Fast and Space Efficient NLA

Lecture 20

Nov 3, 2022

Some topics today

We have seen fast "approximation" algorithms for matrix multiplication

- random sampling
- Using JL

Today:

- Subspace embeddings for faster linear least squares and low-rank approximation
- Frequent directions algorithms for one/two pass approximate SVD

Subspace Embedding

Question: Suppose we have linear subspace E of $\mathbb{R}^n$ of dimension d. Can we find a projection $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^k$ such that for *every* $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$.
- Possible if $k = d$. Pick Π to be an orthonormal basis for E. **Disadvantage:** This requires knowing E and computing orthonormal basis which is slow.

What we really want: *Oblivious* subspace embedding ala JL based on random projections

Oblivious Supspace Embedding

Theorem

Suppose E *is a linear subspace of* $\mathbb{R}^n$ *of dimension* d*. Let* Π *be a DJL matrix* $\Pi \in \mathbb{R}^{k \times n}$ *with* $k = O(\frac{d}{\epsilon^2}\log(1/\delta))$ *rows. Then with probability* $(1-\delta)$ *for every* $x \in E$*,*

$$\|\frac{1}{\sqrt{k}}\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2.$$

In other words JL Lemma extends from one dimension to arbitrary number of dimensions in a graceful way.

Part I

Faster algorithms via subspace embeddings

Linear least squares/Regression

Linear least squares: Given $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n} \times \boldsymbol{d}}$ and $\boldsymbol{b} \in \mathbb{R}^{\boldsymbol{d}}$ find $\boldsymbol{x}$ to minimize $\|\boldsymbol{Ax} - \boldsymbol{b}\|_2$.

Interesting when $\boldsymbol{n} \gg \boldsymbol{d}$ the over constrained case when there is no solution to $\boldsymbol{Ax} = \boldsymbol{b}$ and want to find best fit.

Geometrically $\boldsymbol{Ax}$ is a linear combination of columns of $\boldsymbol{A}$. Hence we are asking what is the vector $\boldsymbol{z}$ in the column space of $\boldsymbol{A}$ that is closest to vector $\boldsymbol{b}$ in ℓ_2 norm.

Closest vector to $\boldsymbol{b}$ is the projection of $\boldsymbol{b}$ into the column space of $\boldsymbol{A}$ so it is "obvious" geometrically. How do we find it?

Linear least squares/Regression

Linear least squares: Given $A \in \mathbb{R}^{n \times d}$ and $b \in \mathbb{R}^{d}$ find x to minimize $\|Ax - b\|_2$.

Interesting when $n \gg d$ the over constrained case when there is no solution to $Ax = b$ and want to find best fit.

Geometrically Ax is a linear combination of columns of A. Hence we are asking what is the vector z in the column space of A that is closest to vector b in ℓ_2 norm.

Closest vector to b is the projection of b into the column space of A so it is "obvious" geometrically. How do we find it? Find an orthonormal basis $z_1, z_2, \ldots, z_r$ for the columns of A. Compute projection c as $c = \sum_{j=1}^{r} \langle b, z_j \rangle z_j$ and output answer as $\|b - c\|_2$.

Linear least squares via Subspace embeddings

Let $a_1, a_2, \ldots, a_d$ be the columns of A and let E be the subspace spanned by $\{a_1, a_2, \ldots, a_d, b\}$

E has dimension at most $d + 1$.

Use subspace embedding on E. Applying JL matrix Π with $k = O(\frac{d}{\epsilon^2})$ rows we reduce $a_1, a_2, \ldots, a_d, b$ to $a'_1, a'_2, \ldots, a'_d, b'$ which are vectors in $\mathbb{R}^k$.

