Render

```
โจทย์ข้อนี้กำหนดให้มีหออยู่ $N$ หอ โดยแต่ละหอมีพิกัด $x_i$ กับ $y_i$ และมีนักเรียน $s_i$ คน โจทย์ต้องการให้หาว่าหากเลือกพิกัด $x$ $y$ ให้ดีที่สุดจะทำให้ $\Sigma_i s_i (| x_i - x| + |y_i - y|)$ ได้ต่ำสุดเท่าไหร่

สำหรับข้อนี้สามารถสังเกตว่า $\Sigma_i s_i * (| x_i - x| + |y_i - y|) = \Sigma_i s_i | x_i - x|  + \Sigma_i s_i |y_i - y|$ จึงสามารถพิจารณาระยะทางในพิกัด $x$ กับพิกัด $y$ แยกกันในการหาคำตอบ

### ค่ามัธยฐาน

คุณสมบัติที่สำคัญที่ต้องใช้การแก้ข้อนี้คือ $\Sigma_{i=1}^N | a_i - a|$ จะมีค่าต่ำสุดเมื่อเลือก $a$ เป็นค่ามัธยฐานของ $(a_1, a_2, \dots, a_N)$ โดยนิยามค่ามัธยฐาน $m$ เป็นค่าที่ทำให้มีจำนวน $a_i$ ที่ไม่น้อยกว่า $m$ อย่างน้อยครึ่ง และจำนวน $a_i$ ที่มีค่าไม่เกิน $m$ อย่างน้อยครึ่ง (ตามนิยามนี้อาจมีค่ามัธยฐานมากกว่า 1 ค่า เช่น สำหรับ $1, 2, 3, 4, 5, 6$ ทั้ง $3, 4$ และ $3.5$ ต่างเป็นค่ามัธยฐาน)

เนื่องจากข้อนี้ยังมีค่า $s_i$ ที่ต้องพิจารณาด้วยเราจะพิสูจน์ว่าพิกัด $x$ ที่ต้องการคือค่า $m$ ที่ทำให้ $\Sigma_{i,m \leq x_i} s_i \geq \frac{\Sigma_i s_i}{2}$ และ $\Sigma_{i,m \geq x_i} s_i \geq \frac{\Sigma_i s_i}{2}$ กล่าวคือนักเรียนที่อยู่ที่พิกัด $x$ ที่ไม่น้อยกว่า $m$ มีอย่างน้อยครึ่ง และไม่มากกว่า $m$ อย่างน้อยครึ่ง

สำหรับการหา $\Sigma_i s_i | x_i - m|$  สังเกตว่า $\Sigma_i s_i | x_i - m| = \Sigma_{i,m \geq x_i} s_i (m - x_i) +\Sigma_{i,m < x_i} s_i (x_i -m) $ สมมติว่า $\Sigma_{i,m \geq x_i} s_i < \Sigma_{i,m < x_i} s_i $ จะได้ว่าถ้าเพิ่มค่า $m$ เป็น $m + \epsilon$ โดยที่ $0< \epsilon < \min_{x_i<m} (m-x_i)$ จะทำให้ผลรวมนี้ลดลง เนื่องจาก $\Sigma_{i,m \geq x_i} s_i (m  + \epsilon - x_i) +\Sigma_{i,m < x_i} s_i (x_i -m - \epsilon) = \Sigma_i s_i | x_i - m|  + \epsilon (\Sigma_{i,m \geq x_i} s_i - \Sigma_{i,m < x_i} s_i) < \Sigma_i s_i | x_i - m| $ (เลือก $ \epsilon < \min_{x_i<m} (m-x_i)$ เพราะจะทำให้ทุก $x_i$ ยังอยู่ในทางซ้ายหรือขวาของ $m$ อยู่ฝั่งเดิมเมื่อเปลี่ยนเป็น $m + \epsilon$)

ดังนั้นสำหรับ $m$ ที่ทำให้ผลรวมที่ต้องการมีค่าต่ำสุดจะต้องได้ว่า $\Sigma_{i,m \geq x_i} s_i \geq \Sigma_{i,m < x_i} s_i \implies \Sigma_{i,m \geq x_i} s_i + \Sigma_{i,m \geq x_i} s_i \geq \Sigma_{i,m < x_i} s_i  + \Sigma_{i,m \geq x_i} s_i = \Sigma_i s_i \implies  \Sigma_{i,m \geq x_i} s_i \geq \frac{\Sigma_i s_i}{2}$ ตามที่ต้องการ เราสามารถพิสูจน์ด้วยวิธีเดียวกันว่าจะต้องได้ว่า $\Sigma_{i,m \leq x_i} s_i \geq \frac{\Sigma_i s_i}{2}$  เช่นกัน

โดยสรุปคือเราต้องการเลือกพิกัด $x=m$ ที่ทำให้ และ $\Sigma_{i,m \geq x_i} s_i \geq \frac{\Sigma_i s_i}{2}$ และ $\Sigma_{i,m \leq x_i} s_i \geq \frac{\Sigma_i s_i}{2}$ กล่าวคือต้องการเลือกค่ามัธยฐานของพิกัด $x$ ของนักเรียนทุกคนนั่นเองซึ่งพิกัด $x$ ที่เป็นอันดับที่ $\lfloor \frac{\Sigma_i s_i}{2} \rfloor$ จะเข้าทั้งสองเงื่อนไขไม่ว่า $\Sigma_i s_i$ จะเป็นจำนวนคู่หรือจำนวนคี่

การเลือกค่ามัธยฐานโดยตรงด้วยการ sort ทั้ง $x$ และ $y$ และนับ
```

