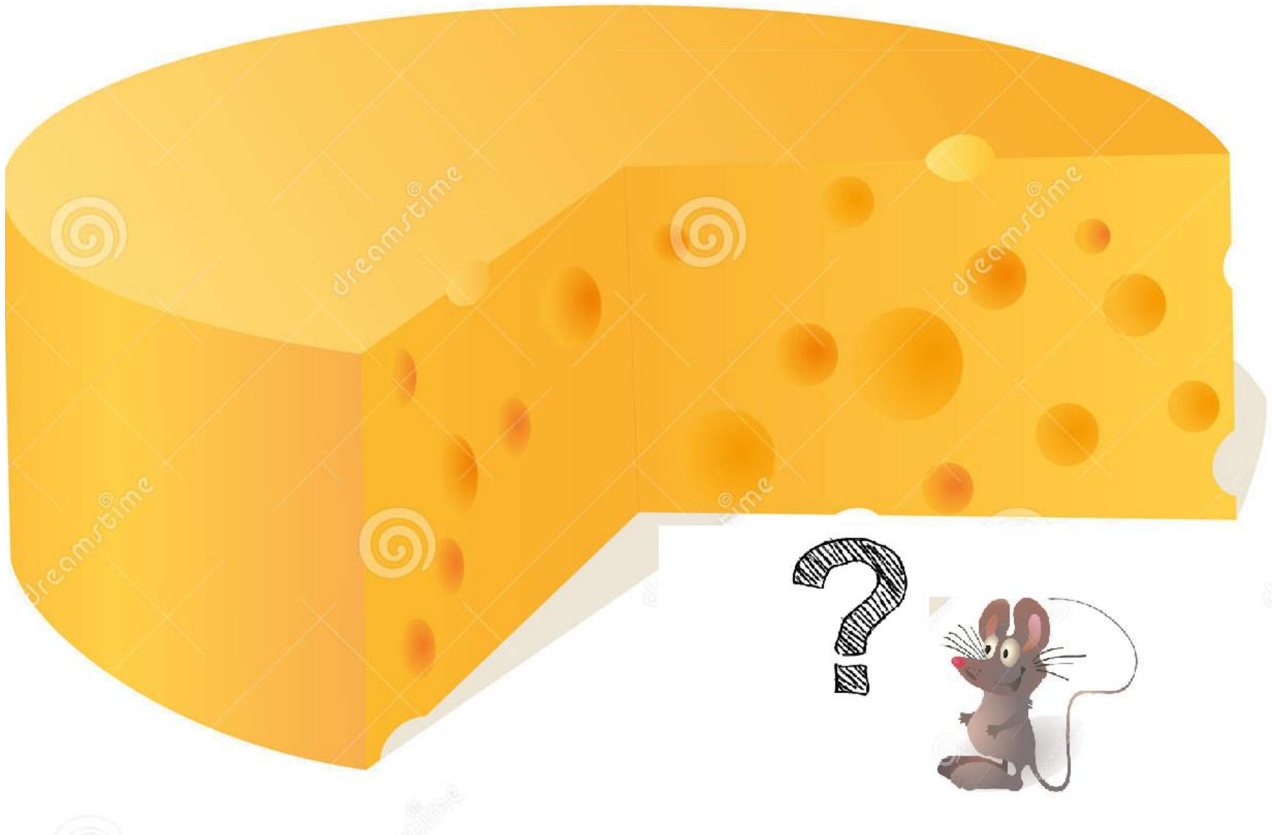
Big Data Class



LECTURER: DAN FELDMAN

TEACHING ASSISTANTS:

IBRAHIM JUBRAN

ALAA MAALOUF

Department of Computer Science, University of Haifa.

Bounding Sensitivity using Bicreteria (Intuition)

Reminder:

In the previous lecture, we saw how to bound the sensitivity of a point and the total sensitivity (sum for all points), using Bicreteria approximation.

$$s(p) \leq \rho\alpha \frac{dist(p, p')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha + 1) \cdot dist(p', T)}{dist(A', T)}$$

Bounding Sensitivity using Bicreteria (Intuition)

Reminder:

In the previous lecture, we saw how to bound the sensitivity of a point and the total sensitivity (sum for all points), using Bicreteria approximation.

$$s(p) \leq \rho\alpha \frac{dist(p,p')}{dist(A,A')} + \max_{T\in Q} \frac{\rho^2(\alpha+1)\cdot dist(p',T)}{dist(A',T)}$$

What is the intuition
behind this terms ?

Bounding Sensitivity using Bicreteria (Intuition)

Reminder:

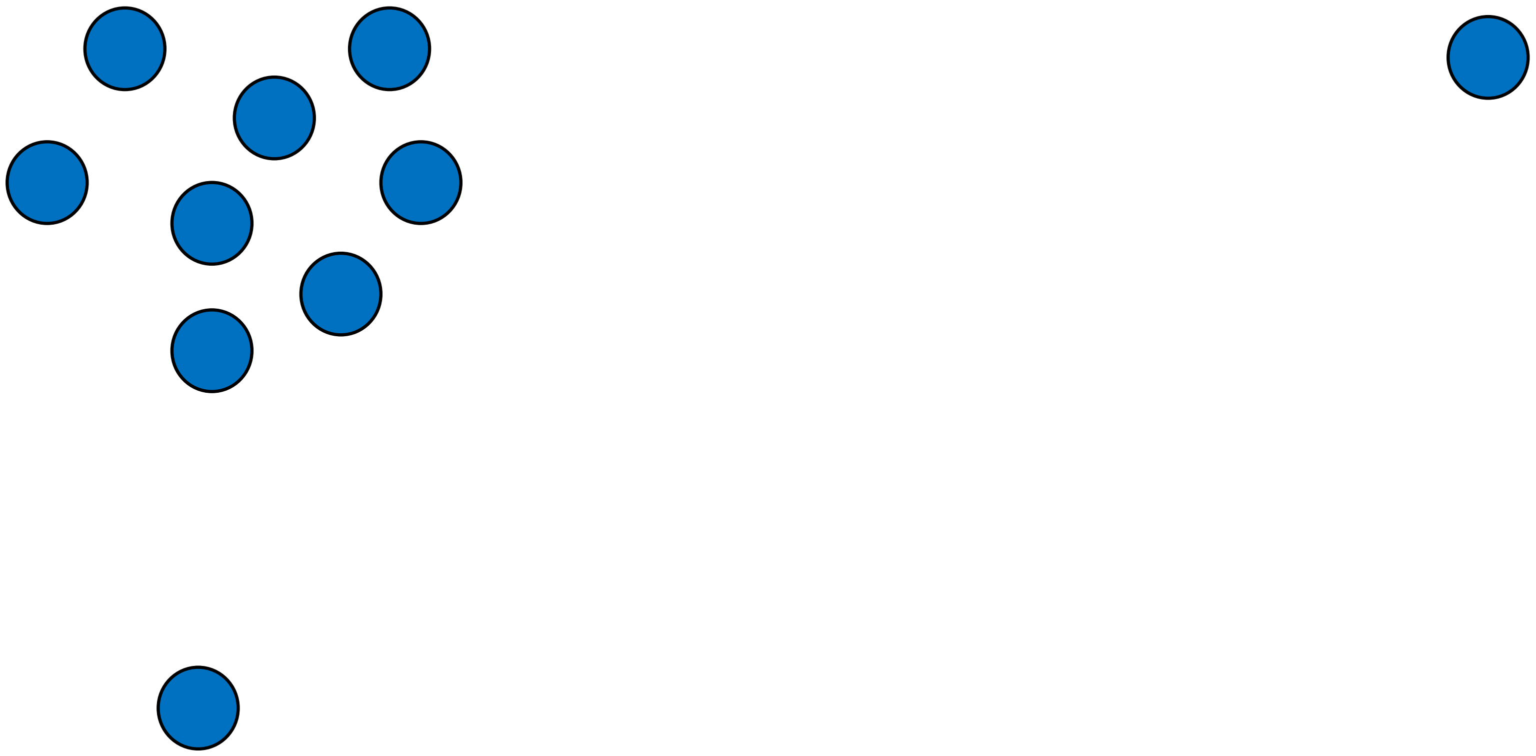
In the previous lecture, we saw how to bound the sensitivity of a point and the total sensitivity (sum for all points), using Bicreteria approximation.

$$s(p) \leq \rho\alpha \frac{dist(p, p')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha + 1) \cdot dist(p', T)}{dist(A', T)}$$

What is the intuition behind this terms ?

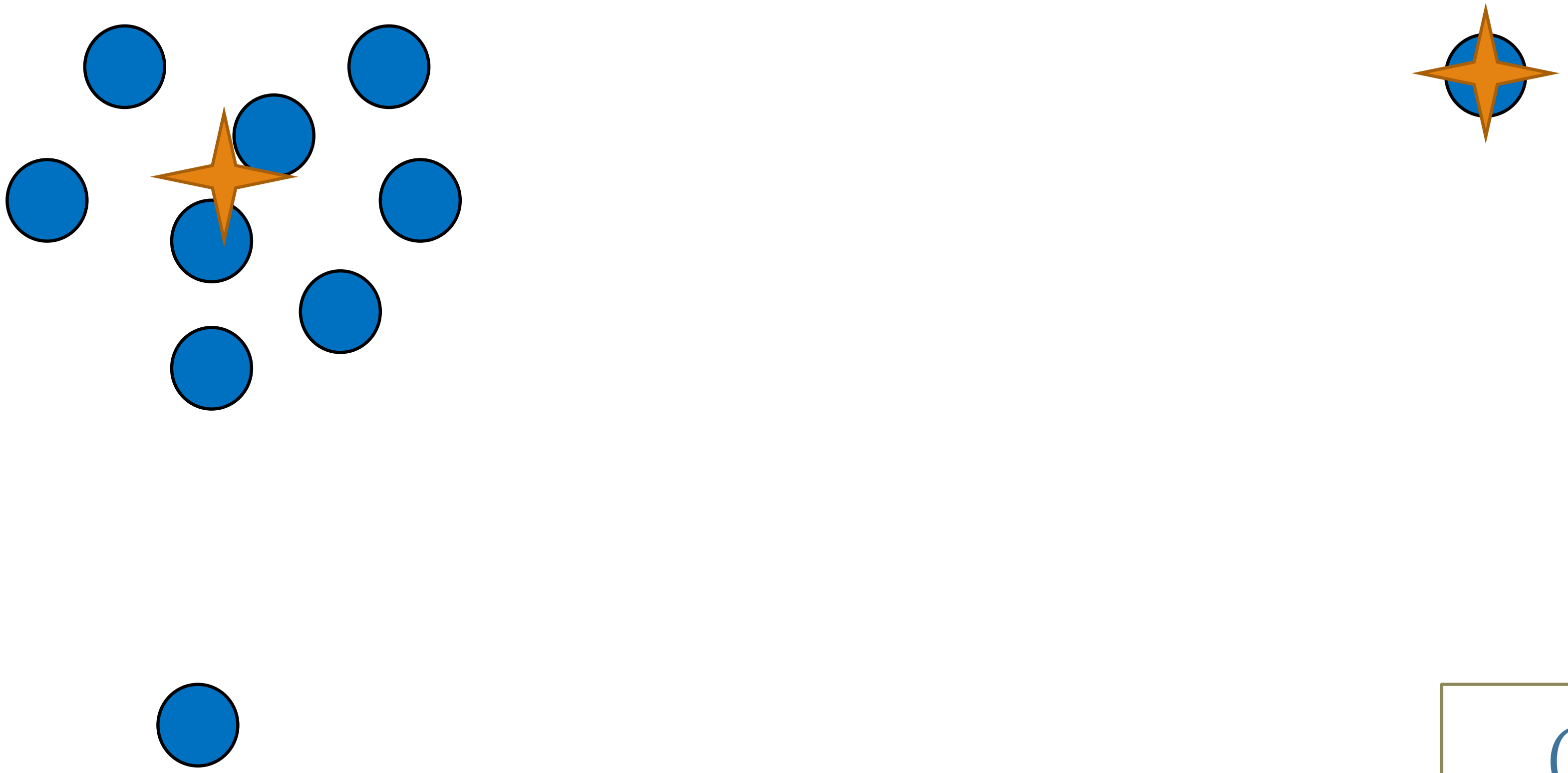
Bounding Sensitivity using Bicreteria (Intuition for k-means)

$k = 1$



Bounding Sensitivity using Bicreteria (Intuition for k-means)

