EE270
Large scale matrix computation, optimization and learning

Instructor : Mert Pilanci

Stanford University

Tuesday, Feb 11 2020

Randomized Linear Algebra and Optimization
Lecture 11: Spectral Approximation, Subspace Embedding and Fast JL Transforms

Approximating Matrices

Approximate matrix product $A^T A \approx A^T S^T S A$

sampling based vs projection based methods

Let $A = U \Sigma V^T$ be the Singular Value Decomposition of A

- **Sampling based**
  - Uniform
  - Row norm scores $p_i = \frac{\|a_i\|_2^2}{\sum_j \|a_j\|_2^2}$
  - Leverage scores $p_i = \frac{\|u_i\|_2^2}{\sum_j \|u_j\|_2^2}$

Approximating Matrices

Approximate matrix product $A^T A \approx A^T S^T S A$

sampling based vs projection based methods

Let $A = U \Sigma V^T$ be the Singular Value Decomposition of A

- **Sampling based**
  - Uniform
  - Row norm scores $p_i = \frac{\|a_i\|_2^2}{\sum_j \|a_j\|_2^2}$
  - Leverage scores $p_i = \frac{\|u_i\|_2^2}{\sum_j \|u_j\|_2^2}$
- **Projection based**
  - Gaussian $N(0, 1)$ random projection
  - Rademacher ± 1 random projection
  - Haar (uniform orthogonal) random projection
  - Sparse Johnson Lindenstrauss (CountSketch) Embeddings
  - Fast Johnson Lindenstrauss (Randomized Hadamard) Transform

Leverage Scores

- Let $A = U\Sigma V^T$ be the Singular Value Decomposition of A implies Least Squares cost approximation
- Importance sampling: proportional to the rows norms of U
- Leverage scores: $\ell_i := \|u_i\|_2^2$ for $i = 1, \ldots n$
- $\sum_i \ell_i = \sum_i \|u_i\|_2^2 = \|U\|_F^2 = trU^T U = trI_d = d$ when A is full column rank
- Sampling probabilities: $p_i = \frac{1}{d}\|u_i\|_2^2$
  $\sum_i p_i = 1$
- Can be non-uniform or uniform $A = [I; 0]$
- Approximate Matrix Multiplication for $U^T U$ i.e,
  $\|U^T S^T S U - U^T U\|_F = \|U^T S^T S U - I\|_F \leq \epsilon$

Interpretation of Leverage Scores: Spectral Approximation

- Let $A = U\Sigma V^T$ be the Singular Value Decomposition of A
- S be the leverage score sampling matrix
- Approximate Matrix Multiplication for $U^T U$ i.e,

$$\|U^T S^T S U - U^T U\|_F = \|U^T S^T S U - I\|_F \leq \epsilon \tag{1}$$

Interpretation of Leverage Scores: Spectral Approximation

- Let $A = U\Sigma V^T$ be the Singular Value Decomposition of A
- S be the leverage score sampling matrix
- Approximate Matrix Multiplication for $U^T U$ i.e,

$$\|U^T S^T S U - U^T U\|_F = \|U^T S^T S U - I\|_F \leq \epsilon \tag{1}$$

- (1) implies $\sigma_{max}\left(U^T S^T S U - I\right) \leq \epsilon$
  Singular values of a symmetric matrix are the absolute values of the eigenvalues

Interpretation of Leverage Scores: Spectral Approximation

- Let $A = U\Sigma V^T$ be the Singular Value Decomposition of A
- S be the leverage score sampling matrix
- Approximate Matrix Multiplication for $U^T U$ i.e,

$$\|U^T S^T S U - U^T U\|_F = \|U^T S^T S U - I\|_F \leq \epsilon \qquad (1)$$

- (1) implies $\sigma_{max}\left(U^T S^T S U - I\right) \leq \epsilon$
  Singular values of a symmetric matrix are the absolute values of the eigenvalues
- $\max_{i=1,..d} \left|\lambda_i\Big(U^T S^T S U - I\Big)\right| \leq \epsilon$
- (1) implies $1 - \epsilon \leq \lambda_i\Big(U^T S^T S U\Big) \leq 1 + \epsilon$ for all i

Interpretation of Leverage Scores: Spectral Approximation

- Let $A = U\Sigma V^T$ be the Singular Value Decomposition of A
- S be the leverage score sampling matrix
- Approximate Matrix Multiplication for $U^T U$ i.e,

$$\|U^T S^T SU - U^T U\|_F = \|U^T S^T SU - I\|_F \leq \epsilon \qquad (1)$$

- (1) implies $\sigma_{max}\left(U^T S^T SU - I\right) \leq \epsilon$
  Singular values of a symmetric matrix are the absolute values of the eigenvalues
- $\max_{i=1,..d} \left|\lambda_i\Big(U^T S^T SU - I\Big)\right| \leq \epsilon$
- (1) implies $1 - \epsilon \leq \lambda_i\Big(U^T S^T SU\Big) \leq 1 + \epsilon$ for all i
- $(A^T S^T SA)^{-1}$ exists whenever $(A^T A)^{-1}$ exists
- sketched least squares solution
  $\arg\min_x \|SAx - Sb\|_2 = (A^T S^T SA)^{-1} S^T Sb$ is well defined