โจทย์ข้อนี้กำหนดให้มีหออยู่ $N$ หอ โดยแต่ละหอมีพิกัด $x_i$ กับ $y_i$ และมีนักเรียน $s_i$ คน โจทย์ต้องการให้หาว่าหากเลือกพิกัด $x$ $y$ ให้ดีที่สุดจะทำให้ $\Sigma_i s_i (| x_i - x| + |y_i - y|)$ ได้ต่ำสุดเท่าไหร่

สำหรับข้อนี้สามารถสังเกตว่า $\Sigma_i s_i * (| x_i - x| + |y_i - y|) = \Sigma_i s_i | x_i - x| + \Sigma_i s_i |y_i - y|$ จึงสามารถพิจารณาระยะทางในพิกัด $x$ กับพิกัด $y$ แยกกันในการหาคำตอบ

### ค่ามัธยฐาน

คุณสมบัติที่สำคัญที่ต้องใช้การแก้ข้อนี้คือ $\Sigma_{i=1}^N | a_i - a|$ จะมีค่าต่ำสุดเมื่อเลือก $a$ เป็นค่ามัธยฐานของ $(a_1, a_2, \dots, a_N)$ โดยนิยามค่ามัธยฐาน $m$ เป็นค่าที่ทำให้มีจำนวน $a_i$ ที่ไม่น้อยกว่า $m$ อย่างน้อยครึ่ง และจำนวน $a_i$ ที่มีค่าไม่เกิน $m$ อย่างน้อยครึ่ง (ตามนิยามนี้อาจมีค่ามัธยฐานมากกว่า 1 ค่า เช่น สำหรับ $1, 2, 3, 4, 5, 6$ ทั้ง $3, 4$ และ $3.5$ ต่างเป็นค่ามัธยฐาน)

เนื่องจากข้อนี้ยังมีค่า $s_i$ ที่ต้องพิจารณาด้วยเราจะพิสูจน์ว่าพิกัด $x$ ที่ต้องการคือค่า $m$ ที่ทำให้ $\Sigma_{i,m \leq x_i} s_i \geq \frac{\Sigma_i s_i}{2}$ และ $\Sigma_{i,m \geq x_i} s_i \geq \frac{\Sigma_i s_i}{2}$ กล่าวคือนักเรียนที่อยู่ที่พิกัด $x$ ที่ไม่น้อยกว่า $m$ มีอย่างน้อยครึ่ง และไม่มากกว่า $m$ อย่างน้อยครึ่ง

