COL866: Foundations of Data Science

Ragesh Jaiswal, IITD

Ranking and Social Choice

- Problem: Merge multiple ranked lists in a meaningful manner.
- Here is a simple example that brings the difficulty of such a task.

Individual	rank 1	rank 2	rank3
1	a	b	c
2	b	c	a
3	c	a	b

Ranking and Social Choice

- Problem: Merge multiple ranked lists in a meaningful manner.
- Here is a simple example that brings the difficulty of such a task.

Individual	rank 1	rank 2	rank3
1	a	b	c
2	b	c	a
3	c	a	b

- Is a ranked higher than b?
- Is b ranked higher than c?
- Is a ranked higher than c?

Ranking and Social Choice

- Problem: Merge multiple ranked lists in a meaningful manner.
- Here is a simple example that brings the difficulty of such a task.

Individual	rank 1	rank 2	rank3
1	a	b	c
2	b	c	a
3	c	a	b

- Is a ranked higher than b? yes since two people prefer a
- Is b ranked higher than c? yes since two people prefer b
- Is a ranked higher than c? no since two people prefer c
- So, such a task of combining individual rankings to come up with global ranking might be difficult in general. It would be great if we could argue this in general.

Ranking and Social Choice

- Problem: Merge multiple ranked lists in a meaningful manner.
- Here is a simple example that brings the difficulty of such a task.

Individual	rank 1	rank 2	rank3
1	a	b	c
2	b	c	a
3	c	a	b

- Is a ranked higher than b? yes since two people prefer a
- Is b ranked higher than c? yes since two people prefer b
- Is a ranked higher than c? no since two people prefer c
- So, such a task of combining individual rankings to come up with global ranking might be difficult in general. It would be great if we could argue this in general.
- For such an argument we need to fix the axioms of ranking, or some basic conditions that a global ranking should satisfy.

Ranking and Social Choice

- Problem: Merge multiple ranked lists in a meaningful manner.
- Axioms of ranking: The method of producing a global ranking should satisfy the following:
  - Nondictatorship: The algorithm cannot always select one individual's ranking as the global ranking.
  - Unanimity: If every individual prefers a to b, then the global ranking should prefer a to b.
  - Independent of irrelevant alternatives: If individuals modify their rankings but keep the order of a and b unchanged, then the global order of a and b should not change.
- We will argue that it is not possible to satisfy all three axioms simultaneously (Arrow's Theorem).
- We start with a lemma.

- Problem: Merge multiple ranked lists in a meaningful manner.
- Axioms of ranking: The method of producing a global ranking should satisfy the following:
  - Nondictatorship: The algorithm cannot always select one individual's ranking as the global ranking.
  - Unanimity: If every individual prefers a to b, then the global ranking should prefer a to b.
  - Independent of irrelevant alternatives: If individuals modify their rankings but keep the order of a and b unchanged, then the global order of a and b should not change.
- We will argue that it is not possible to satisfy all three axioms simultaneously (Arrow's Theorem).
- We start with a lemma.

Lemma

For any set of rankings in which each individual ranks an item first or last, a global ranking satisfying the three axioms must put b first or last.

- Problem: Merge multiple ranked lists in a meaningful manner.
- Axioms of ranking: The method of producing a global ranking should satisfy the following:
  - Nondictatorship: The algorithm cannot always select one individual's ranking as the global ranking.
  - Unanimity: If every individual prefers a to b, then the global ranking should prefer a to b.
  - Independent of irrelevant alternatives: If individuals modify their rankings but keep the order of a and b unchanged, then the global order of a and b should not change.

Lemma

For any set of rankings in which each individual ranks an item first or last, a global ranking satisfying the three axioms must put b first or last.

Theorem (Arrow's impossibility theorem)

Any deterministic algorithm for creating a global ranking from individual rankings of three or more elements in which the global ranking satisfies unanimity and independence of irrelevant alternatives is a dictatorship.