Preserving Spectral Properties

$$\|U^T S^T S U - U^T U\|_F = \|U^T S^T S U - I\|_F \leq \epsilon \tag{2}$$

- also implies that

  $$(1-\epsilon)\|Ax\|_2^2 \leq \|SAx\|_2^2 \leq (1+\epsilon)\|Ax\|_2^2$$

  for all $x \in \mathbb{R}^d$

  Johnson-Lindenstrauss embedding property for the whole subspace range(A)
- we utilized this in the basic inequality method

Interpretation of Leverage Scores: Subspace Embedding

$$\|U^T S^T SU - U^T U\|_F = \|U^T S^T SU - I\|_F \leq \epsilon$$

implies

$$(1-\epsilon)\|Ax\|_2^2 \leq \|SAx\|_2^2 \leq (1+\epsilon)\|Ax\|_2^2$$

for all $x \in \mathbb{R}^d$

- Weyl's Inquality $\left|\lambda_i(M) - \lambda_i(M')\right| \leq \sigma_{\max}\left(M - M'\right)$ for all i
- $\left|\lambda_i(A^T S^T SA) - \lambda_i(A^T A)\right| \leq \epsilon$, i.e., all eigenvalues are approximately preserved

Interpretation of Leverage Scores: Sensitivity of the loss function

- Consider $\|Ax - b\|_2^2 = \sum_i (a_i^T x - b_i)^2$
  suppose that $b = Ax^*$ for simplicity
- Consider the worst-case ratio

Fast Johnson Lindenstrauss Transform

- Let H denote the $n \times n$ Hadamard Transform matrix constructed as follows
  $H_2 := \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$
  $H_{n+1} = \begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix}$
- let D be an $n \times n$ diagonal matrix of random ± 1 uniform signs
- Uniform $m \times n$ sub-sampling matrix P scaled with $\frac{\sqrt{n}}{\sqrt{m}}$
- Let $S = \frac{1}{\sqrt{n}} PHD$.
- Note that $\mathbb{E} S^T S = I$ since $DH^T HD = nI$ and $\mathbb{E} P^T P = I$

Fast Johnson Lindenstrauss Transform Analysis

- Leverage scores of a matrix $A = U\Sigma V^T$ are given by $\ell_i = \|U^T e_i\|_2^2 = e_i^T U U^T e_i$
- Another expression: $\ell_i = e_i^T A(A^T A)^{-1} A^T e_i$

Fast Johnson Lindenstrauss Transform Analysis

- Leverage scores of a matrix $A = U\Sigma V^T$ are given by $\ell_i = \|U^T e_i\|_2^2 = e_i^T UU^T e_i$
- Another expression: $\ell_i = e_i^T A(A^T A)^{-1} A^T e_i$
- Compare with leverage scores of $\frac{1}{\sqrt{n}} HDA$ denoted by $\tilde{\ell}_i$

$$\tilde{\ell}_i := e_i^T HDA(A^T DH^T HDA)^{-1} A^T DH^T e_i \quad (3)$$

$$= \frac{1}{n} e_i^T HDA(A^T A)^{-1} A^T DH^T e_i \quad (4)$$

$$= \frac{1}{n} e_i^T HDUU^T DH^T e_i \quad (5)$$

$$= \frac{1}{n} h_i^T DUU^T Dh_i \quad (6)$$

- where we have used $H^T H = nI$
- $\tilde{\ell}_i$ is distributed as $\frac{1}{n} r^T UU^T r$ where r is i.i.d. ± 1
- $\mathbb{E} \frac{1}{n} r^T UU^T r = \frac{d}{n}$

Fast Johnson Lindenstrauss Transform Analysis

- Chernoff's method (as in Chernoff Bound) implies that

$$\mathbb{P}\left[\left|\frac{1}{n}h_i^T D u_j\right| \geq t\right] \leq 2e^{-t^2 n/2}$$

for every fixed i and j.

- Applying union bound

$$\tilde{\ell}_i = \frac{1}{n}h_i^T D U U^T D h_i \leq \text{const}\, \frac{d \log(nd)}{n}$$

with high probability

note that $\ell_i = \frac{d}{n}$ for all i when leverage scores are exactly uniform

Randomized Hadamard Transform HD preconditions leverage scores

Apply HD to data A

- $PHDA$ is a uniformly subsampled version HDA

Leverage scores of $\frac{1}{\sqrt{n}}HDU$ are near uniform

uniform sampling $\frac{1}{\sqrt{n}}HDA$ works!

in other works SA where $S = \frac{1}{\sqrt{n}}PHD$ is a subspace embedding

Questions?