vec{a}^{(i)}\|_2^2 \|\vec{b}^{(j)}\|_2^2$$

$$\sum_i \|\vec{a}^{(i)}\|_2^2 \sum_j \|\vec{b}^{(j)}\|_2^2$$

$$\|A\|_F^2 \|B\|_F^2$$

By the triangle inequality,

$$\|C - \widetilde{C}\|_F \le \sqrt{2}\varepsilon \left((2+\varepsilon)\|C\|_F + 2(1+\varepsilon)\|A\|_F \|B\|_F \right) \quad \blacksquare$$