Solve $\min_{x' \in \mathbb{R}^d} \|A'x' - b'\|_2$

Low-rank approximation

Recall: Given $A \in \mathbb{R}^{n \times d}$ and integer k want to find best rank matrix B to minimize $\|A - B\|_F$

- SVD gives optimum for all k. If $A = UDV^T = \sum_{i=1}^{d} \sigma_i u_i v_i^T$ then $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$ is optimum for every k.
- $\|A - A_k\|_F^2 = \sum_{i>k} \sigma_i^2$.
- $v_1, v_2, \ldots, v_k$ are k orthogonal unit vectors from $\mathbb{R}^d$ and maximize the sum of squares of the projection of the **rows** of A onto the space spanned by them
- $u_1, u_2, \ldots, u_k$ are k orthogonal unit vectors from $\mathbb{R}^n$ that maximize the sum of squares of the projections of the **columns** of A onto the space spanned

Low-rank approximation via subspace embeddings

Column view of SVD: $u_1, u_2, \ldots, u_k$ are k orthogonal unit vectors from $\mathbb{R}^n$ that maximize the sum of squares of the projections of the columns of A onto the space spanned

Let $a_1, a_2, \ldots, a_d$ be the columns of A and let E be subspace spanned by them. $\dim(E) \leq d$ obviously.

Wlog $u_1, u_2, \ldots, u_k \in E$. Why?

Low-rank approximation via subspace embeddings

Column view of SVD: $u_1, u_2, \ldots, u_k$ are k orthogonal unit vectors from $\mathbb{R}^n$ that maximize the sum of squares of the projections of the columns of A onto the space spanned

Let $a_1, a_2, \ldots, a_d$ be the columns of A and let E be subspace spanned by them. $\dim(E) \leq d$ obviously.

Wlog $u_1, u_2, \ldots, u_k \in E$. Why?
If $u_1, u_2, \ldots, u_k$ fixed then $v_1, v_2, \ldots, v_k$ are determined. Why?

Low-rank approximation via subspace embeddings

Column view of SVD: $u_1, u_2, \ldots, u_k$ are k orthogonal unit vectors from $\mathbb{R}^n$ that maximize the sum of squares of the projections of the columns of A onto the space spanned

Let $a_1, a_2, \ldots, a_d$ be the columns of A and let E be subspace spanned by them. $\dim(E) \leq d$ obviously.

Wlog $u_1, u_2, \ldots, u_k \in E$. Why?
If $u_1, u_2, \ldots, u_k$ fixed then $v_1, v_2, \ldots, v_k$ are determined. Why?

Low-rank approximation via subspace embeddings

Column view of SVD: $u_1, u_2, \ldots, u_k$ are k orthogonal unit vectors from $\mathbb{R}^n$ that maximize the sum of squares of the projections of the columns of A onto the space spanned

Let $a_1, a_2, \ldots, a_d$ be the columns of A and let E be subspace spanned by them. $\dim(E) \leq d$ obviously.

Wlog $u_1, u_2, \ldots, u_k \in E$. Why?
If $u_1, u_2, \ldots, u_k$ fixed then $v_1, v_2, \ldots, v_k$ are determined. Why?
Let Π be an ϵ-approximate subspace preserving embedding for E

Claim: $\|(\Pi A) - (\Pi A)_k\|_F \leq (1 + \epsilon)\|A - A_k\|_F$

Analysis

Claim: $\|(\Pi A) - (\Pi A)_k\|_F \leq (1+\epsilon)\|A - A_k\|_F$

Analysis

Claim: $\|(\Pi A) - (\Pi A)_k\|_F \leq (1+\epsilon)\|A - A_k\|_F$

Proof sketch: Let $a'_1, \ldots, a'_d$ be columns of ΠA and let $u'_1, \ldots, u'_k$ be $\Pi u_1, \ldots, \Pi u_k$.

Analysis

Claim: $\|(\Pi A) - (\Pi A)_k\|_F \leq (1+\epsilon)\|A - A_k\|_F$

Proof sketch: Let $a'_1, \ldots, a'_d$ be columns of ΠA and let $u'_1, \ldots, u'_k$ be $\Pi u_1, \ldots, \Pi u_k$.

