Algorithms for Data Science

Anil Maheshwari

anil@scs.carleton.ca
School of Computer Science
Carleton University

Outline

Topics

Required Background

Blended Teaching

Course Evaluation

Help Me

Topics

Algorithms for **Data Science**

Input	Algorithm	Output
01000011		
10100001		
01110010		
01101010	$\Rightarrow$	?
01100101		
01110100		
01101111		
01101110		

Topics

1. Probability Basics: Linearity of Expectation, Concentration Bounds, Balls and Bins, Hashing.
2. Data Streaming - CMS, Estimating Frequency Moments, What is Trending?
3. Locality-Sensitive Hashing
4. Geometric Approximation (Nearest Neighbors, Core Sets)
5. Dimensionality Reduction
6. Online Algorithms: Bipartite Matching, Adwords
7. Regret Minimization (Multiplicative Weight Algorithm)
8. Graph Partitioning
9. Linear Algebra: Eigenvalues + SVD
10. Recommendation Systems
11. Randomized Linear Algebra
12. . . .

Topics

1. Probability Basics: Linearity of Expectation, Concentration Bounds, Balls and Bins, Hashing.
2. Data Streaming - CMS, Estimating Frequency Moments, What is Trending?
3. Locality-Sensitive Hashing
4. Geometric Approximation (Nearest Neighbors, Core Sets)
5. Dimensionality Reduction
6. Online Algorithms: Bipartite Matching, Adwords
7. Regret Minimization (Multiplicative Weight Algorithm)
8. Graph Partitioning
9. Linear Algebra: Eigenvalues + SVD
10. Recommendation Systems
11. Randomized Linear Algebra
12. . . .

Topics

1. Probability Basics: Linearity of Expectation, Concentration Bounds, Balls and Bins, Hashing.
2. Data Streaming - CMS, Estimating Frequency Moments, What is Trending?
3. Locality-Sensitive Hashing
4. Geometric Approximation (Nearest Neighbors, Core Sets)
5. Dimensionality Reduction
6. Online Algorithms: Bipartite Matching, Adwords
7. Regret Minimization (Multiplicative Weight Algorithm)
8. Graph Partitioning
9. Linear Algebra: Eigenvalues + SVD
10. Recommendation Systems
11. Randomized Linear Algebra
12. . . .

Topics

1. Probability Basics: Linearity of Expectation, Concentration Bounds, Balls and Bins, Hashing.
2. Data Streaming - CMS, Estimating Frequency Moments, What is Trending?
3. Locality-Sensitive Hashing
4. Geometric Approximation (Nearest Neighbors, Core Sets)
5. Dimensionality Reduction
6. Online Algorithms: Bipartite Matching, Adwords
7. Regret Minimization (Multiplicative Weight Algorithm)
8. Graph Partitioning
9. Linear Algebra: Eigenvalues + SVD
10. Recommendation Systems
11. Randomized Linear Algebra
12. . . .

Topics

1. Probability Basics: Linearity of Expectation, Concentration Bounds, Balls and Bins, Hashing.
2. Data Streaming - CMS, Estimating Frequency Moments, What is Trending?
3. Locality-Sensitive Hashing
4. Geometric Approximation (Nearest Neighbors, Core Sets)
5. Dimensionality Reduction
6. Online Algorithms: Bipartite Matching, Adwords
7. Regret Minimization (Multiplicative Weight Algorithm)
8. Graph Partitioning
9. Linear Algebra: Eigenvalues + SVD
10. Recommendation Systems
11. Randomized Linear Algebra
12. . . .

Topics

1. Probability Basics: Linearity of Expectation, Concentration Bounds, Balls and Bins, Hashing.
2. Data Streaming - CMS, Estimating Frequency Moments, What is Trending?
3. Locality-Sensitive Hashing
4. Geometric Approximation (Nearest Neighbors, Core Sets)
5. Dimensionality Reduction
6. Online Algorithms: Bipartite Matching, Adwords
7. Regret Minimization (Multiplicative Weight Algorithm)
8. Graph Partitioning
9. Linear Algebra: Eigenvalues + SVD
10. Recommendation Systems
11. Randomized Linear Algebra
12. . . .

Topics

1. Probability Basics: Linearity of Expectation, Concentration Bounds, Balls and Bins, Hashing.
2. Data Streaming - CMS, Estimating Frequency Moments, What is Trending?
3. Locality-Sensitive Hashing
4. Geometric Approximation (Nearest Neighbors, Core Sets)
5. Dimensionality Reduction
6. Online Algorithms: Bipartite Matching, Adwords
7. Regret Minimization (Multiplicative Weight Algorithm)
8. Graph Partitioning
9. Linear Algebra: Eigenvalues + SVD
10. Recommendation Systems
11. Randomized Linear Algebra
12. . . .

Topics

1. Probability Basics: Linearity of Expectation, Concentration Bounds, Balls and Bins, Hashing.
2. Data Streaming - CMS, Estimating Frequency Moments, What is Trending?
3. Locality-Sensitive Hashing
4. Geometric Approximation (Nearest Neighbors, Core Sets)
5. Dimensionality Reduction
6. Online Algorithms: Bipartite Matching, Adwords
7. Regret Minimization (Multiplicative Weight Algorithm)
8. Graph Partitioning
9. Linear Algebra: Eigenvalues + SVD
10. Recommendation Systems
11. Randomized Linear Algebra
12. . . .