- Problem: Merge multiple ranked lists in a meaningful manner.
- Axioms of ranking: The method of producing a global ranking should satisfy the following:
  - Nondictatorship: The algorithm cannot always select one individual's ranking as the global ranking.
  - Unanimity: If every individual prefers a to b, then the global ranking should prefer a to b.
  - Independent of irrelevant alternatives: If individuals modify their rankings but keep the order of a and b unchanged, then the global order of a and b should not change.

Theorem (Arrow's impossibility theorem)

Any deterministic algorithm for creating a global ranking from individual rankings of three or more elements in which the global ranking satisfies unanimity and independence of irrelevant alternatives is a dictatorship.

- Example: Borda count
  - Each item gets points from an individual in reverse order of the ranking. The global ranking is done based on the total number of points received.
  - Give an example in which independence of irrelevant alternatives fails.

- Problem: Merge multiple ranked lists in a meaningful manner.
- Axioms of ranking: The method of producing a global ranking should satisfy the following:
  - Nondictatorship: The algorithm cannot always select one individual's ranking as the global ranking.
  - Unanimity: If every individual prefers a to b, then the global ranking should prefer a to b.
  - Independent of irrelevant alternatives: If individuals modify their rankings but keep the order of a and b unchanged, then the global order of a and b should not change.

Theorem (Arrow's impossibility theorem)

Any deterministic algorithm for creating a global ranking from individual rankings of three or more elements in which the global ranking satisfies unanimity and independence of irrelevant alternatives is a dictatorship.

- Example: Borda count
  - Each item gets points from an individual in reverse order of the ranking. The global ranking is done based on the total number of points received.
  - Here is an example in which independence of irrelevant alternatives fails:

Individual	Ranking
1	abcd
2	abcd
3	bacd

Table: Individual 3 changing his ranking to bcda, changes the global ranking.

Compressed Sensing and Sparse Vectors

Compressed Sensing and Sparse Vectors

- A signal in the current context is a vector $\mathbf{x}$ of length d and a measurement of signal $\mathbf{x}$ is taking the dot product of $\mathbf{x}$ with a known vector $\mathbf{a}_i$.
- Claim: For uniquely reconstructing $\mathbf{x}$ without any assumptions, d linearly independent measurements are necessary and sufficient.
  - Given $A\mathbf{x} = \mathbf{b}$, solve for $\mathbf{x}$ by computing $\mathbf{x} = A^{-1}\mathbf{b}$.
- If there are fewer than d measurements and A has rank $< d$, there may be multiple solutions.
- Informal claim: If $\mathbf{x}$ is sparse with $s << d$ non-zero elements, then we might be able to reconstruct $\mathbf{x}$ with far fewer measurements.
- This is popularly known as compressed sensing and has applications in photography (where it reduces the number of sensors) and magnetic resonance imaging.

- Sparse vector: A vector $\mathbf{x} \in \mathbb{R}^d$ is said to be s-sparse if it has at most $s \leq d$ non-zero elements.
- Let us examine the conditions under which $A\mathbf{x} = \mathbf{b}$ has a unique sparse solution. The matrix A is an $n \times d$ matrix with $n < d$.
- Claim 1: Suppose there are two s-sparse solutions $\mathbf{x}_1$ and $\mathbf{x}_2$. Then $\mathbf{x}_1 - \mathbf{x}_2$ will be a $2s$-sparse solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$.

- Sparse vector: A vector $\mathbf{x} \in \mathbb{R}^d$ is said to be s-sparse if it has at most $s \leq d$ non-zero elements.
- Let us examine the conditions under which $A\mathbf{x} = \mathbf{b}$ has a unique sparse solution. The matrix A is an $n \times d$ matrix with $n < d$.
- Claim 1: Suppose there are two s-sparse solutions $\mathbf{x}_1$ and $\mathbf{x}_2$. Then $\mathbf{x}_1 - \mathbf{x}_2$ will be a $2s$-sparse solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$.
- Claim 2: Existence of a $2s$-sparse solution to $A\mathbf{x} = \mathbf{0}$ implies the existence of $2s$ columns of A that are linearly dependent.
- Combining claims 1 and 2, we get that if no $2s$ columns of A are linearly dependent, then there can only be one s-sparse solutions to $A\mathbf{x} = \mathbf{b}$.