$k = 1$

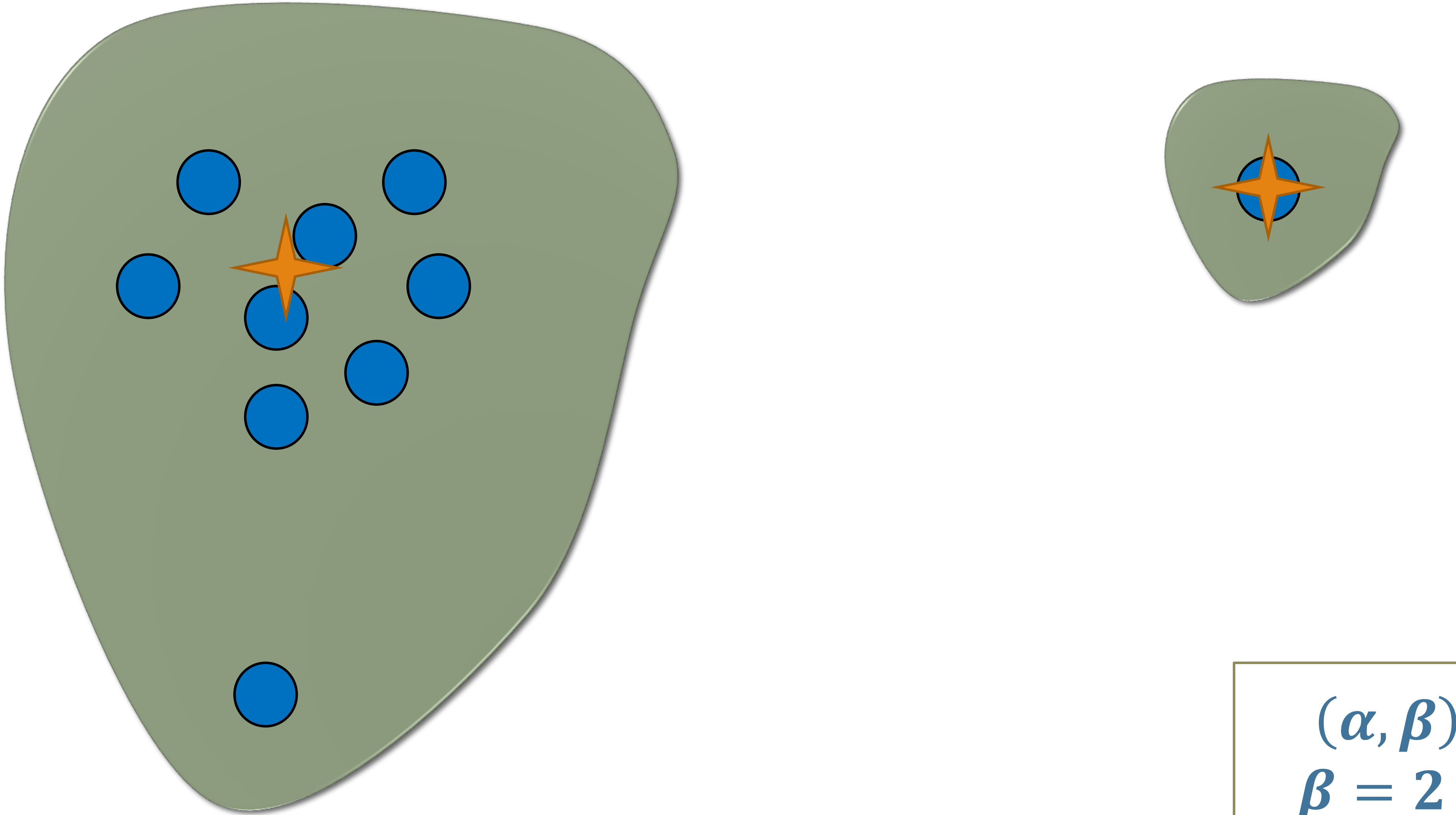


$(\alpha, \beta) - approx$

$\beta = 2$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

$k = 1$

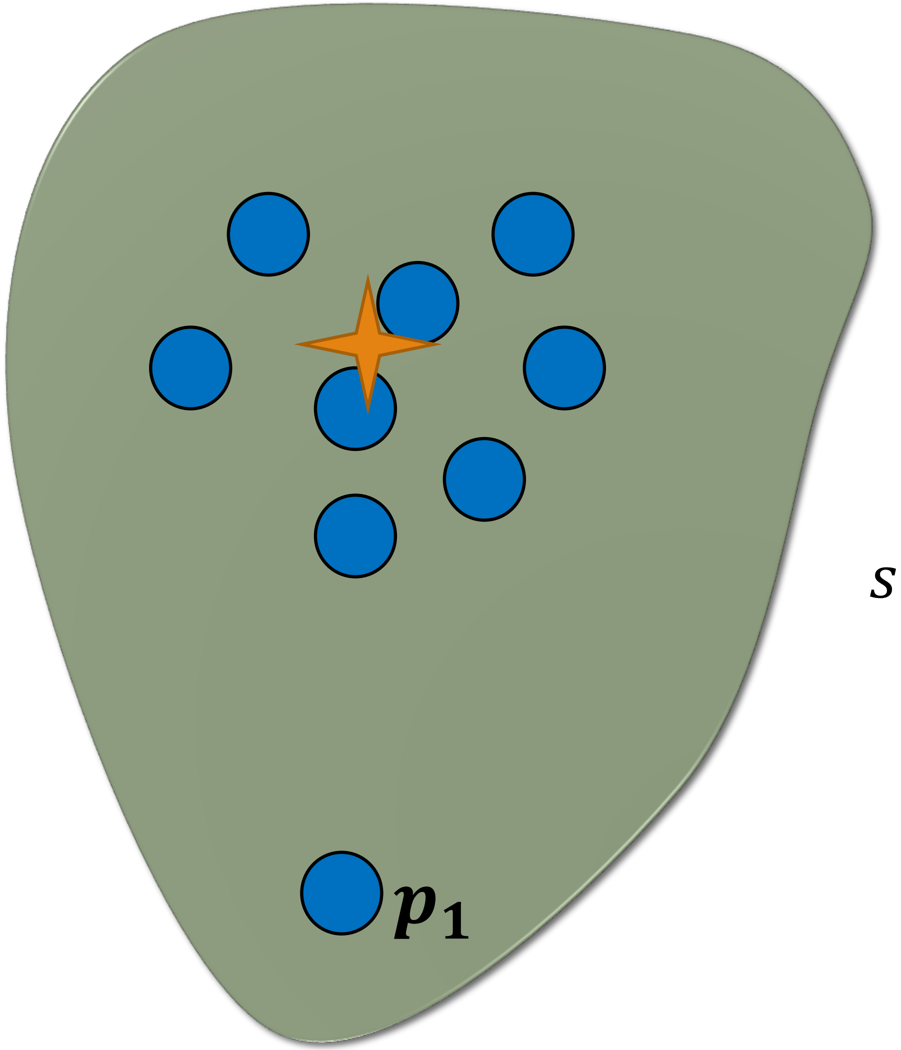


$(\alpha, \beta) - approx$

$\beta = 2$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

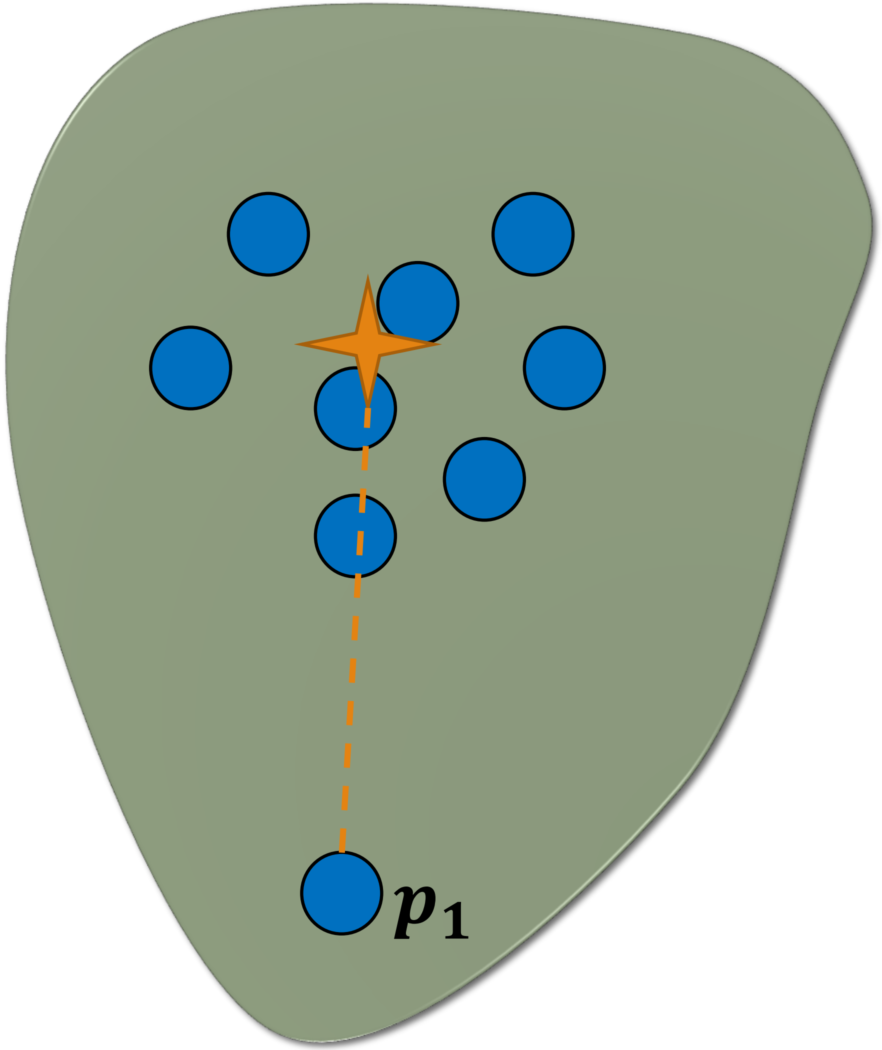
$\boldsymbol{k = 1}$



$$s(p_1) \leq \rho\alpha \frac{dist(p_1, p_1')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha + 1) \cdot dist(p_1', T)}{dist(A', T)}$$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

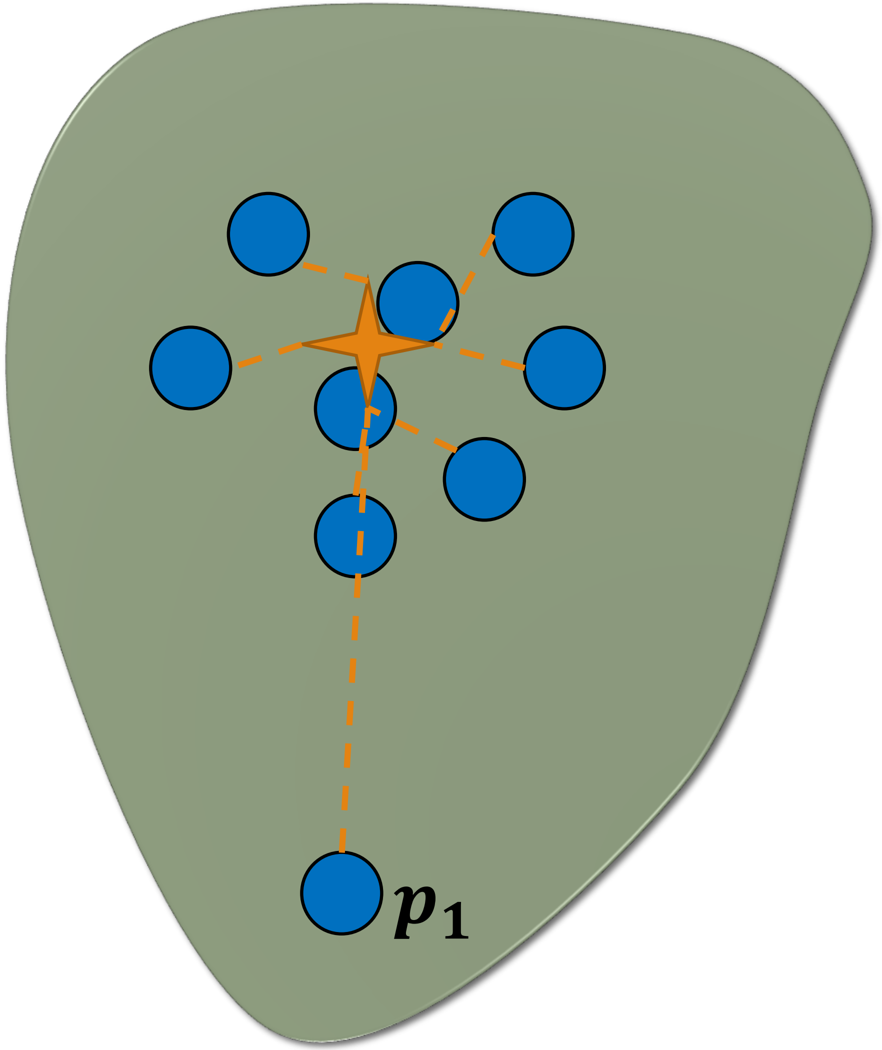
$k = 1$



$$s(p_1) \leq \rho\alpha \frac{dist(p_1, p_1')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha+1) \cdot dist(p_1', T)}{dist(A', T)}$$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

$k = 1$



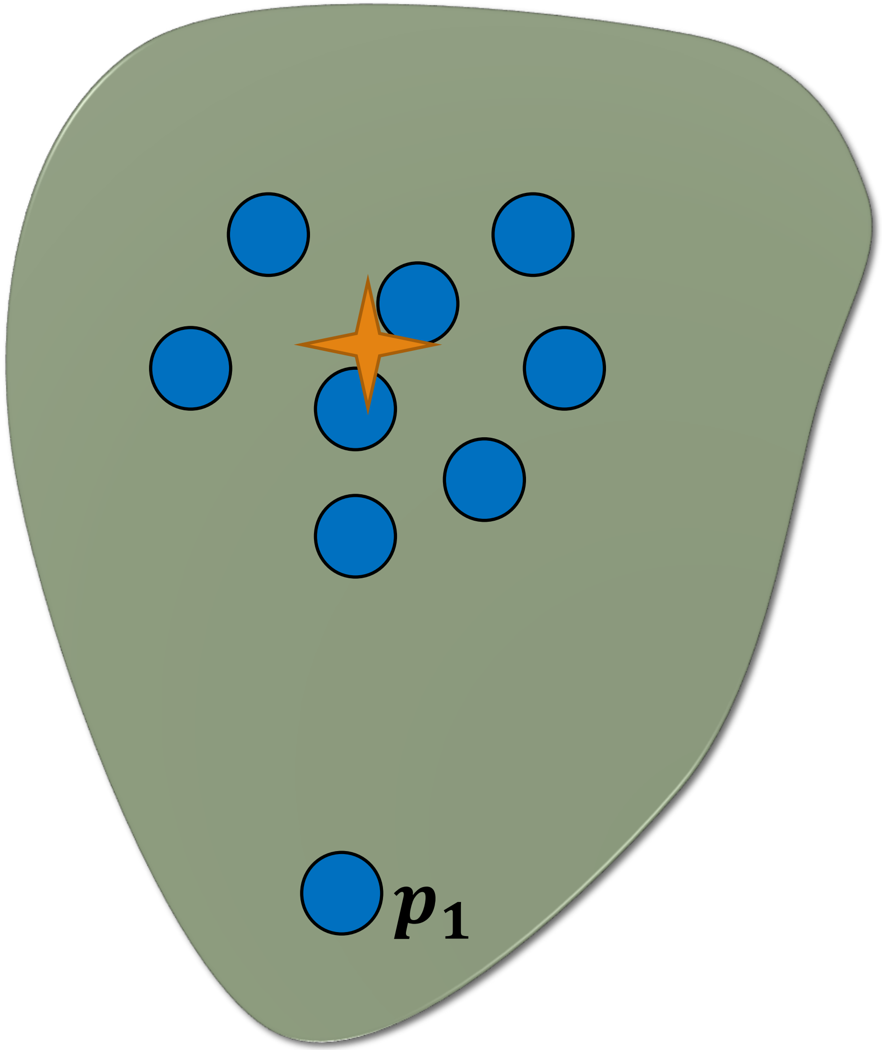
$\cong 1$

$$s(p_1) \leq \rho\alpha \frac{dist(p_1, p_1')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha+1) \cdot dist(p', T)}{dist(A', T)}$$

