CIS 700: "algorithms for Big Data"

Lecture 7: Sketching for Linear Algebra

Slides at http://grigory.us/big-data-class.html

Grigory Yaroslavtsev

http://grigory.us

Penn
UNIVERSITY of PENNSYLVANIA

Least Squares Regression

- Solving an overconstrained linear system
- For $d \ll n$ given:
  - matrix $\boldsymbol{A} \in \mathbb{R}^{n \times d}$
  - vector $\boldsymbol{b} \in \mathbb{R}^{n}$
- Find $\mathbf{x}^* \in \mathbb{R}^d$ that minimizes: $||\boldsymbol{Ax} - \boldsymbol{b}||_2$
- Normal equation: $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}^T\boldsymbol{b}$
- If $\boldsymbol{A}$ has rank d then $\boldsymbol{x}^* = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}$
- Takes $O(nd^2)$ time to compute (using naïve matrix multiplication)

Sketching for Least Squares Regression

- Use JL matrix $\boldsymbol{S} \in \mathbb{R}^{r \times n}$ where $r = \Theta\left(\frac{d}{\epsilon^2}\right) \ll n$
- Solve $\min_x \left|\left|\boldsymbol{SAx} - \boldsymbol{Sb}\right|\right|_2$ instead
- Standard JL: time $O(nrd + rd^2) > O(nd^2)$
- Sparse JL: time $O(nd^2/\epsilon + rd^2)$
- Fast JL: time $O(nd \log n + rd^2)$
- Subspace embeddings from JL:
  - JL only gives a guarantee for a fixed vector
  - We need the guarantee for the column space of A

Oblivious Subspace Embeddings

- Subspace embedding for A:
$$||SAx||_2^2 = (1 \pm \epsilon)||Ax||_2^2$$
- SE for $A \equiv$ SE for U where U is the orthonormal basis for the column space of A
- Least Squares Regression: use SE for (A,b)

$$\min_x ||\boldsymbol{Ax} - \boldsymbol{b}||_2 \to \min_x ||\boldsymbol{SAx} - \boldsymbol{Sb}||_2 = \min_x ||\boldsymbol{S}(\boldsymbol{Ax} - \boldsymbol{b})||_2$$

- Oblivious Subspace Embedding (OSE): matrix S
chosen independently of A, works for any fixed A
- JL transforms can be used as oblivious subspace embeddings

JLT(ϵ, δ, f)

- JLT(ϵ, δ, f): $S \in \mathbb{R}^{k \times n}$ that for any f-element subset $V \subseteq \mathbb{R}^n$ for all $v, v' \in V$ satisfies that:
$$|\langle Sv, Sv' \rangle - \langle v, v' \rangle| \leq \epsilon \left|\left|v\right|\right|_2 \left|\left|v'\right|\right|_2$$
- For unit vectors v, v':
$$|\langle Sv, Sv' \rangle - \langle v, v' \rangle| \leq \epsilon$$
- $\langle Sv, Sv' \rangle =$
$$\frac{1}{2}\left(\left|\left|S(v + v')\right|\right|_2^2 - \left|\left|Sv\right|\right|_2^2 - S\left|\left|v'\right|\right|_2^2\right)$$
$$= \frac{1}{2}\left((1 \pm \epsilon)\left|\left|v + v'\right|\right|_2^2 - (1 \pm \epsilon)\left|\left|v\right|\right|_2^2 - (1 \pm \epsilon)\left|\left|v'\right|\right|_2^2\right)$$
$$= \langle v, v' \rangle \pm O(\epsilon)$$
- Suffices to take regular JL of dimension $d = \Omega(1/\epsilon^2 \log f/\delta)$

OSE construction

- $S = \{y \in \mathbb{R}^n \mid \exists x\colon\ y = Ax, ||y||_2 = 1\}$
- ϵ-net argument: find a set $N \subseteq S$ such that if

$$\langle \boldsymbol{Sw}, \boldsymbol{Sw}' \rangle = \langle \boldsymbol{w}, \boldsymbol{w}' \rangle \pm \epsilon \qquad \forall \boldsymbol{w}, \boldsymbol{w}' \in N$$

then $||\boldsymbol{Sy}||_2^2 = (1 \pm \epsilon)||\boldsymbol{y}||_2^2\ \ \forall \boldsymbol{y} \in S$

- $N = 1/2$-net:

$$\forall y \in S\ \exists\ w \in N\colon ||y - w||_2 \leq \frac{1}{2}$$

- $\boldsymbol{y} = \boldsymbol{y}^0 + \boldsymbol{y}^1 + \boldsymbol{y}^2 + \cdots$, where $||\boldsymbol{y}^i|| \leq \frac{1}{2^i}$ and each $\boldsymbol{y}^i$

is a multiple of a vector in N.