- Sparse vector: A vector $\mathbf{x} \in \mathbb{R}^d$ is said to be s-sparse if it has at most $s \leq d$ non-zero elements.
- Let us examine the conditions under which $A\mathbf{x} = \mathbf{b}$ has a unique sparse solution. The matrix A is an $n \times d$ matrix with $n < d$.
- Claim 1: Suppose there are two s-sparse solutions $\mathbf{x}_1$ and $\mathbf{x}_2$. Then $\mathbf{x}_1 - \mathbf{x}_2$ will be a $2s$-sparse solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$.
- Claim 2: Existence of a $2s$-sparse solution to $A\mathbf{x} = \mathbf{0}$ implies the existence of $2s$ columns of A that are linearly dependent.
- Combining claims 1 and 2, we get that if no $2s$ columns of A are linearly dependent, then there can only be one s-sparse solutions to $A\mathbf{x} = \mathbf{b}$.
- Consider the $2s \times d$ matrix A constructed as follows:

  Select each entry of A independently from the standard Gaussian.

- Claim 3: With probability 1, no $2s$ columns of A constructed above are linearly dependent.

- Sparse vector: A vector $\mathbf{x} \in \mathbb{R}^d$ is said to be s-sparse if it has at most $s \leq d$ non-zero elements.
- Let us examine the conditions under which $A\mathbf{x} = \mathbf{b}$ has a unique sparse solution. The matrix A is an $n \times d$ matrix with $n < d$.
- Claim 1: Suppose there are two s-sparse solutions $\mathbf{x}_1$ and $\mathbf{x}_2$. Then $\mathbf{x}_1 - \mathbf{x}_2$ will be a $2s$-sparse solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$.
- Claim 2: Existence of a $2s$-sparse solution to $A\mathbf{x} = \mathbf{0}$ implies the existence of $2s$ columns of A that are linearly dependent.
- Combining claims 1 and 2, we get that if no $2s$ columns of A are linearly dependent, then there can only be one s-sparse solutions to $A\mathbf{x} = \mathbf{b}$.
- Consider the $2s \times d$ matrix A constructed as follows:

  Select each entry of A independently from the standard Gaussian.

- Claim 3: With probability 1, no $2s$ columns of A constructed above are linearly dependent.
- So, for matrix A constructed above $A\mathbf{x} = \mathbf{b}$ has a unique s-sparse solution.
- Question: How do we obtain the s-sparse solution? Think brute-force.

Compressed Sensing and Sparse Vectors
Unique reconstruction of a space vector

- Sparse vector: A vector $\mathbf{x} \in \mathbb{R}^d$ is said to be s-sparse if it has at most $s \leq d$ non-zero elements.
- Let us examine the conditions under which $A\mathbf{x} = \mathbf{b}$ has a unique sparse solution. The matrix A is an $n \times d$ matrix with $n < d$.
- Claim 1: Suppose there are two s-sparse solutions $\mathbf{x}_1$ and $\mathbf{x}_2$. Then $\mathbf{x}_1 - \mathbf{x}_2$ will be a $2s$-sparse solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$.
- Claim 2: Existence of a $2s$-sparse solution to $A\mathbf{x} = \mathbf{0}$ implies the existence of $2s$ columns of A that are linearly dependent.
- Combining claims 1 and 2, we get that if no $2s$ columns of A are linearly dependent, then there can only be one s-sparse solutions to $A\mathbf{x} = \mathbf{b}$.
- Consider the $2s \times d$ matrix A constructed as follows:

  Select each entry of A independently from the standard Gaussian.