High sensitivity for p_1

Bounding Sensitivity using Bicreteria (Intuition for k-means)

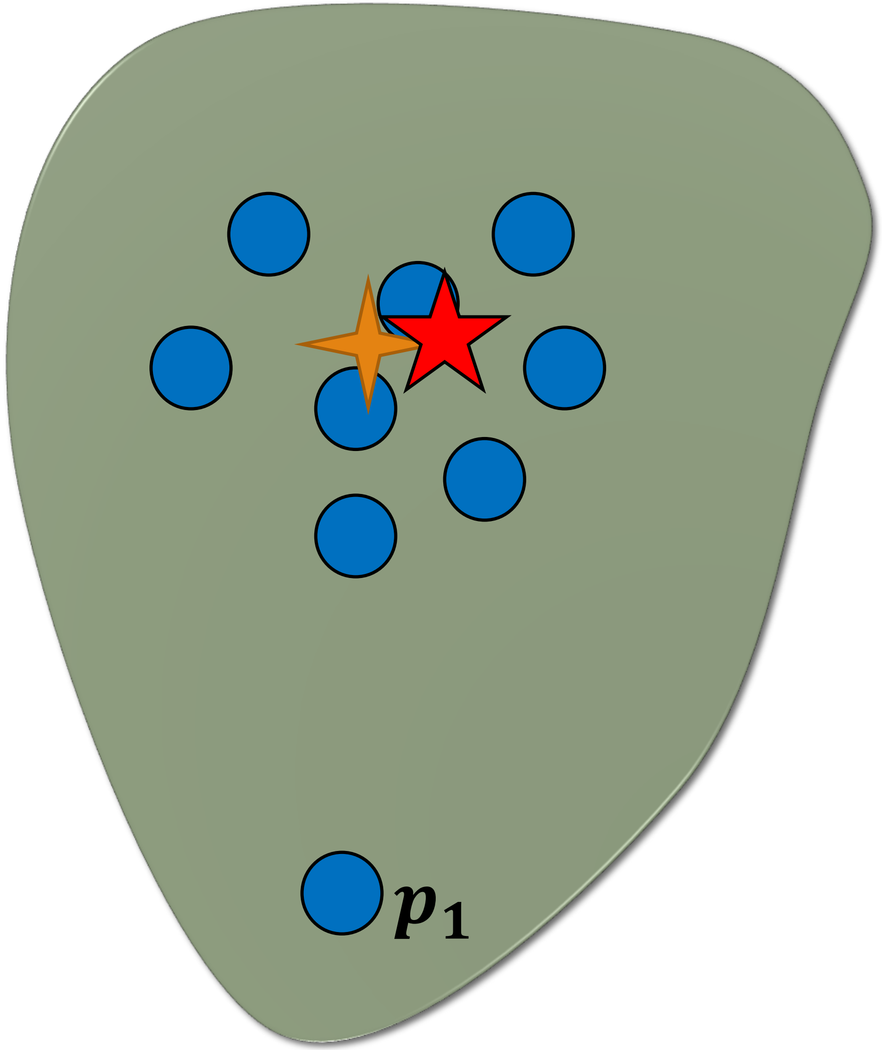
$\boldsymbol{k = 1}$



$$s(p_2) \leq \rho\alpha \frac{dist(p_2, p_2')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha + 1) \cdot dist(p_2', T)}{dist(A', T)}$$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

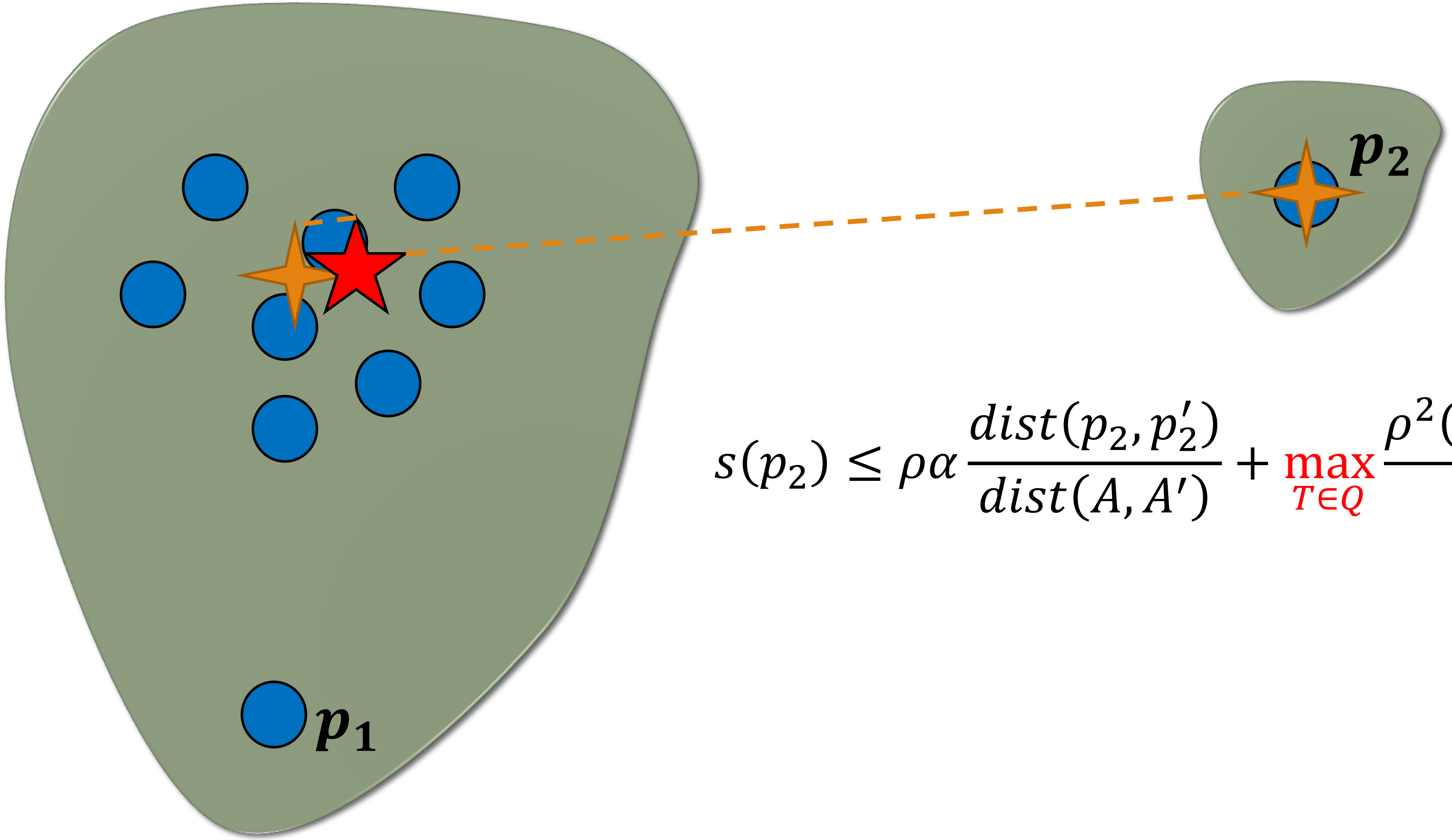
$k = 1$



$$s(p_2) \leq \rho\alpha \frac{dist(p_2, p_2')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha+1) \cdot dist(p_2', T)}{dist(A', T)}$$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

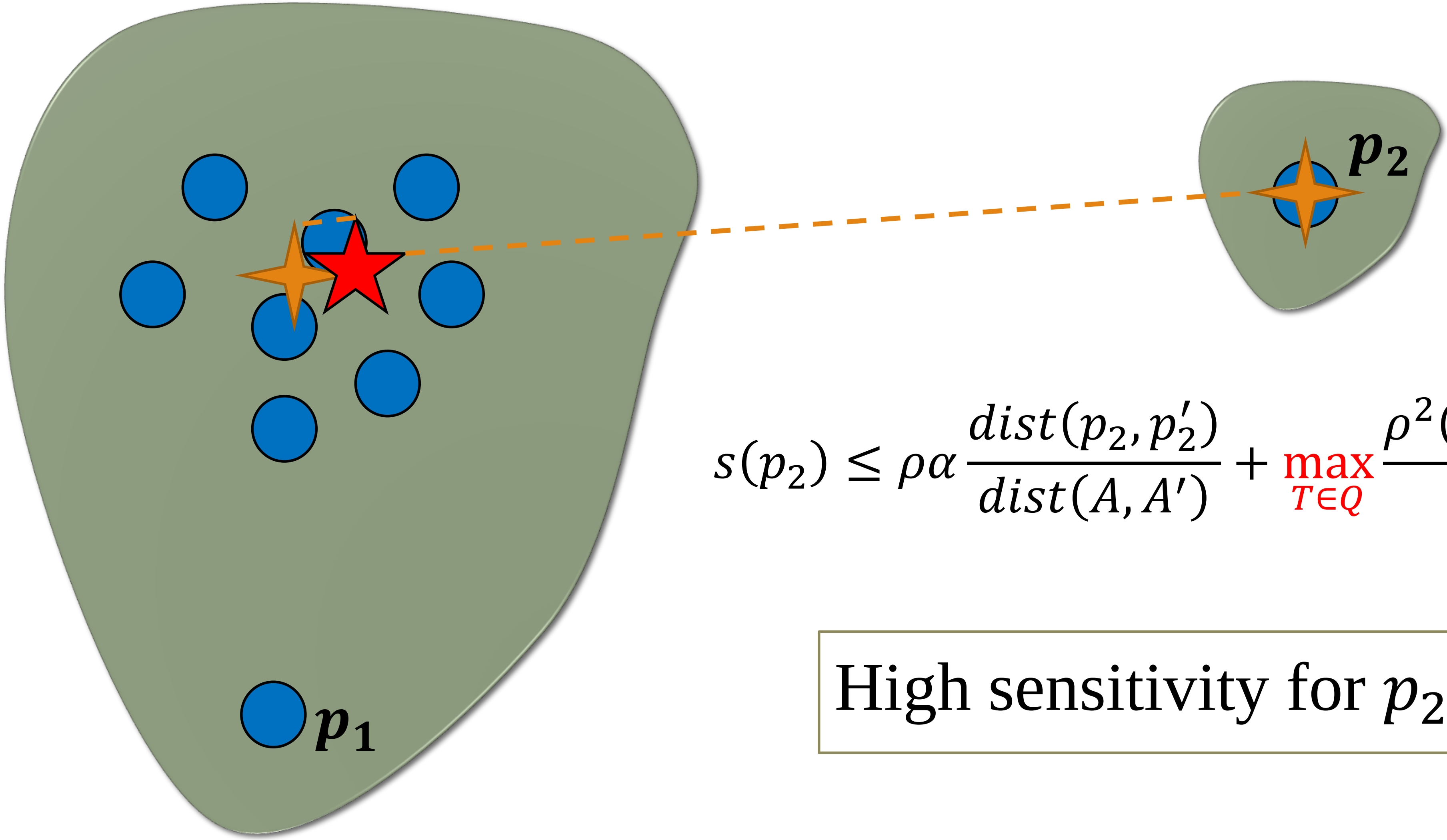
$k = 1$



$$s(p_2) \leq \rho\alpha \frac{dist(p_2, p_2')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha+1) \cdot dist(p_2', T)}{dist(A', T)}$$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