Topics

1. Probability Basics: Linearity of Expectation, Concentration Bounds, Balls and Bins, Hashing.
2. Data Streaming - CMS, Estimating Frequency Moments, What is Trending?
3. Locality-Sensitive Hashing
4. Geometric Approximation (Nearest Neighbors, Core Sets)
5. Dimensionality Reduction
6. Online Algorithms: Bipartite Matching, Adwords
7. Regret Minimization (Multiplicative Weight Algorithm)
8. Graph Partitioning
9. Linear Algebra: Eigenvalues + SVD
10. Recommendation Systems
11. Randomized Linear Algebra
12. . . .

Topics

1. Probability Basics: Linearity of Expectation, Concentration Bounds, Balls and Bins, Hashing.
2. Data Streaming - CMS, Estimating Frequency Moments, What is Trending?
3. Locality-Sensitive Hashing
4. Geometric Approximation (Nearest Neighbors, Core Sets)
5. Dimensionality Reduction
6. Online Algorithms: Bipartite Matching, Adwords
7. Regret Minimization (Multiplicative Weight Algorithm)
8. Graph Partitioning
9. Linear Algebra: Eigenvalues + SVD
10. Recommendation Systems
11. Randomized Linear Algebra
12. . . .

Topics

1. Probability Basics: Linearity of Expectation, Concentration Bounds, Balls and Bins, Hashing.
2. Data Streaming - CMS, Estimating Frequency Moments, What is Trending?
3. Locality-Sensitive Hashing
4. Geometric Approximation (Nearest Neighbors, Core Sets)
5. Dimensionality Reduction
6. Online Algorithms: Bipartite Matching, Adwords
7. Regret Minimization (Multiplicative Weight Algorithm)
8. Graph Partitioning
9. Linear Algebra: Eigenvalues + SVD
10. Recommendation Systems
11. Randomized Linear Algebra
12. . . .

Topics

1. Probability Basics: Linearity of Expectation, Concentration Bounds, Balls and Bins, Hashing.
2. Data Streaming - CMS, Estimating Frequency Moments, What is Trending?
3. Locality-Sensitive Hashing
4. Geometric Approximation (Nearest Neighbors, Core Sets)
5. Dimensionality Reduction
6. Online Algorithms: Bipartite Matching, Adwords
7. Regret Minimization (Multiplicative Weight Algorithm)
8. Graph Partitioning
9. Linear Algebra: Eigenvalues + SVD
10. Recommendation Systems
11. Randomized Linear Algebra
12. . . .

Our Focus

Warnings: First Offering - Online - Plans May Derail!
Office may become inaccessible

We will try to understand

- Algorithmic Techniques
- Key Ideas
- Learn bit of Math
- High-level details

Required Background

Background

- ## Basic Data Structures

  Lists, Stacks, BST, Hashing, Searching and Sorting

  See Pat Morin's `https://opendatastructures.org/`

- ## Basic Probability Theory

  Expectation, Variance, Concentration Bounds

  Michiel Smid's COMP 2804 Notes `http://cglab.ca/~michiel/DiscreteStructures/`

  Joe Blitzstein `https://projects.iq.harvard.edu/stat110/youtube`

- ## Linear Algebra

  Eigenvalues and Eigenvectors, Null Space, Positive Definitive Matrices, $Q\Lambda Q^T$, SVD

  Gilbert Strang

  `https://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/video-lectures/`

- ## Algorithmic Techniques

  Pseudocode, Correctness, Analysis (Recurrences + O, Ω-notation), Analysis of (randomized) Quicksort. Algorithms for Connectivity, Strong Connectivity, MST, SSSP, Graph Traversal. Techniques: D& C, Greedy, Dynamic Programming.

Blended Teaching

Online Course

- Lectures (with recordings) on Tuesdays & Fridays from 06:30 AM EST.
- Timing may change when Day Light Saving Hours apply in March.
- E-mail: anil@scs.carleton.ca
- Post queries during the lectures using the zoom chat.

References

Useful References

1. Foundations of Data Science by Blum, Hopcroft and Kannan.

   https://www.cs.cornell.edu/jeh/book.pdf

2. Mining of Massive Data Sets http://www.mmds.org/
3. Notes on Algorithm Design on my web-page

   http://people.scs.carleton.ca/~maheshwa/

4. Numerous research articles.
5. Pointers to useful Videos.

Course Evaluation

Evaluation Parameters

Principles

- Open-ended course by design
- Contents evolve
- Enjoyable learning experience
- Use zoom chat effectively

Evaluation Components

- Assignments
- Programming Assignments
- Tests
- On the spot quizzes!

Help Me

PLEASE, PLEASE, PLEASE, PLEASE

1. Post your questions on Zoom
2. Post useful links/references/ideas
3. Take initiative to learn
4. E-mail now if you want to learn some specific algorithmic technique(s)
5. Seek Help, Have Fun & Smile

Season - SUMMER

Season - FALL

Season - WINTER