- Claim 3: With probability 1, no $2s$ columns of A constructed above are linearly dependent.
- So, for matrix A constructed above $A\mathbf{x} = \mathbf{b}$ has a unique s-sparse solution.
- Question: How do we obtain the s-sparse solution? Think brute-force.
  - Try all possible $\binom{d}{s}$ locations for non-zero elements in $\mathbf{x}$ and solve $A\mathbf{x} = \mathbf{b}$. Unfortunately, this takes $\Omega(d^s)$ time.

- Sparse vector: A vector $\mathbf{x} \in \mathbb{R}^d$ is said to be s-sparse if it has at most $s \leq d$ non-zero elements.
- Let us examine the conditions under which $A\mathbf{x} = \mathbf{b}$ has a unique sparse solution. The matrix A is an $n \times d$ matrix with $n < d$.
- Claim 1: Suppose there are two s-sparse solutions $\mathbf{x}_1$ and $\mathbf{x}_2$. Then $\mathbf{x}_1 - \mathbf{x}_2$ will be a $2s$-sparse solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$.
- Claim 2: Existence of a $2s$-sparse solution to $A\mathbf{x} = \mathbf{0}$ implies the existence of $2s$ columns of A that are linearly dependent.
- Combining claims 1 and 2, we get that if no $2s$ columns of A are linearly dependent, then there can only be one s-sparse solutions to $A\mathbf{x} = \mathbf{b}$.
- Consider the $2s \times d$ matrix A constructed as follows:

  Select each entry of A independently from the standard Gaussian.

- Claim 3: With probability 1, no $2s$ columns of A constructed above are linearly dependent.
- So, for matrix A constructed above $A\mathbf{x} = \mathbf{b}$ has a unique s-sparse solution.
- Question: How do we obtain the s-sparse solution? Yes in $\Omega(d^s)$ time.
- Question: Can we find a sparse solution efficiently?

- Finding a sparse solution to $A\mathbf{x} = \mathbf{b}$ can be written as the following program:

$$\text{minimize } ||\mathbf{x}||_0$$
$$\text{subject to: } A\mathbf{x} = \mathbf{b}$$

- Unfortunately, this is not a convex program. Instead, the next program is a convex program. In fact, it can written as a linear program.

$$\text{minimize } ||\mathbf{x}||_1$$
$$\text{subject to: } A\mathbf{x} = \mathbf{b}$$

- Claim 1: The following linear program is equivalent to the above program.

$$\text{minimize } \sum_i u_i + \sum_i v_i$$
$$\text{subject to: } A\mathbf{u} - A\mathbf{v} = \mathbf{b}, \mathbf{u} \geq 0, \mathbf{v} \geq 0$$

- Finding a sparse solution to $A\mathbf{x} = \mathbf{b}$ can be written as the following program:

$$\text{minimize } ||\mathbf{x}||_0$$
$$\text{subject to: } A\mathbf{x} = \mathbf{b}$$

- Unfortunately, this is not a convex program. Instead, the next program is a convex program. In fact, it can written as a linear program.

$$\text{minimize } ||\mathbf{x}||_1$$
$$\text{subject to: } A\mathbf{x} = \mathbf{b}$$

- Claim 1: The following linear program is equivalent to the above program.

$$\text{minimize } \sum_i u_i + \sum_i v_i$$
$$\text{subject to: } A\mathbf{u} - A\mathbf{v} = \mathbf{b}, \mathbf{u} \geq 0, \mathbf{v} \geq 0$$

- Question: How does solving the above program help in finding a sparse solution to $A\mathbf{x} = \mathbf{b}$?