$k = 1$



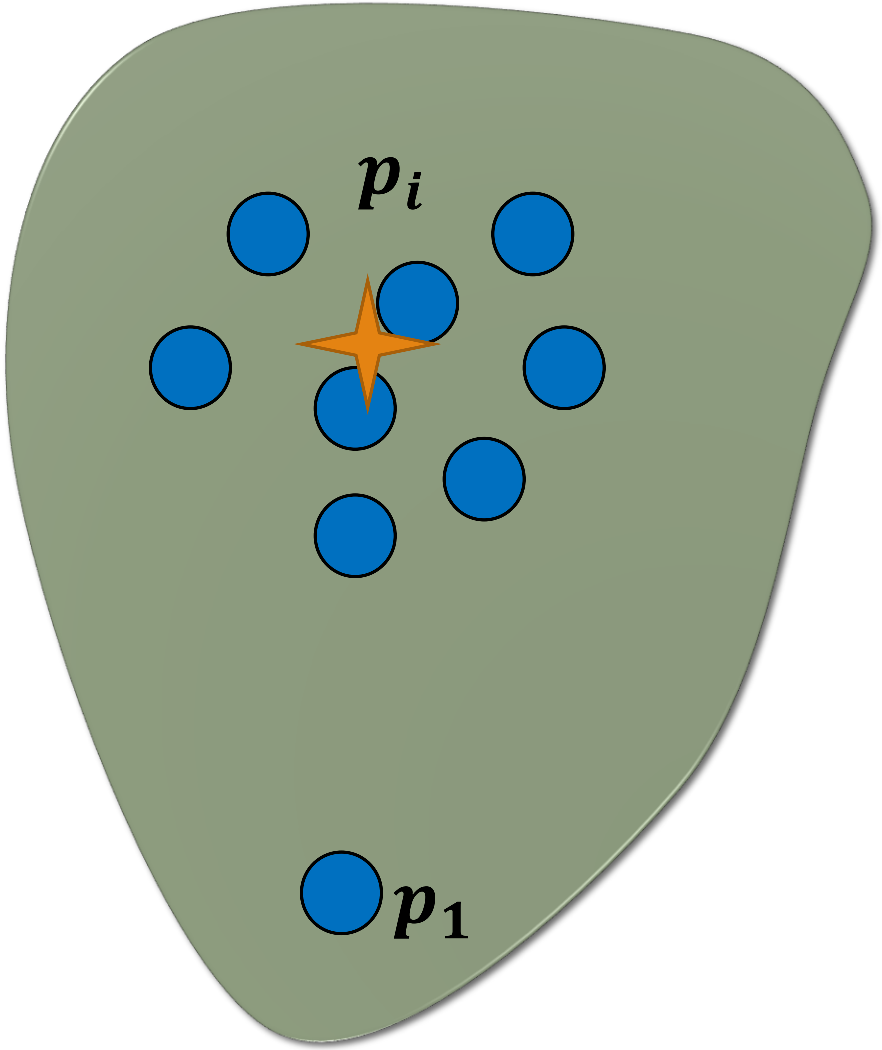
$\cong 1$

$$s(p_2) \leq \rho\alpha \frac{dist(p_2, p_2')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha+1) \cdot dist(p_2', T)}{dist(A', T)}$$

High sensitivity for p_2

Bounding Sensitivity using Bicreteria (Intuition for k-means)

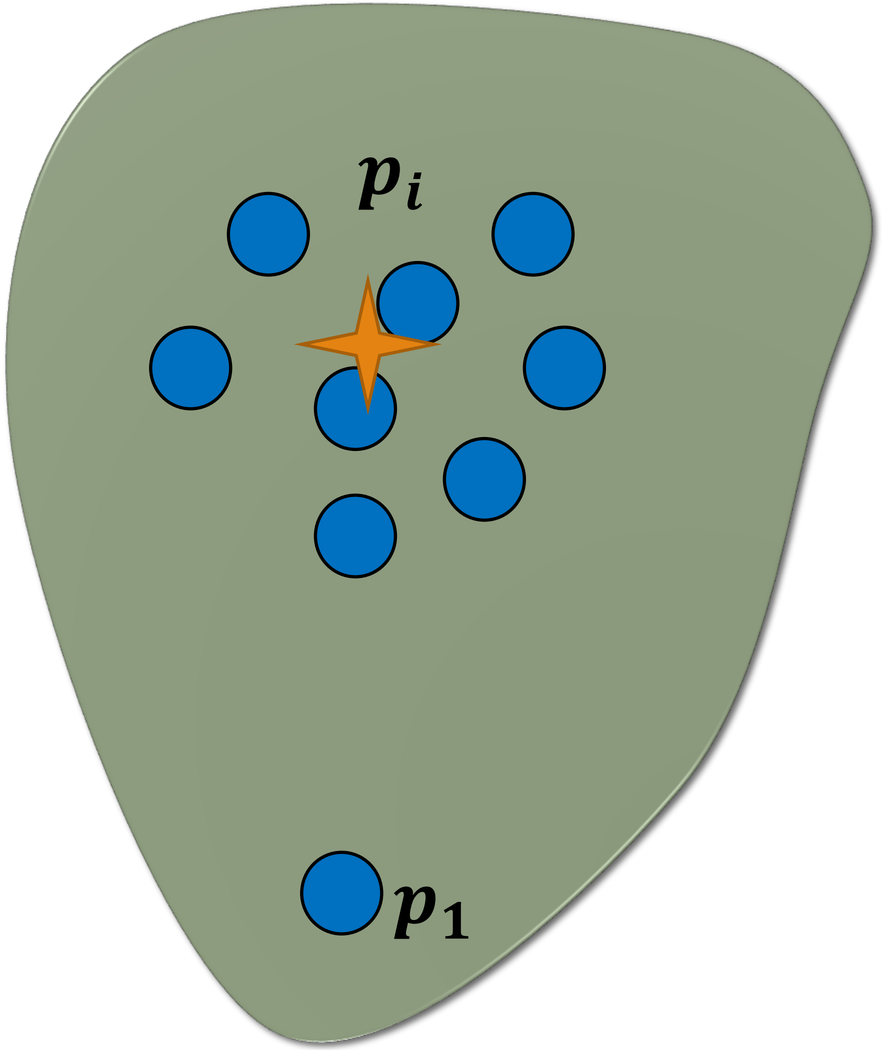
$\boldsymbol{k = 1}$



$$s(p_i) \leq \rho\alpha \frac{dist(p_i, p_i')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha + 1) \cdot dist(p_i', T)}{dist(A', T)}$$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

$\boldsymbol{k = 1}$

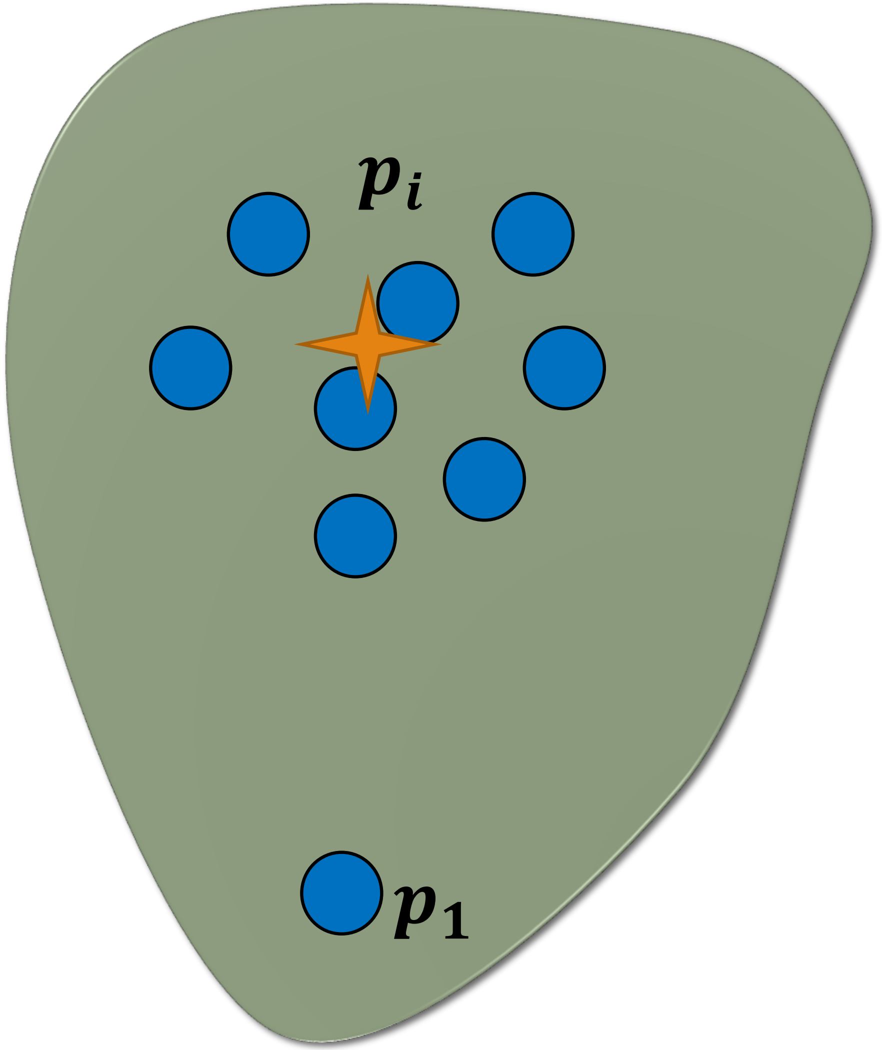


$\cong \mathbf{0}$

$$s(p_i) \leq \rho\alpha \frac{dist(p_i, p_i')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha + 1) \cdot dist(p_i', T)}{dist(A', T)}$$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

$k = 1$

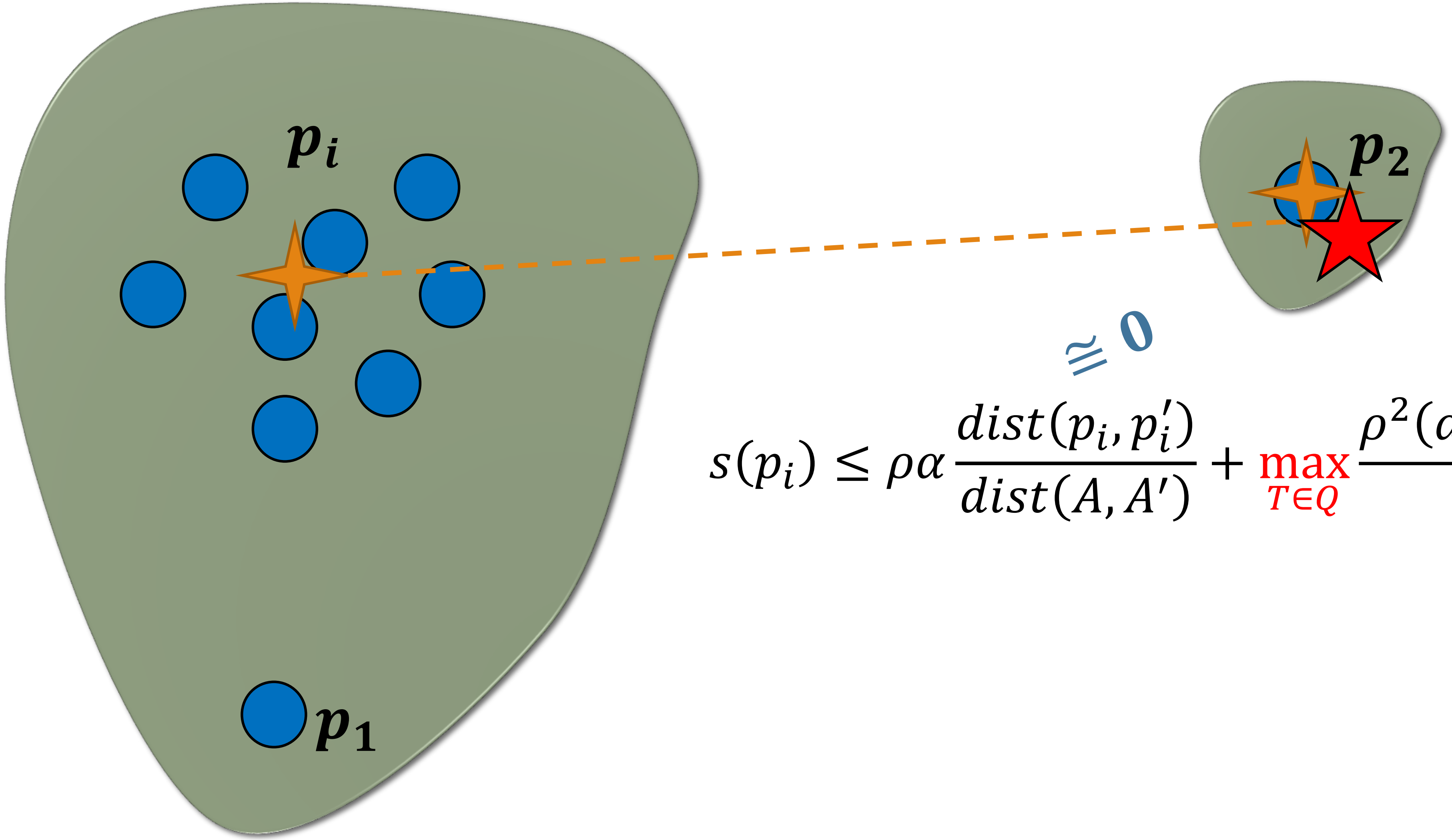


$\cong 0$

$$s(p_i) \leq \rho\alpha \frac{dist(p_i, p_i')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha+1) \cdot dist(p_i', T)}{dist(A', T)}$$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