Net argument

- $\boldsymbol{y} = \boldsymbol{y}^0 + \boldsymbol{y}^1 + \boldsymbol{y}^2 + \cdots$, where $\left|\left|\boldsymbol{y}^i\right|\right| \leq \frac{1}{2^i}$ and each $\boldsymbol{y}^i$ is a multiple of a vector in N.
- $\boldsymbol{y} = \boldsymbol{y}^{\mathbf{0}} + (\boldsymbol{y} - \boldsymbol{y}^{\mathbf{0}})$ where $\boldsymbol{y}_{\mathbf{0}} \in \boldsymbol{N}$, $||\boldsymbol{y} - \boldsymbol{y}^{\mathbf{0}}||_{\mathbf{2}} \leq \frac{1}{2}$
- $(\boldsymbol{y} - \boldsymbol{y}^{\mathbf{0}}) = \boldsymbol{y}^{\mathbf{1}} + \left((\boldsymbol{y} - \boldsymbol{y}^{\mathbf{0}}) - \boldsymbol{y}^{\mathbf{1}}\right)$ where $\boldsymbol{y}^{\mathbf{1}} \in N$ and $\left|\left|\left((\boldsymbol{y} - \boldsymbol{y}^{\mathbf{0}}) - \boldsymbol{y}^{\mathbf{1}}\right)\right|\right|_{\mathbf{2}} \leq \frac{\left|\left|\boldsymbol{y} - \boldsymbol{y}^{\mathbf{0}}\right|\right|}{2} \leq 1/4$
- $\left|\left|\boldsymbol{S}\boldsymbol{y}\right|\right|_2^2 = \left|\left|\boldsymbol{S}(\boldsymbol{y}^0 + \boldsymbol{y}^1 + \boldsymbol{y}^2 + \cdots)\right|\right|_2^2$

$$= \sum_{0 \leq i < j < \infty} \left|\left|\boldsymbol{S}\boldsymbol{y}^i\right|\right|_2^2 + 2\langle \boldsymbol{S}\boldsymbol{y}^{\boldsymbol{i}}, \boldsymbol{S}\boldsymbol{y}^{\boldsymbol{j}} \rangle$$

$$\leq \left(\sum_{0 \leq i < j < \infty} \left|\left|\boldsymbol{y}^i\right|\right|_2^2 + 2\langle \boldsymbol{y}^{\boldsymbol{i}}, \boldsymbol{y}^{\boldsymbol{j}} \rangle \right) \pm 2\epsilon\Big(\sum_{0 \leq i \leq j < \infty} \left|\left|\boldsymbol{y}^i\right|\right|_2 \left|\left|\boldsymbol{y}^j\right|\right|_2 \Big)$$

$$= 1 \pm O(\epsilon)$$

½ -Net construction

- For $0 < \gamma < 1$ there is a γ-net for S of size $\leq \left(1 + \frac{2}{\gamma}\right)^d$
- Choose a maximal set N' of points on S^d such that no two points are within γ of each other
- Balls of radius $\frac{\gamma}{2}$ around the points are disjoint
- Ball of radius $1 + \frac{\gamma}{2}$ around the origin contains all balls
- # points $\leq \left(\frac{1+\frac{\gamma}{2}}{\frac{\gamma}{2}}\right)^d = \left(1 + \frac{2}{\gamma}\right)^d$
- Size of ½-net $\leq 5^d$
- JLT of dimension $\Omega((d + \log\frac{1}{\delta})/\epsilon^2)$ gives OSE

OSE constructions Running Times

nnz(A) = # non-zero entries in A

- OSE from Sparse JL: time $O(nnz(A)d/\epsilon)$
- Fast JL: time $O(nd \log n)$
- [Clarkson, Woodruff'13] possible to construct OSE in time $O(nnz(A))$