สำหรับการหา $\Sigma_i s_i | x_i - m|$ สังเกตว่า $\Sigma_i s_i | x_i - m| = \Sigma_{i,m \geq x_i} s_i (m - x_i) +\Sigma_{i,m < x_i} s_i (x_i -m)$ สมมติว่า $\Sigma_{i,m \geq x_i} s_i < \Sigma_{i,m < x_i} s_i$ จะได้ว่าถ้าเพิ่มค่า $m$ เป็น $m + \epsilon$ โดยที่ $0< \epsilon < \min_{x_i<m} (m-x_i)$ จะทำให้ผลรวมนี้ลดลง เนื่องจาก $\Sigma_{i,m \geq x_i} s_i (m + \epsilon - x_i) +\Sigma_{i,m < x_i} s_i (x_i -m - \epsilon) = \Sigma_i s_i | x_i - m| + \epsilon (\Sigma_{i,m \geq x_i} s_i - \Sigma_{i,m < x_i} s_i) < \Sigma_i s_i | x_i - m|$ (เลือก $\epsilon < \min_{x_i<m} (m-x_i)$ เพราะจะทำให้ทุก $x_i$ ยังอยู่ในทางซ้ายหรือขวาของ $m$ อยู่ฝั่งเดิมเมื่อเปลี่ยนเป็น $m + \epsilon$)

ดังนั้นสำหรับ $m$ ที่ทำให้ผลรวมที่ต้องการมีค่าต่ำสุดจะต้องได้ว่า $\Sigma_{i,m \geq x_i} s_i \geq \Sigma_{i,m < x_i} s_i \implies \Sigma_{i,m \geq x_i} s_i + \Sigma_{i,m \geq x_i} s_i \geq \Sigma_{i,m < x_i} s_i + \Sigma_{i,m \geq x_i} s_i = \Sigma_i s_i \implies \Sigma_{i,m \geq x_i} s_i \geq \frac{\Sigma_i s_i}{2}$ ตามที่ต้องการ เราสามารถพิสูจน์ด้วยวิธีเดียวกันว่าจะต้องได้ว่า $\Sigma_{i,m \leq x_i} s_i \geq \frac{\Sigma_i s_i}{2}$ เช่นกัน

โดยสรุปคือเราต้องการเลือกพิกัด $x=m$ ที่ทำให้ และ $\Sigma_{i,m \geq x_i} s_i \geq \frac{\Sigma_i s_i}{2}$ และ $\Sigma_{i,m \leq x_i} s_i \geq \frac{\Sigma_i s_i}{2}$ กล่าวคือต้องการเลือกค่ามัธยฐานของพิกัด $x$ ของนักเรียนทุกคนนั่นเองซึ่งพิกัด $x$ ที่เป็นอันดับที่ $\lfloor \frac{\Sigma_i s_i}{2} \rfloor$ จะเข้าทั้งสองเงื่อนไขไม่ว่า $\Sigma_i s_i$ จะเป็นจำนวนคู่หรือจำนวนคี่

การเลือกค่ามัธยฐานโดยตรงด้วยการ sort ทั้ง $x$ และ $y$ และนับผลรวม $s_i$ จะช้าเกินไปสำหรับ $N = 500000$ เนื่องจากข้อนี้ให้เวลาเพียง 0.5 วินาที และการ sort ใช้เวลา $\mathcal{O}(N \log N)$

### Quickselect

ขั้นตอนวิธีมาตรฐานในการเลือกค่าอันดับที่ $k$ คือ Quickselect ซึ่งมี Expected Time Complexity $\mathcal{O}(N)$

ผลรวม $s_i$ จะช้าเกินไปสำหรับ $N=500000$ เนื่องจากข้อนี้ให้เวลาเพียง 0.5 วินาที และการ sort ใช้เวลา $\mathcal{O}(N \log N)$

### Quickselect

ขั้นตอนวิธีมาตรฐานในการเลือกค่าอันดับที่ $k$ คือ Quickselect ซึ่งมี Expected Time Complexity $\mathcal{O}(N)$

Quickselect จะรับค่าเป็น Array $A[b..e]$ และค่า $k$ ซึ่งแทนอันดับที่ต้องการ และ return ค่าที่เป็นอันดับ $k$ ใน $A[b..e]$