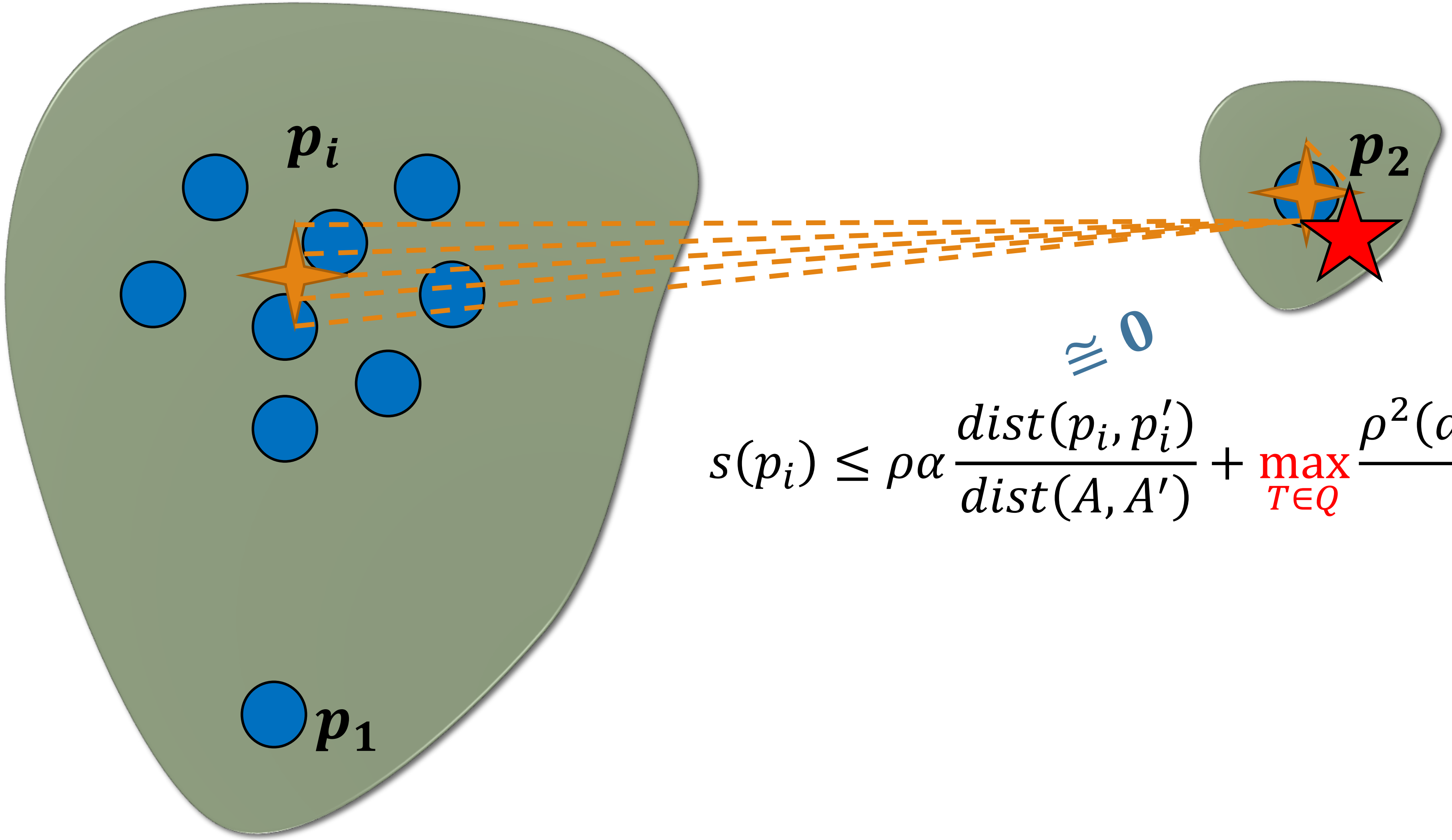
$\boldsymbol{k = 1}$



$$s(p_i) \leq \rho\alpha \frac{dist(p_i, p_i')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha+1) \cdot dist(p_i', T)}{dist(A', T)}$$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

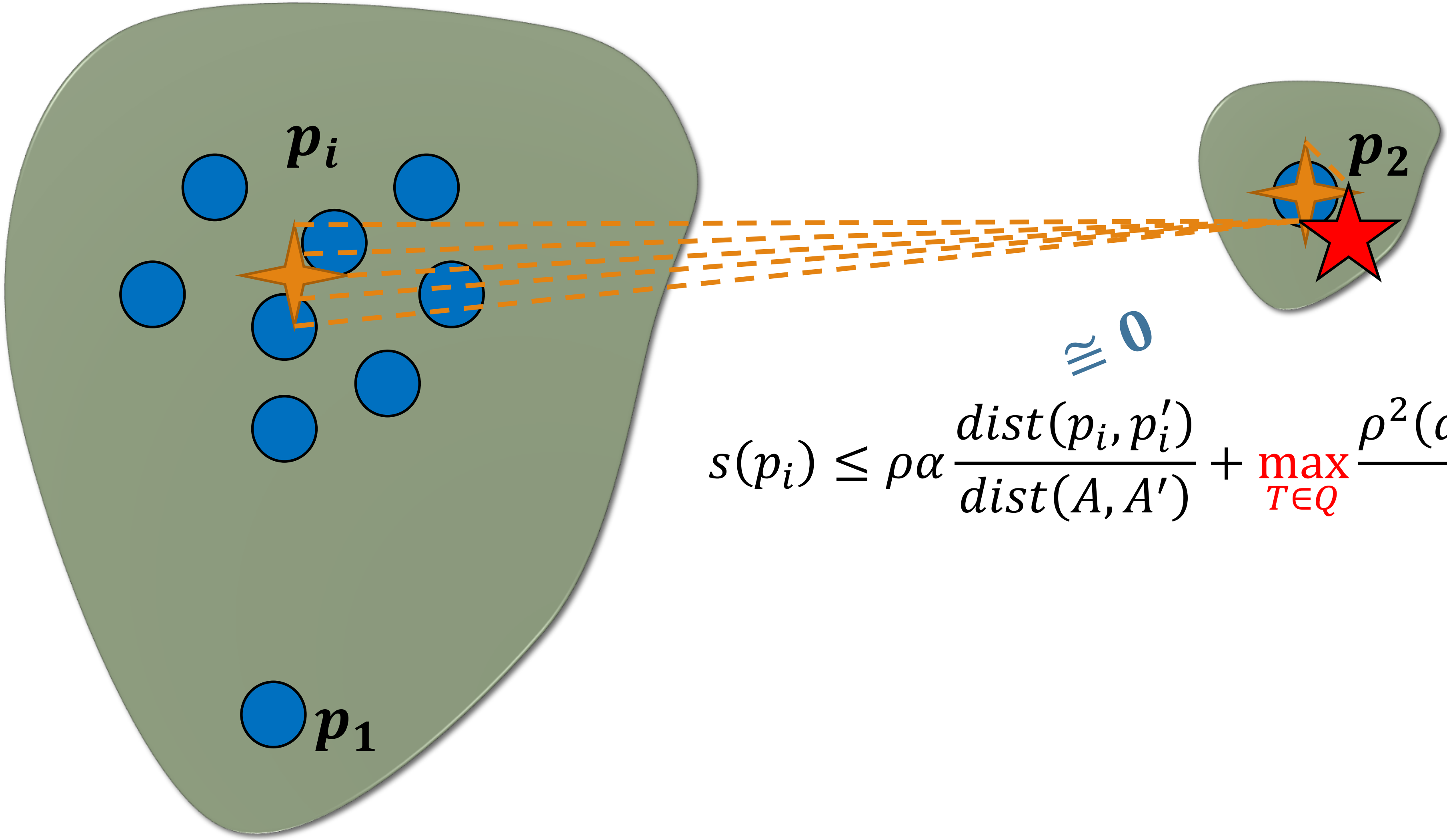
$k = 1$



$$s(p_i) \leq \rho\alpha \frac{dist(p_i, p_i')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha + 1) \cdot dist(p_i', T)}{dist(A', T)}$$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

$\boldsymbol{k = 1}$

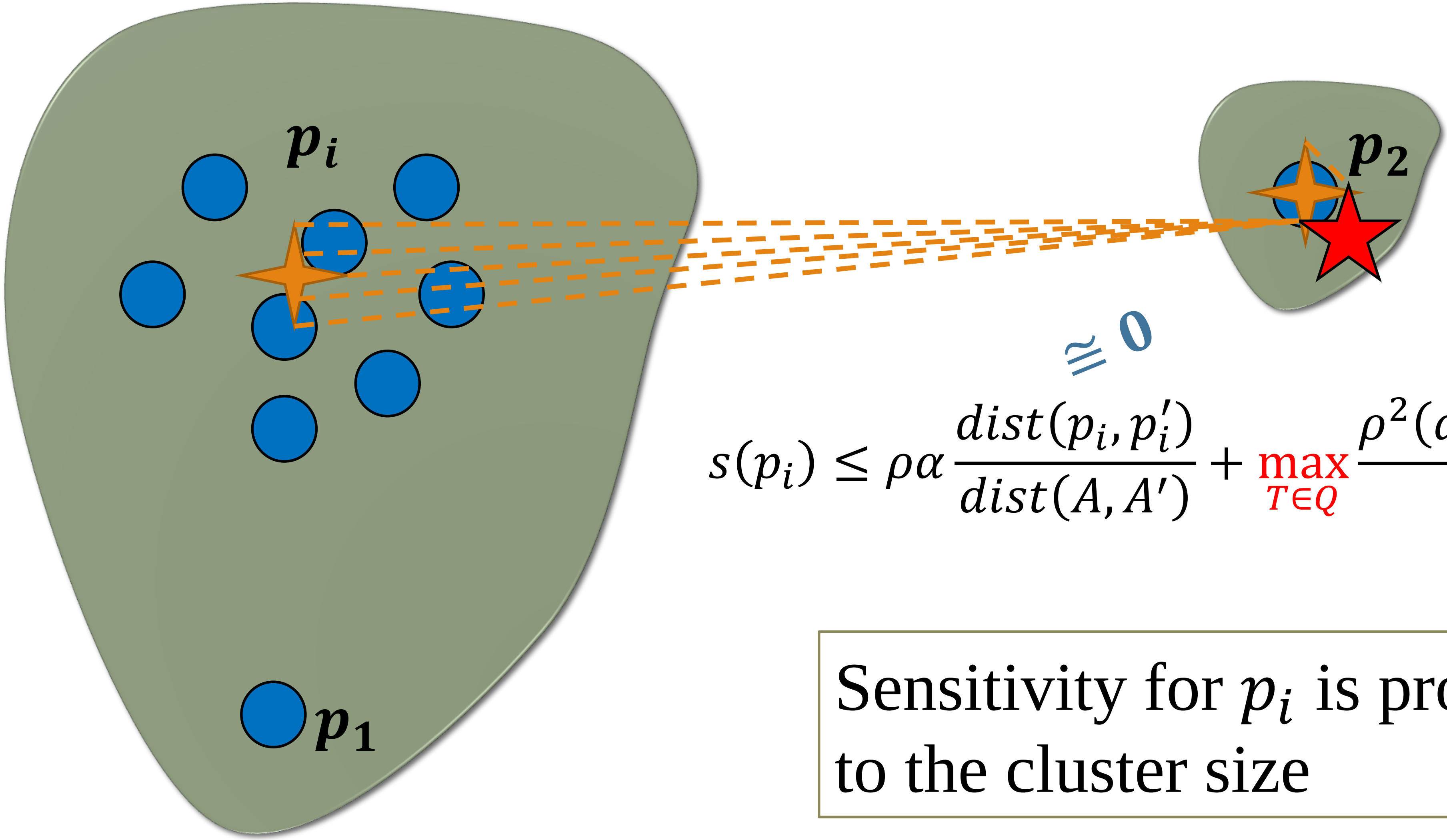


$\cong 0$ $\qquad$ $\cong \frac{1}{|A_i|}$

$$s(p_i) \le \rho\alpha \frac{dist(p_i, p_i')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha+1) \cdot dist(p_i', T)}{dist(A', T)}$$

Bounding Sensitivity using Bicreteria (Intuition for k-means)

$k = 1$



$\cong 0$ $\qquad$ $\cong \frac{1}{|A_i|}$

$$s(p_i) \le \rho\alpha \frac{dist(p_i, p_i')}{dist(A, A')} + \max_{T \in Q} \frac{\rho^2(\alpha + 1) \cdot dist(p_i', T)}{dist(A', T)}$$

Sensitivity for p_i is proportional to the cluster size

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \mathrm{proj}(\mathrm{p}, \mathrm{B})$.

Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$\left|\sum_{p \in P} dist(p, q) - \sum_{s \in S} \mu(s) \cdot (s, q)\right| \leq \epsilon \cdot \sum_{p \in P} dist(p, q)$$

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \mathrm{proj(p, B)}$.

Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$\left|\sum_{p\in P} dist(p,q) - \sum_{s\in S} \mu(s)\cdot(s,q)\right| \leq \epsilon \cdot \sum_{p\in P} dist(p,q)$$

$$\rightarrow \left|\sum_{p\in P} dist(p,q) - \sum_{p'\in P'} dist(p',q) + \sum_{p'\in P'} dist(p',q) - \sum_{s\in S}(s,q)\right| \leq \epsilon \cdot \sum_{p\in P} dist(p,q)$$

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \mathrm{proj(p, B)}$.

Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$\left| \sum_{p \in P} dist(p, q) - \sum_{s \in S} \mu(s) \cdot (s, q) \right| \leq \epsilon \cdot \sum_{p \in P} dist(p, q)$$

$$\rightarrow \left| \sum_{p \in P} dist(p, q) - \sum_{p' \in P'} dist(p', q) + \sum_{p' \in P'} dist(p', q) - \sum_{s \in S} \mu(s) \cdot (s, q) \right| \leq \epsilon \cdot \sum_{p \in P} dist(p, q)$$

$$\rightarrow \left| \frac{\sum_{p \in P} dist(p, q) - \sum_{p' \in P'} dist(p', q) + \sum_{p' \in P'} dist(p', q) - \sum_{s \in S} \mu(s) \cdot (s, q)}{\sum_{p \in P} dist(p, q)} \right| \leq \epsilon$$

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \mathrm{proj}(\mathrm{p}, \mathrm{B})$.

Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$\rightarrow \left| \frac{\sum_{p \in P} dist(p, q) - \sum_{p' \in P'} dist(p', q) + \sum_{p' \in P'} dist(p', q) - \sum_{s \in S} \mu(s) \cdot (s, q)}{\sum_{p \in P} dist(p, q)} \right| \le \epsilon$$

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \mathrm{proj}(\mathrm{p}, \mathrm{B})$.

Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$\rightarrow \left| \frac{\sum_{p \in P} dist(p, q) - \sum_{p' \in P'} dist(p', q) + \sum_{p' \in P'} dist(p', q) - \sum_{s \in S} \mu(s) \cdot (s, q)}{\sum_{p \in P} dist(p, q)} \right| \leq \epsilon$$

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \text{proj}(\text{p}, \text{B})$.

Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$\rightarrow \left| \frac{\sum_{p \in P} dist(p, q) - \sum_{p' \in P'} dist(p', q) + \sum_{p' \in P'} dist(p', q) - \sum_{s \in S} \mu(s) \cdot (s, q)}{\sum_{p \in P} dist(p, q)} \right| \leq \epsilon$$

Add $\text{S}_1 = B = \{b_1, \ldots, b_{\beta k}\}$ to the coreset with weight $\mu(b_i) = clusterSize$:

$$\rightarrow \sum_{p' \in P'} dist(p', q) = \sum_{s \in S_1} \mu(s) \cdot (s, q)$$

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \mathrm{proj}(\mathrm{p}, \mathrm{B})$.

Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$\rightarrow \left| \frac{\sum_{p \in P} dist(p, q) - \sum_{p' \in P'} dist(p', q) + \sum_{p' \in P'} dist(p', q) - \sum_{s \in S} \mu(s) \cdot (s, q)}{\sum_{p \in P} dist(p, q)} \right| \leq \epsilon$$

Add $\mathrm{S}_1 = B = \{b_1, \ldots, b_{\beta k}\}$ to the coreset with weight $\mu(b_i) = clusterSize$:

$$\rightarrow \sum_{p' \in P'} dist(p', q) = \sum_{s \in S_1} \mu(s) \cdot (s, q)$$

$$\rightarrow \frac{\sum_{p' \in P'} dist(p', q) - \sum_{s \in S_1} \mu(s) \cdot (s, q)}{\sum_{p \in P} dist(p, q)} = 0$$

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \mathrm{proj(p, B)}$.

Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$\rightarrow \left|\frac{\sum_{p \in P} dist(p,q) - \sum_{p' \in P'} dist(p',q) + \sum_{p' \in P'} dist(p',q) - \sum_{s \in S} \mu(s) \cdot (s,q)}{\sum_{p \in P} dist(p,q)}\right| \leq \epsilon$$

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \mathrm{proj(p, B)}$.

Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$\rightarrow \left| \frac{\sum_{p \in P} dist(p,q) - \sum_{p' \in P'} dist(p',q) + \sum_{p' \in P'} dist(p',q) - \sum_{s \in S} \mu(s) \cdot (s,q)}{\sum_{p \in P} dist(p,q)} \right| \leq \epsilon$$

Define $f(p,q) = \frac{dist(p,q) - dist(p',q)}{\sum_{p \in P} dist(p,q)}$

Goal: $\left| \sum_{p \in P} f(p,q) - \sum_{s \in S_2} \mu(s) \cdot (s,q) \right| \leq \epsilon$

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \mathrm{proj}(\mathrm{p}, \mathrm{B})$.

Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$\rightarrow \left| \frac{\sum_{p \in P} dist(p,q) - \sum_{p' \in P'} dist(p',q) + \sum_{p' \in P'} dist(p',q) - \sum_{s \in S} \mu(s) \cdot (s,q)}{\sum_{p \in P} dist(p,q)} \right| \leq \epsilon$$

Define $f(p,q) = \frac{dist(p,q) - dist(p',q)}{\sum_{p \in P} dist(p,q)}$

Goal: $\left| \sum_{p \in P} f(p,q) - \sum_{s \in S_2} \mu(s) \cdot (s,q) \right| \leq \epsilon$

$$s(p) = \max_{q \in Q} f(p,q)$$

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \mathrm{proj(p, B)}$.

Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$\rightarrow \left| \frac{\sum_{p \in P} dist(p,q) - \sum_{p' \in P'} dist(p',q) + \sum_{p' \in P'} dist(p',q) - \sum_{s \in S} \mu(s) \cdot (s,q)}{\sum_{p \in P} dist(p,q)} \right| \leq \epsilon$$

$$s(p) = \max_{q \in Q} f(p,q)$$

$$\sum_{p \in P} s(p) = \sum_{p \in P} \max_{q \in Q} \frac{dist(p,q) - dist(p',q)}{\sum_{p \in P} dist(p,q)} \leq \sum_{p \in P} \frac{dist(p,p')}{OPT} \leq \alpha$$

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \mathrm{proj}(\mathrm{p}, \mathrm{B})$.

Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$\rightarrow \left| \frac{\sum_{p \in P} dist(p, q) - \sum_{p' \in P'} dist(p', q) + \sum_{p' \in P'} dist(p', q) - \sum_{s \in S} \mu(s) \cdot (s, q)}{\sum_{p \in P} dist(p, q)} \right| \leq \epsilon$$

Define $f(p, q) = \frac{dist(p,q) - dist(p',q)}{\sum_{p \in P} dist(p,q)}$

Goal: $\left| \sum_{p \in P} f(p, q) - \sum_{s \in S_2} \mu(s) \cdot (s, q) \right| \leq \epsilon$

Bounded total sensitivity → we can compute a coreset (S_2, μ_2) that satisfies

Coreset for k-median

Input: (P, Q) and an (α, β)-approximation B. Let $p' = \mathrm{proj(p, B)}$.

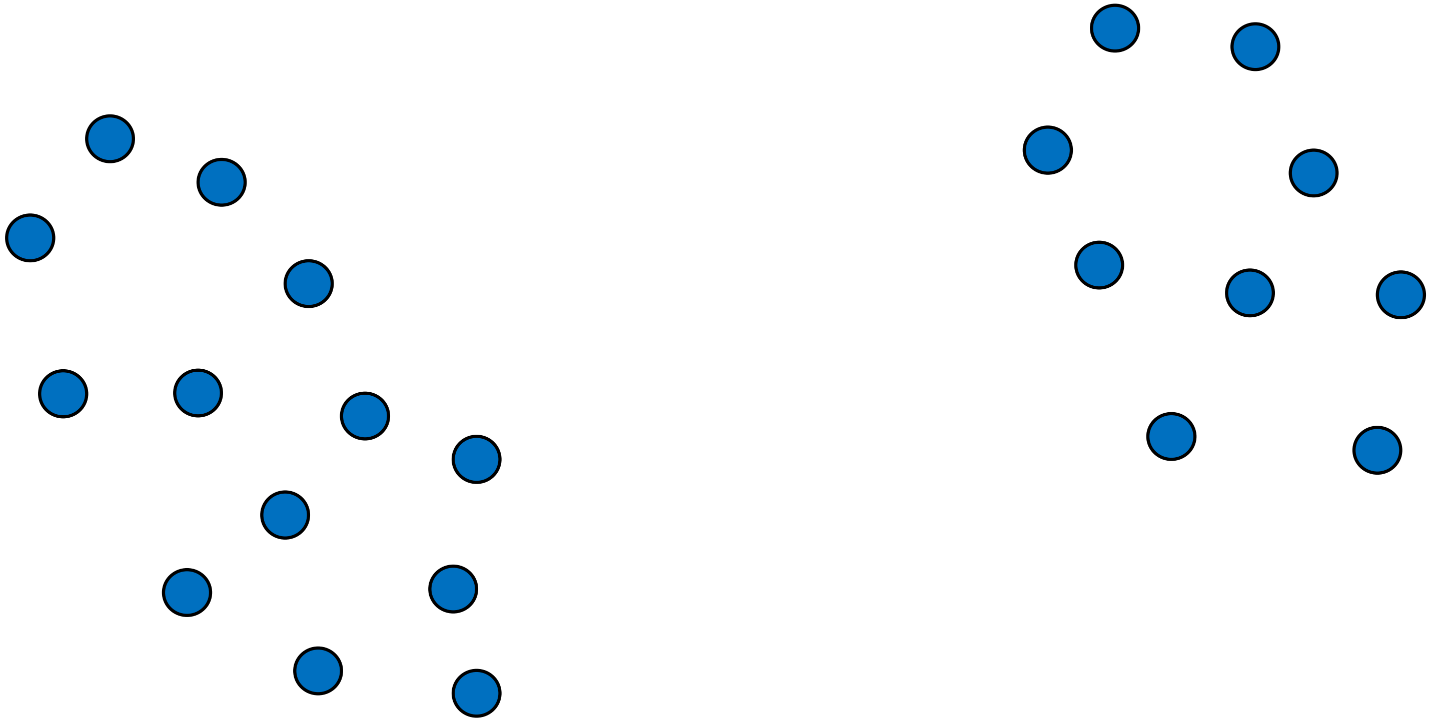
Goal: To compute a set (S, μ) such that for every $q \subseteq Q$:

$$S = S_1 \cup S_2$$

$$\rightarrow \left| \frac{\sum_{p \in P} dist(p, q) - \sum_{p' \in P'} dist(p', q) + \sum_{p' \in P'} dist(p', q) - \sum_{s \in S} \mu(s) \cdot (s, q)}{\sum_{p \in P} dist(p, q)} \right| \leq \epsilon$$

Sensitivity for k-means

Example: $k = 2$

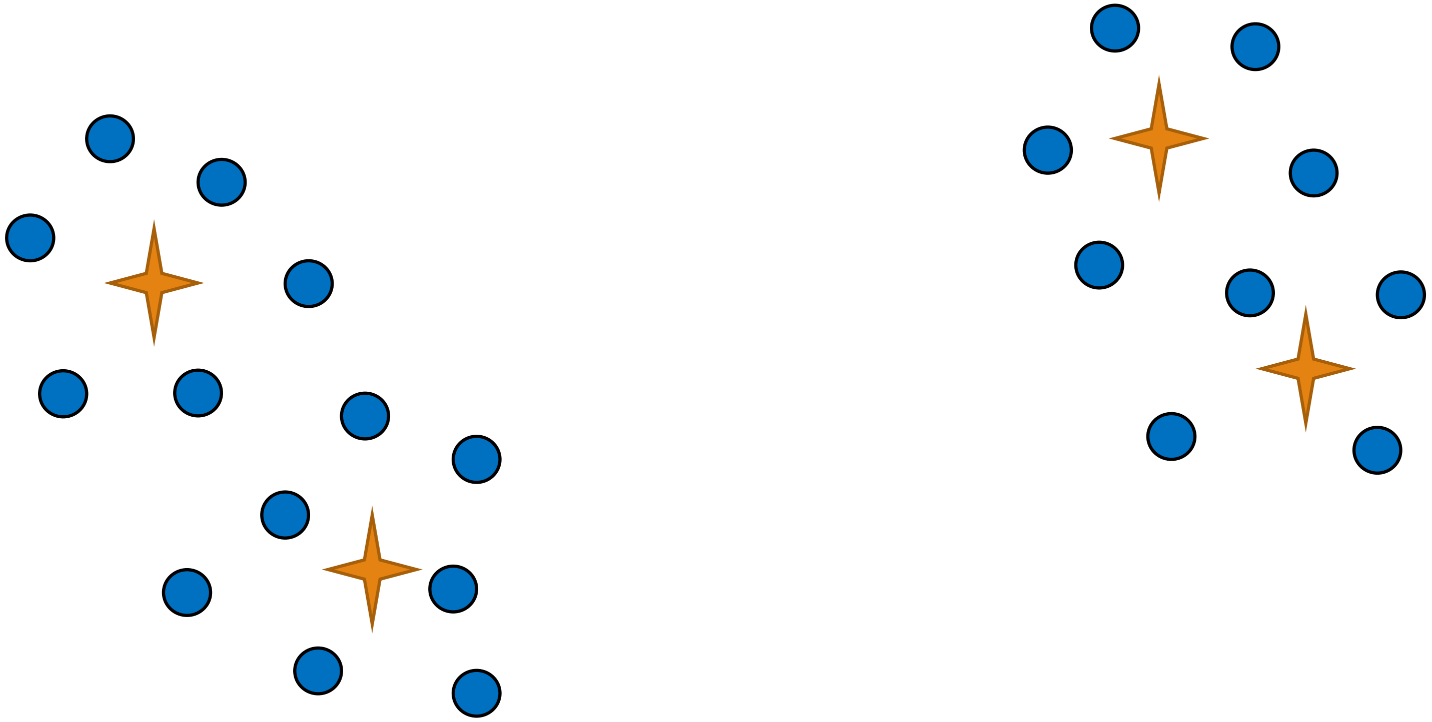


Sensitivity for k-means

<u>Example: $k = 2$</u>

Compute an (α, β)-approximation

2



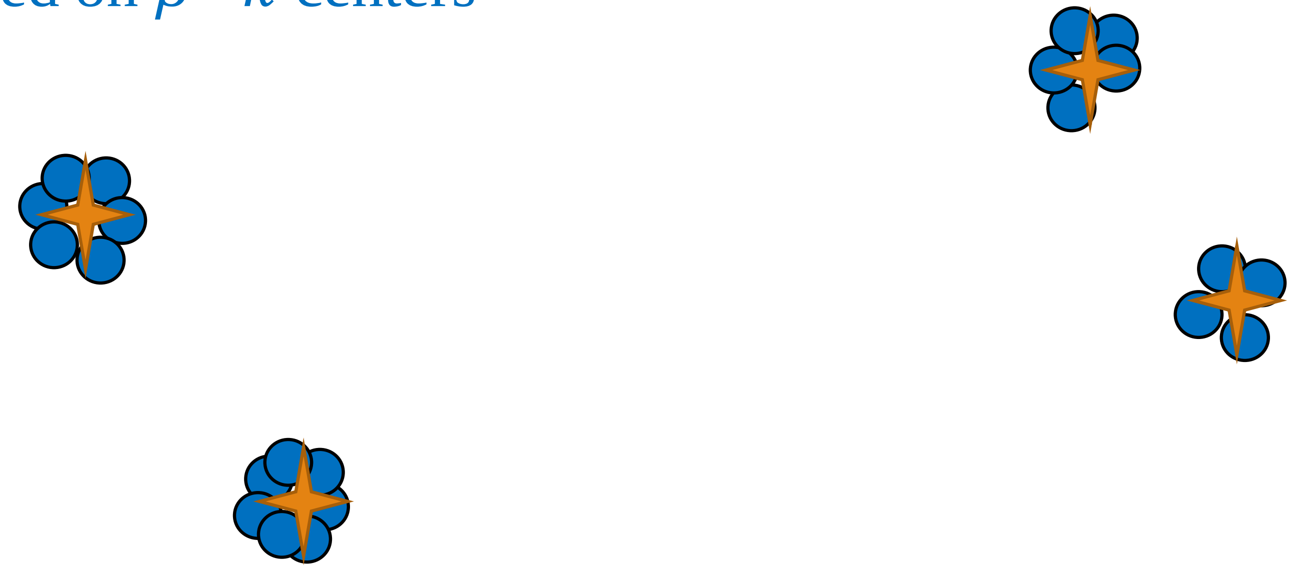
Sensitivity for k-means

Example: $k = 2$

Compute an (α, β)-approximation

2

→ n points projected on $\beta \cdot k$ centers



Sensitivity for k-mea

Using the last Lemma we have that:

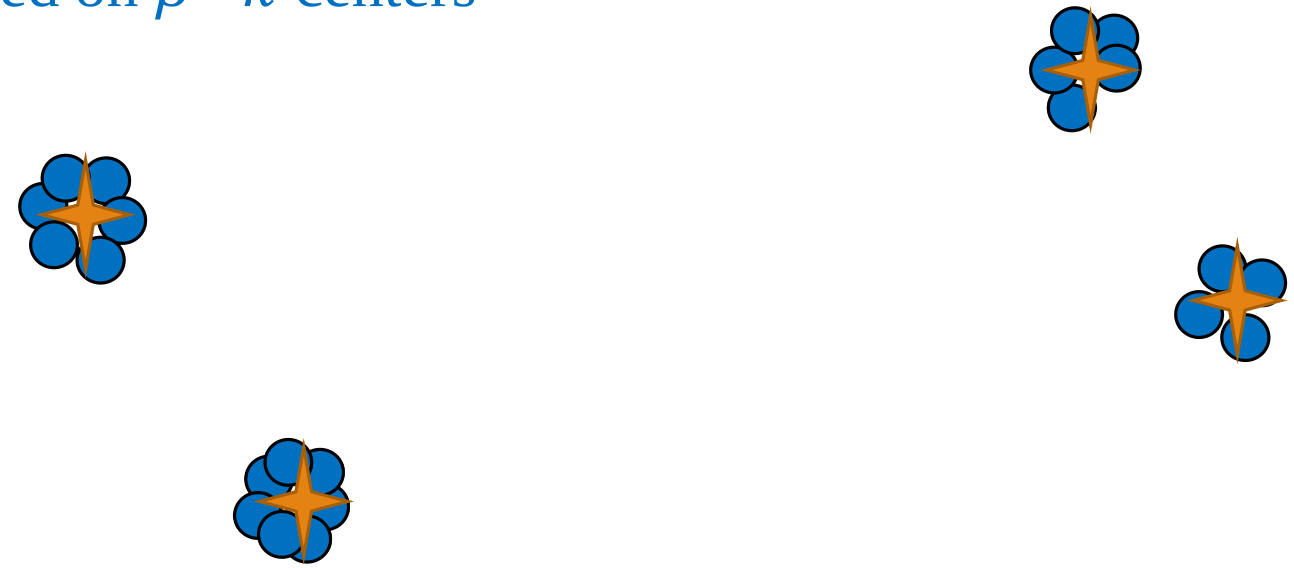
$$\sum_{p \in A} s(p) \leq \rho\alpha + \rho^2(1+\alpha) \sum_{p' \in A'} \max_{T \in Q} \frac{dist(p', T)}{dist(A', T)}$$

Example: $k = 2$

Compute an (α, β)-approximation

2

→ n points projected on $\beta \cdot k$ centers



Sensitivity for k-mea

Using the last Lemma we have that:

$$\sum_{p \in A} s(p) \leq \rho\alpha + \rho^2(1+\alpha) \underbrace{\sum_{p' \in A'} \max_{T \in Q} \frac{dist(p',T)}{dist(A',T)}}$$

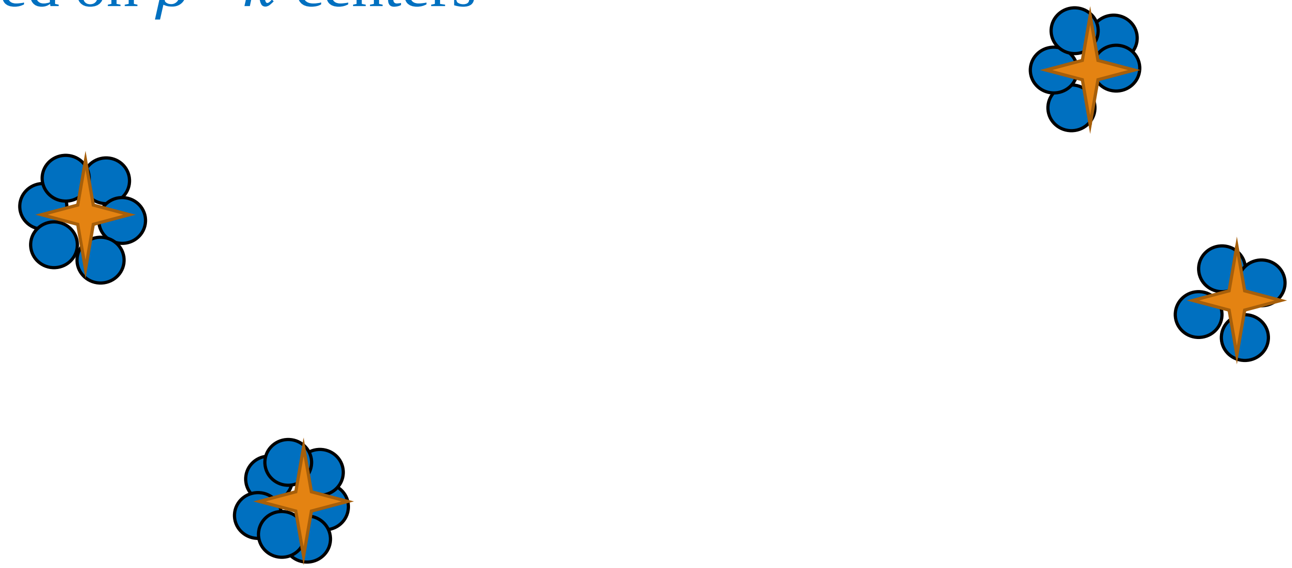
Sensitivity of the projected points A'

<u>Example: $k = 2$</u>

Compute an (α, β)-approximation

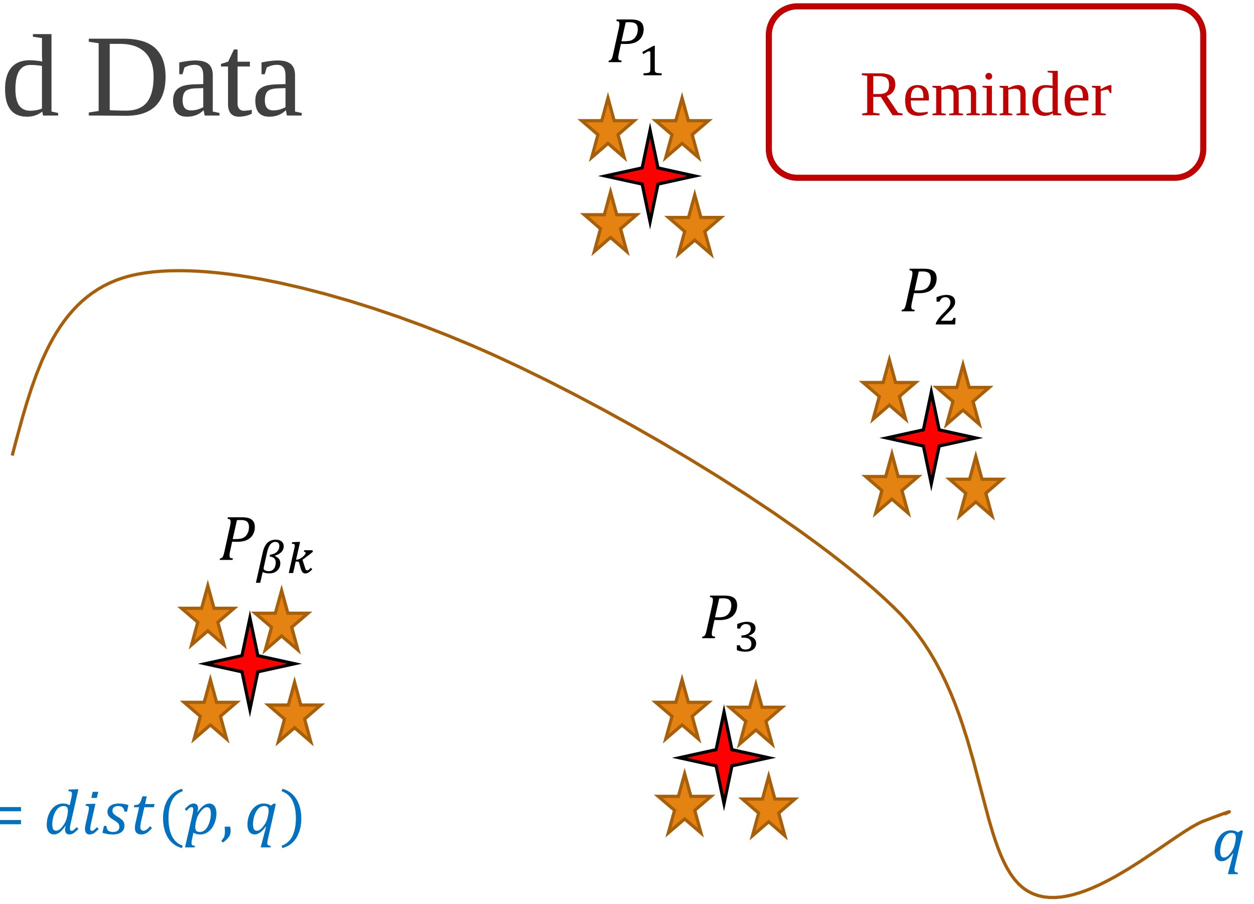
2

$\rightarrow$ n points projected on $\beta \cdot k$ centers



Sensitivity for Clustered Data

Reminder



Let:

- $p_1, \cdots p_{\beta \cdot k} \in R^d$ be $\beta \cdot k$ centers.
- $P_i = \{p_i, p_i, \dots, p_i\}$, $|P_i| = \frac{n}{\beta \cdot k}$.
- $P = P_1 \cup P_2 \cup \cdots \cup P_{\beta \cdot k}$

Query: a function q.
Cost function: For every $p \in P$: $f(p,q) = dist(p,q)$
Output: $\sum_{p \in P} f(p,q)$

$$s(p_i) = \max_q \frac{f(p_i,q)}{\sum_{p' \in P} f(p',q)} \leq \max_q \frac{f(p_i,q)}{\sum_{p_i' \in P_i} f(p_i',q)} = \frac{1}{|P_i|}$$

$$\sum_{p_i \in P_i} s(p_i) = \sum_{p_i \in P_i} \frac{1}{|P_i|} = 1$$

$$\sum_{p \in P} s(p_i) = \beta \cdot k$$

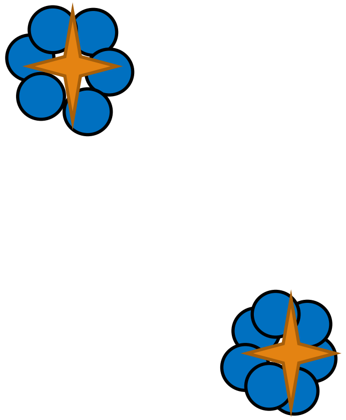
Sensitivity for k-mea

<u>Example: $k = 2$</u>

Compute an (α, β)-approximation

2

$\rightarrow n$ points projected on $\beta \cdot k$ centers



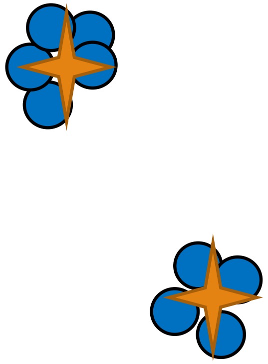
Using the last Lemma we have that:

$$\sum_{p \in A} s(p) \leq \rho\alpha + \rho^2(1+\alpha) \underbrace{\sum_{p' \in A'} \max_{T \in Q} \frac{dist(p', T)}{dist(A', T)}}$$

Sensitivity of the projected points A'

By last lecture:

$$\sum_{p \in A'} s(p) = \beta \cdot k$$



L_∞-coreset → Bound on Total Sensitivity

Lemma:

Suppose that for every $A \in R^{n\times d}$ there is an L_∞-coreset S of precision $\epsilon = \frac{1}{2}$ and of size $|S| \leq g(n)$ that can be computed in time $t(n)$. Then we can compute $\tilde{\boldsymbol{\sigma}}(\boldsymbol{f}_{A_{i*}}) \geq \boldsymbol{\sigma}(\boldsymbol{f}_{A_{i*}})$ for every $i \in [n]$, and

$$G(f) \coloneqq \sum_{i\in[n]} \tilde{\sigma}(f_{A_{i*}}) \in O(\log n) \cdot g(n)$$

Proof: $t = 1$

- Build an L_∞-coreset S_t for A_t.
- $A_{t+1} = A_t \backslash S_t$.

Intuition: The sensitivity of a point A_{i*} is proportional to 1 over the number of coresets in which A_{i*} appears.