$\|A - A_k\|_F^2 = \sum_{i=1}^{d} \|a_i - \sum_{j=1}^{k} v_j(i) u_j\|_2^2$

Analysis

Claim: $\|(\Pi A) - (\Pi A)_k\|_F \leq (1+\epsilon)\|A - A_k\|_F$

Proof sketch: Let $a'_1, \ldots, a'_d$ be columns of ΠA and let $u'_1, \ldots, u'_k$ be $\Pi u_1, \ldots, \Pi u_k$.

$\|A - A_k\|_F^2 = \sum_{i=1}^d \|a_i - \sum_{j=1}^k v_j(i) u_j\|_2^2$

From subspace embedding property of Π,
$\|\Pi(a_i - \sum_{j=1}^k v_j(i) u_j)\|_2 \leq (1+\epsilon)\|a_i - \sum_{j=1}^k v_j(i) u_j\|_2$

Since $u'_1, u'_2, \ldots, u'_k$ is a feasible solution for k-rank approximation to ΠA.

Analysis

Claim: $\|(\Pi A) - (\Pi A)_k\|_F \leq (1+\epsilon)\|A - A_k\|_F$

Proof sketch: Let $a'_1, \ldots, a'_d$ be columns of ΠA and let $u'_1, \ldots, u'_k$ be $\Pi u_1, \ldots, \Pi u_k$.

$\|A - A_k\|_F^2 = \sum_{i=1}^{d} \|a_i - \sum_{j=1}^{k} v_j(i) u_j\|_2^2$

From subspace embedding property of Π,
$\|\Pi(a_i - \sum_{j=1}^{k} v_j(i) u_j)\|_2 \leq (1+\epsilon)\|a_i - \sum_{j=1}^{k} v_j(i) u_j\|_2$

Since $u'_1, u'_2, \ldots, u'_k$ is a feasible solution for k-rank approximation to ΠA.

Claim: $\|(\Pi A) - (\Pi A)_k\|_F \geq (1-\epsilon)\|A - A_k\|_F$.

Analysis

Claim: $\|(\Pi A) - (\Pi A)_k\|_F \leq (1+\epsilon)\|A - A_k\|_F$

Proof sketch: Let $a'_1, \ldots, a'_d$ be columns of ΠA and let $u'_1, \ldots, u'_k$ be $\Pi u_1, \ldots, \Pi u_k$.

$\|A - A_k\|_F^2 = \sum_{i=1}^{d} \|a_i - \sum_{j=1}^{k} v_j(i) u_j\|_2^2$

From subspace embedding property of Π,
$\|\Pi(a_i - \sum_{j=1}^{k} v_j(i) u_j)\|_2 \leq (1+\epsilon)\|a_i - \sum_{j=1}^{k} v_j(i) u_j\|_2$

Since $u'_1, u'_2, \ldots, u'_k$ is a feasible solution for k-rank approximation to ΠA.

Claim: $\|(\Pi A) - (\Pi A)_k\|_F \geq (1-\epsilon)\|A - A_k\|_F$. Prove it!

Running Time

- A has d columns in $\mathbb{R}^n$ and ΠA has d columns in $\mathbb{R}^k$ where $k = O(\frac{d}{\epsilon^2} \ln(1/\delta))$. Hence dimensionality reduction from n to k and one can run SVD on ΠA.
- ΠA can be computed fast in time roughly proportional to $nnz(A)$ (number of non-zeroes of A).

Part II

Frequent Directions Algorithm

Low-rank approximation

Faster low-rank approximation algorithms based on randomized algorithm: sampling and subspace embeddings

- Can we find a deterministic algorithm?
- Streaming algorithm?