- Finding a sparse solution to $A\mathbf{x} = \mathbf{b}$ can be written as the following program:

$$\text{minimize } ||\mathbf{x}||_0$$
$$\text{subject to: } A\mathbf{x} = \mathbf{b}$$

- Unfortunately, this is not a convex program. Instead, the next program is a convex program. In fact, it can written as a linear program.

$$\text{minimize } ||\mathbf{x}||_1$$
$$\text{subject to: } A\mathbf{x} = \mathbf{b}$$

- Claim 1: The following linear program is equivalent to the above program.

$$\text{minimize } \sum_i u_i + \sum_i v_i$$
$$\text{subject to: } A\mathbf{u} - A\mathbf{v} = \mathbf{b}, \mathbf{u} \geq 0, \mathbf{v} \geq 0$$

- Question: How does solving the above program help in finding a sparse solution to $A\mathbf{x} = \mathbf{b}$?
  - If A is of a specific form, then the solution to the program gives a sparse solution.

- Program **P**:

$$\text{minimize } ||\mathbf{x}||_1$$
$$\text{subject to: } A\mathbf{x} = \mathbf{b}$$

- Question: How does solving the above program help in finding a sparse solution to $A\mathbf{x} = \mathbf{b}$?
  - If A is of a specific form, then the solution to the program gives a sparse solution.
- The following theorem states the conditions for matrix A under which the solution to **P** is an s-sparse solution $A\mathbf{x} = \mathbf{b}$.

Theorem

*If matrix A has unit-length columns $\mathbf{a}_1, ..., \mathbf{a}_d$ and the property that $|\mathbf{a}_i^T \mathbf{a}_j| < \frac{1}{2s}$ for all $i \neq j$, then if the equation $A\mathbf{x} = \mathbf{b}$ has a solution with at most s non-zero coordinates, this solution is the unique 1-norm solution to $A\mathbf{x} = \mathbf{b}$ (i.e., solution to program **P**).*

- Program **P**:

$$\text{minimize } ||\mathbf{x}||_1$$
$$\text{subject to: } A\mathbf{x} = \mathbf{b}$$

- Question: How does solving the above program help in finding a sparse solution to $A\mathbf{x} = \mathbf{b}$?
  - If A is of a specific form, then the solution to the program gives a sparse solution.
- The following theorem states the conditions for matrix A under which the solution to **P** is an s-sparse solution $A\mathbf{x} = \mathbf{b}$.

Theorem

*If matrix A has unit-length columns $\mathbf{a}_1, ..., \mathbf{a}_d$ and the property that $|\mathbf{a}_i^T \mathbf{a}_j| < \frac{1}{2s}$ for all $i \neq j$, then if the equation $A\mathbf{x} = \mathbf{b}$ has a solution with at most s non-zero coordinates, this solution is the unique 1-norm solution to $A\mathbf{x} = \mathbf{b}$ (i.e., solution to program **P**).*

- Such a matrix can be constructed efficiently using concepts developed in high dimensional geometry. The next theorem summarises everything.

Theorem

For some absolute constant c, if A has n rows for $n \geq cs^2 \log d$ and each column of A is chosen to be a random unit-length n-dimensional vector, then with high probability A satisfies the conditions of previous theorem and therefore if the equation $A\mathbf{x} = \mathbf{b}$ has a solution with at most s non-zero coordinates, this solution is the unique minimum 1-norm solution to $A\mathbf{x} = \mathbf{b}$.

Theorem

If matrix A has unit-length columns $\mathbf{a}_1, ..., \mathbf{a}_d$ and the property that $|\mathbf{a}_i^T \mathbf{a}_j| < \frac{1}{2s}$ for all $i \neq j$, then if the equation $A\mathbf{x} = \mathbf{b}$ has a solution with at most s non-zero coordinates, this solution is the unique 1-norm solution to $A\mathbf{x} = \mathbf{b}$ (i.e., solution to program $\mathbf{P}$).

Proof sketch

- Claim: Let $\mathbf{x}_0$ denote the unique s-sparse solution to $A\mathbf{x} = \mathbf{b}$ and let $\mathbf{x}_1$ be a solution of smallest possible 1-norm. Let $\mathbf{z} = \mathbf{x}_1 - \mathbf{x}_0$. Then $\mathbf{z} = \mathbf{0}$.

End