Quickselect มีลักษณะคล้าย Quicksort คือในแต่ละขั้น Quickselect จะสุ่มเลือก pivot มาเพื่อแยกข้อมูลที่ยังพิจารณาเป็น 2 กลุ่ม คือกลุ่มที่มาก่อนและมาหลัง pivot จากนั้นจะเรียก Quickselect อีกรอบแบบ recursive ในกลุ่มที่มีค่าอันดับที่ต้องการ สำหรับข้อนี้จะต้องดัดแปลงให้สามารถพิจารณาจำนวนนักเรียน $s_i$ ในแต่ละหอด้วย

ขั้นตอนวิธี Quickselect สามารถเขียนเป็น
 1. หากข้อมูลที่พิจารณามีเพียง 1 หอ ให้ return ค่า $x$ ของหอนี้
 2. สุ่มเลือกหอหนึ่งมาเป็น pivot และแยกข้อมูลเป็น 2 ส่วน คือหอที่มีค่า $x < x_{pivot}$ และหอที่มีค่า $x > x_{pivot}$ (สำหรับหอที่มีค่า $x$ เท่ากันให้สุ่มว่าจะไว้กลุ่มแรกหรือหลังแบบสุ่มครึ่งครึ่ง) ให้ $p$ เป็นจำนวนหอในกลุ่มแรกบวก 1 สามารถเอา pivot ไปไว้ที่ตำแหน่ง $A[p]$ และสลับกลุ่มแรกมาไว้ใน $A[b..(p-1)]$ และกลุ่มหลังไว้ใน $A[(p+1)..e]$ จะทำให้ทุกหอใน $A[b..(p-1)]$ มีพิกัด $x$ ไม่เกิน $x_{pivot}$ และใน $A[(p+1)..e]$ ไม่ต่ำกว่า $x_{pivot}$
 3. ให้ผลรวม $s_i$ ในกลุ่มหน้าเป็น $S_{front}$ และในกลุ่มหลังเป็น $S_{back}$ ถ้า $S_{front} \geq k$ ให้ Quickselect ค่าที่ $k$ ใน $A[b..(p-1)]$ และ return ค่าที่ได้ ถ้า $S_{back} > k$ ให้ Quickselect ค่าที่ $k - s_{pivot} - S_{back}$ ใน $A[(p+1)..e]$ หากไม่เข้าทั้งสองกรณีให้ return ค่า $x_{pivot}$ เพราะจะเป็นค่าที่เป็นอันดับ $k$ แล้ว

สังเกตว่าขั้นตอนวิธีนี้มี Worst Case Time Complexity $\mathcal{O}(N^2)$ เช่นเดียวกับ Quicksort เช่นหากเลือกหอที่มี $x$ ต่ำสุดในทุกรอบ อย่างไรก็ตามเราสามารถพิสูจน์ว่า Expected Time Complexity ของขั้นตอนวิธีนี้เป็นเพียง $\mathcal{O}(N)$

ให้เวลาที่ขั้นตอนวิธีนี้ใช้สำหรับ $N$ ค่าเป็น $T(N)$ จะได้ recurrence ว่า $E[T(N)] = \frac{1}{N}\Sigma_{i=1}^N E[T(i)] + \mathcal{O}(N) $ โดย $E[T(1)] = T(1) = 1$ ซึ่งจะพิสูจน์โดย induction ได้ว่า $E[T(N)] = \mathcal{O}(N)$ (เพราะจะได้

Quickselect จะรับค่าเป็น Array $A[b..e]$ และค่า $k$ ซึ่งแทนอันดับที่ต้องการ และ return ค่าที่เป็นอันดับ $k$ ใน $A[b..e]$

Quickselect มีลักษณะคล้าย Quicksort คือในแต่ละขั้น Quickselect จะสุ่มเลือก pivot มาเพื่อแยกข้อมูลที่ยังพิจารณาเป็น 2 กลุ่ม คือกลุ่มที่มาก่อนและมาหลัง pivot จากนั้นจะเรียก Quickselect อีกรอบแบบ recursive ในกลุ่มที่มีค่าอันดับที่ต้องการ สำหรับข้อนี้จะต้องดัดแปลงให้สามารถพิจารณาจำนวนนักเรียน $s_i$ ในแต่ละหอด้วย