Low-rank approximation and SVD

Given matrix $A \in \mathbb{R}^{n \times d}$ and (small) integer k

Row view of SVD: $v_1, v_2, \ldots, v_k$ are k orthogonal unit vectors from $\mathbb{R}^d$ that maximize the sum of squares of the projections of the rows A onto the space spanned

Let $a_1, a_2, \ldots, a_n$ be the rows of A (treated as vectors in $\mathbb{R}^d$)

$\sigma_j^2 = \sum_{i=1}^n \langle a_i, v_j \rangle^2$ and $\|A - A_k\|_F^2 = \sum_{j>k} \sigma_j^2$

Low-rank approximation and SVD

Given matrix $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n} \times \boldsymbol{d}}$ and (small) integer $\boldsymbol{k}$

Row view of SVD: $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{\boldsymbol{k}}$ are $\boldsymbol{k}$ orthogonal unit vectors from $\mathbb{R}^{\boldsymbol{d}}$ that maximize the sum of squares of the projections of the rows $\boldsymbol{A}$ onto the space spanned

Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{\boldsymbol{n}}$ be the rows of $\boldsymbol{A}$ (treated as vectors in $\mathbb{R}^{\boldsymbol{d}}$)

$\boldsymbol{\sigma}_{\boldsymbol{j}}^2 = \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \langle \boldsymbol{a}_{\boldsymbol{i}}, \boldsymbol{v}_{\boldsymbol{j}} \rangle^2$ and $\|\boldsymbol{A} - \boldsymbol{A}_{\boldsymbol{k}}\|_{\boldsymbol{F}}^2 = \sum_{\boldsymbol{j} > \boldsymbol{k}} \boldsymbol{\sigma}_{\boldsymbol{j}}^2$

Consider matrix $\boldsymbol{D}_{\boldsymbol{k}} \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{T}}$ whose rows are $\boldsymbol{\sigma}_1 \boldsymbol{v}_1, \boldsymbol{\sigma}_2 \boldsymbol{v}_2, \ldots, \boldsymbol{\sigma}_{\boldsymbol{k}} \boldsymbol{v}_{\boldsymbol{k}}$.
$\|\boldsymbol{D}_{\boldsymbol{k}} \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{T}}\|_{\boldsymbol{F}}^2 = \sum_{\boldsymbol{j}=1}^{\boldsymbol{k}} \boldsymbol{\sigma}_{\boldsymbol{j}}^2 = \|\boldsymbol{A}_{\boldsymbol{k}}\|_{\boldsymbol{F}}^2$

Frequent Directions Algorithm

[Liberty] and analyzed for relative error guarantee by [Ghashami-Phillips]
Liberty inspired by Misra-Greis frequent items algorithm.

Rows of A come one by one

Algorithm maintains a matrix $Q \in \mathbb{R}^{\ell \times d}$ where $\ell = k(1 + 1/\epsilon)$.
Hence memory is $O(kd/\epsilon)$

At end of algorithm let Q_k be best rank k-approximation for Q.
Then $\|A - \text{Proj}_{Q_k}(A)\|_F \leq (1 + \epsilon)\|A - A_k\|_F$.

Thus a $(1 + \epsilon)$-approximate k-dimensional subspace for rows of A be identified by storing $O(k/\epsilon)$ rows.

FD Algorithm

Frequent-Directions

Initialize Q^0 as an all zeroes $\ell \times d$ matrix
For each row $a_i \in A$ do
  Set $Q_+ \leftarrow Q^{i-1}$ with last row replaced by a_i
  Compute SVD of Q_+ as UDV^T
  $C^i = DV^T$ (for analysis)
  $\delta_i = \sigma_\ell^2$ (for analysis)
  $D' = \mathrm{diag}(\sqrt{\sigma_1^2 - \delta_i}, \sqrt{\sigma_2^2 - \delta_i}, \ldots, \sqrt{\sigma_{\ell-1}^2 - \delta_i}, 0)$
  $Q^i = D'V^T$
EndFor
Return $Q = Q^n$

If $\ell = \lceil k(1 + 1/\epsilon) \rceil$ and Q_k is the rank k approximation to output Q then

$$\|A - \mathrm{Proj}_{Q_k}(A)\|_F \le (1 + \epsilon)\|A - A_k\|_F$$

Running time

- One pass algorithm but requires second pass to compute actual singular values etc
- Space $O(kd/\epsilon)$
- Run time: n computations of SVD on $k/\epsilon \times d$ matrix. Can be improved (see home work problem).

Interesting even when $k = 1$. Alternative to power method to find top singular value/vector. Deterministic.