ขั้นตอนวิธี Quickselect สามารถเขียนเป็น

1. หากข้อมูลที่พิจารณามีเพียง 1 หอ ให้ return ค่า $x$ ของหอนี้
2. สุ่มเลือกหอหนึ่งมาเป็น pivot และแยกข้อมูลเป็น 2 ส่วน คือหอที่มีค่า $x < x_{pivot}$ และหอที่มีค่า $x > x_{pivot}$ (สำหรับหอที่มีค่า $x$ เท่ากันให้สุ่มว่าจะไว้กลุ่มแรกหรือหลังแบบสุ่มครึ่งครึ่ง) ให้ $p$ เป็นจำนวนหอในกลุ่มแรกบวก 1 สามารถเอา pivot ไปไว้ที่ตำแหน่ง $A[p]$ และสลับกลุ่มแรก

มาไว้ใน $A[b..(p-1)]$ และกลุ่มหลังไว้ใน $A[(p+1)..e]$ จะทำให้ทุกหอใน $A[b..(p-1)]$ มีพิกัด $x$ ไม่เกิน $x_{pivot}$ และใน $A[(p+1)..e]$ ไม่ต่ำกว่า $x_{pivot}$

3. ให้ผลรวม $s_i$ ในกลุ่มหน้าเป็น $S_{front}$ และในกลุ่มหลังเป็น $S_{back}$ ถ้า $S_{front} \geq k$ ให้ Quickselect ค่าที่ $k$ ใน $A[b..(p-1)]$ และ return ค่าที่ได้ ถ้า $S_{back} > k$ ให้ Quickselect ค่าที่ $k - s_{pivot} - S_{back}$ ใน $A[(p+1)..e]$ หากไม่เข้าทั้งสองกรณีให้ return ค่า $x_{pivot}$ เพราะจะเป็นค่าที่เป็นอันดับ $k$ แล้ว

สังเกตว่าขั้นตอนวิธีนี้มี Worst Case Time Complexity $\mathcal{O}(N^2)$ เช่นเดียวกับ Quicksort เช่นหากเลือกหอที่มี $x$ ต่ำสุดในทุกรอบ อย่างไรก็ตามเราสามารถพิสูจน์ว่า Expected Time Complexity ของขั้นตอนวิธีนี้เป็นเพียง $\mathcal{O}(N)$

ให้เวลาที่ขั้นตอนวิธีนี้ใช้สำหรับ $N$ ค่าเป็น $T(N)$ จะได้ recurrence ว่า $E[T(N)] = \frac{1}{N}\Sigma_{i=1}^{N} E[T(i)] + \mathcal{O}(N)$ โดย $E[T(1)] = T(1) = 1$ ซึ่งจะพิสูจน์โดย induction ได้ว่า $E[T(N)] = \mathcal{O}(N)$ (เพราะจะได้ $E[T(N)] = \frac{1}{N}\Sigma_{i=1}^{N} E[T(i)] + \mathcal{O}(N) = \frac{1}{N}\Sigma_{i=1}^{N}\mathcal{O}(i) + \mathcal{O}(N) \leq \frac{1}{N}\Sigma_{i=1}^{N} Ci + \mathcal{O}(N) = \frac{C\frac{N(N-1)}{2}}{N} + \mathcal{O}(N) = C\frac{N-1}{2} + \mathcal{O}(N) = \mathcal{O}(N)$ )

## Solution

เมื่อรู้แล้วว่าค่า $x$ และ $y$ ที่จะทำให้ระยะทางต่ำสุดเป็นค่ามัธยฐานของพิกัดนักเรียนในแต่ละแกน เราสามารถใช้ Quickselect เลือกค่า $x$ และ $y$ ดังกล่าวและคำนวน $\Sigma_i s_i |x_i - x| + \Sigma_i s_i |y_i - y|$ โดยตรง โดยทั้งหมดใช้ Expected Time Complexity $\mathcal{O}(N)$

Home Tasks Learn About