L_∞-coreset → Bound on Total Sensitivity

Proof:

Consider the sequence of subsets $A_\ell \subseteq A_{\ell-1} \subseteq \cdots \subseteq A_1 = A$ (where $|A_\ell| \leq g(n)$) and the sequence of L_∞-coresets $S_1, \ldots, S_\ell$.

Consider the input point A_{i*} and let $v \in [\ell]$ be the largest index of a coreset S_v that contains A_{i*}. Let $q \in Q$. Let $1 \leq u \leq v$. Let $A_{i_u*} \in S_u$ be the point with maximal distance to q.

By the L_∞-coreset property:

$$dist(A_{i*}, q) \leq (1+\epsilon) \cdot dist(A_{i_u*}, q)$$

$$\rightarrow \frac{dist^2(A_{i*}, q)}{(1+\epsilon)^2} \leq dist^2(A_{i_u*}, q)$$

L_∞-coreset → Bound on Total Sensitivity

Proof:

$$\rightarrow dist^2\left(A_{i_{u^*}}, q\right) \geq \frac{dist^2(A_{i^*}, q)}{(1+\epsilon)^2}$$

L_∞-coreset → Bound on Total Sensitivity

Proof:

$\rightarrow dist^2(A_{i_u*}, q) \geq \frac{dist^2(A_{i*}, q)}{(1+\epsilon)^2}$

Since $\{A_{i_1*}, \dots, A_{i_v*}\}$ is a subset of A

L_∞-coreset → Bound on Total Sensitivity

Proof:

$$\rightarrow dist^2(A_{i_u*}, q) \geq \frac{dist^2(A_{i*}, q)}{(1+\epsilon)^2}$$

Since $\{A_{i_1*}, \dots, A_{i_v*}\}$ is a subset of A

For $A_{i*} \in S_v$:

$$\sigma(f_{A_{i*}}) \coloneqq \max_{q \in Q} \frac{dist^2(A_{i*}, q)}{dist^2(A, q)} \leq \frac{dist^2(A_{i*}, q)}{\frac{v}{(1+\epsilon)^2} \cdot dist^2(A_{i*}, q)} = \frac{(1+\epsilon)^2}{v}$$

L_∞-coreset → Bound on Total Sensitivity

Proof:

For $A_{i*} \in S_v$:

$$\sigma\left(f_{A_{i*}}\right) := \max_{q \in Q} \frac{dist^2(A_{i*}, q)}{dist^2(A, q)} \leq \frac{dist^2(A_{i*}, q)}{\frac{v}{(1+\epsilon)^2} \cdot dist^2(A_{i*}, q)} = \frac{(1+\epsilon)^2}{v}$$

L_∞-coreset → Bound on Total Sensitivity

Proof:

For $A_{i*} \in S_v$:

$$\sigma\left(f_{A_{i*}}\right) \coloneqq \max_{q \in Q} \frac{dist^2(A_{i*}, q)}{dist^2(A, q)} \leq \frac{dist^2(A_{i*}, q)}{\frac{v}{(1+\epsilon)^2} \cdot dist^2(A_{i*}, q)} = \frac{(1+\epsilon)^2}{v}$$

Define $\tilde{\sigma}\left(f_{A_{i*}}\right) = \frac{(1+\epsilon)^2}{v}$

$$G(f) = \sum_{v=1}^{\ell} \sum_{A_{i*} \in A_v} \tilde{\sigma}\left(f_{A_{i*}}\right)$$

L_∞-coreset → Bound on Total Sensitivity

Proof:

For $A_{i*} \in S_v$:

$$\sigma(f_{A_{i*}}) := \max_{q \in Q} \frac{dist^2(A_{i*}, q)}{dist^2(A, q)} \leq \frac{dist^2(A_{i*}, q)}{\frac{v}{(1+\epsilon)^2} \cdot dist^2(A_{i*}, q)} = \frac{(1+\epsilon)^2}{v}$$

Define $\tilde{\sigma}(f_{A_{i*}}) = \frac{(1+\epsilon)^2}{v}$

$$G(f) = \sum_{v=1}^{\ell} \sum_{A_{i*} \in A_v} \tilde{\sigma}(f_{A_{i*}}) \leq \sum_{v=1}^{\ell} \frac{|S_v|(1+\epsilon)^2}{v}$$

$|S_v| \leq g(n)$ ←

$$\leq g(n) \cdot (1+\epsilon)^2 \cdot \sum_{v=1}^{\ell} \frac{1}{v}$$

L_∞-coreset → Bound on Total Sensitivity

Proof:

For $A_{i*} \in S_v$:

$$\sigma\left(f_{A_{i*}}\right) := \max_{q \in Q} \frac{dist^2(A_{i*}, q)}{dist^2(A, q)} \leq \frac{dist^2(A_{i*}, q)}{\frac{v}{(1+\epsilon)^2} \cdot dist^2(A_{i*}, q)} = \frac{(1+\epsilon)^2}{v}$$

Define $\tilde{\sigma}\left(f_{A_{i*}}\right) = \frac{(1+\epsilon)^2}{v}$

$$G(f) = \sum_{v=1}^{\ell} \sum_{A_{i*} \in A_v} \tilde{\sigma}\left(f_{A_{i*}}\right) \leq \sum_{v=1}^{\ell} \frac{|S_v|(1+\epsilon)^2}{v}$$

$$\leq g(n) \cdot (1+\epsilon)^2 \cdot \sum_{v=1}^{\ell} \frac{1}{v} \leq g(n) \cdot (1+\epsilon)^2 \cdot \sum_{v=1}^{n} \frac{1}{v}$$

L_∞-coreset → Bound on Total Sensitivity

Proof:

For $A_{i*} \in S_v$:

$$\sigma\left(f_{A_{i*}}\right) := \max_{q \in Q} \frac{dist^2(A_{i*}, q)}{dist^2(A, q)} \le \frac{dist^2(A_{i*}, q)}{\frac{v}{(1+\epsilon)^2} \cdot dist^2(A_{i*}, q)} = \frac{(1+\epsilon)^2}{v}$$

Define $\tilde{\sigma}\left(f_{A_{i*}}\right) = \frac{(1+\epsilon)^2}{v}$

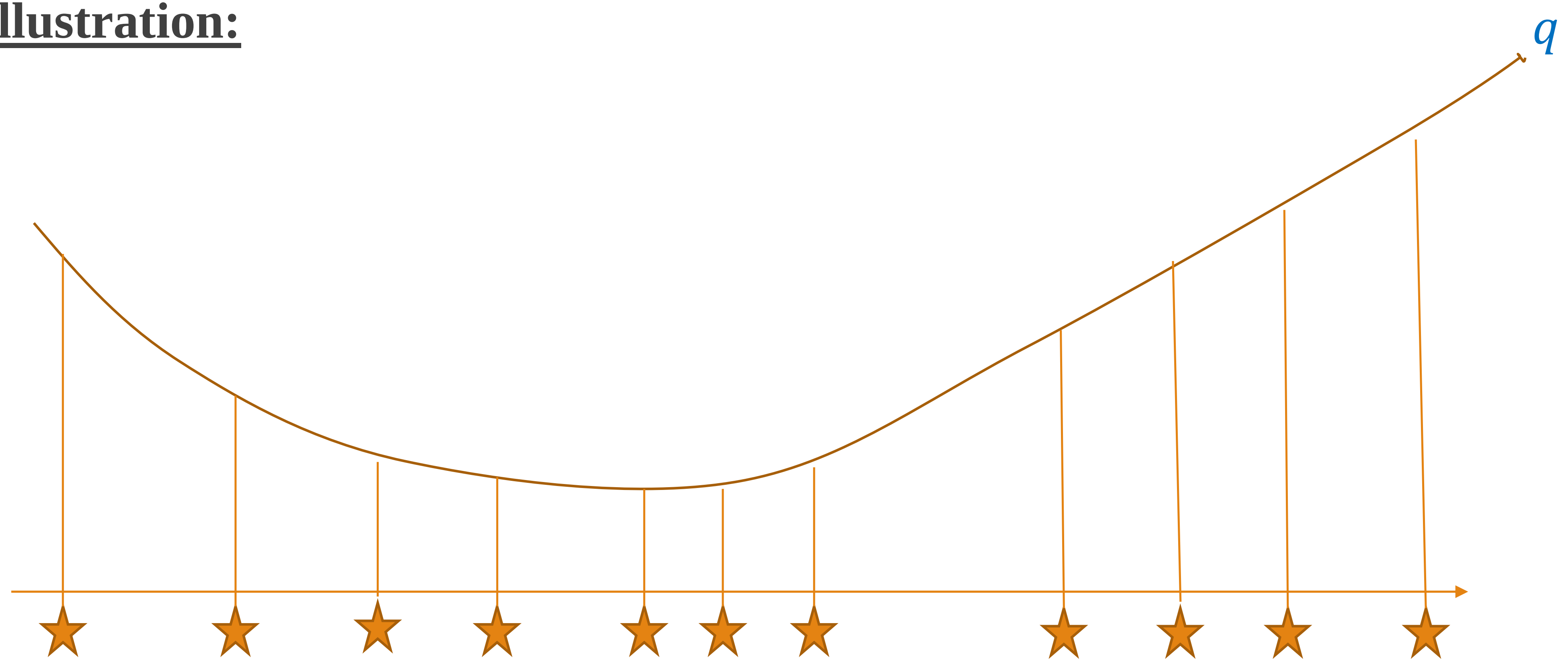
$$G(f) = \sum_{v=1}^{\ell} \sum_{A_{i*} \in A_v} \tilde{\sigma}\left(f_{A_{i*}}\right) \le \sum_{v=1}^{\ell} \frac{|S_v|(1+\epsilon)^2}{v}$$

$$\le g(n) \cdot (1+\epsilon)^2 \cdot \sum_{v=1}^{\ell} \frac{1}{v} \le g(n) \cdot (1+\epsilon)^2 \cdot \sum_{v=1}^{n} \frac{1}{v}$$

$$\le g(n) \cdot (1+\epsilon)^2 \cdot (\ln n + 1)$$

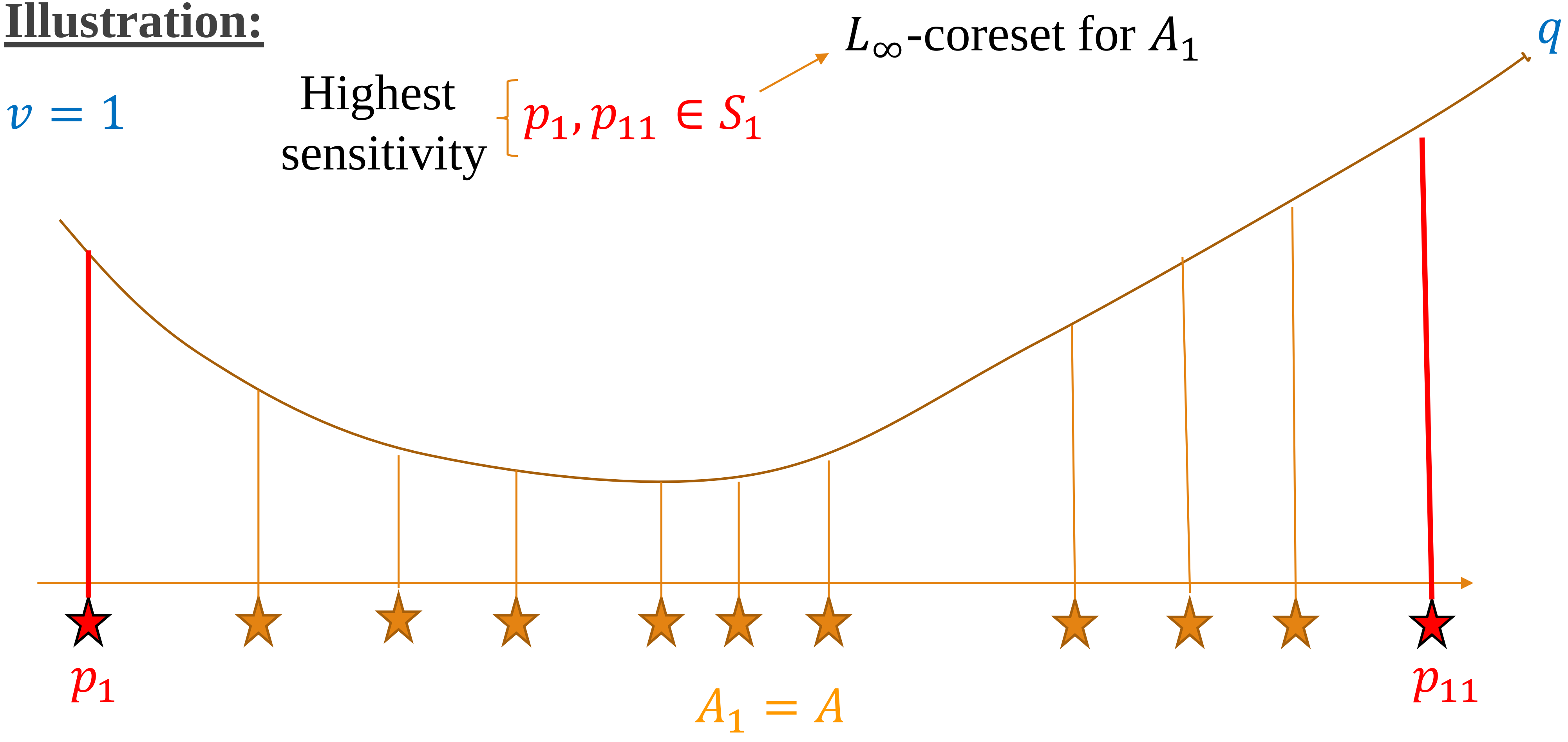
L_∞-coreset → Bound on Total Sensitivity

Illustration:



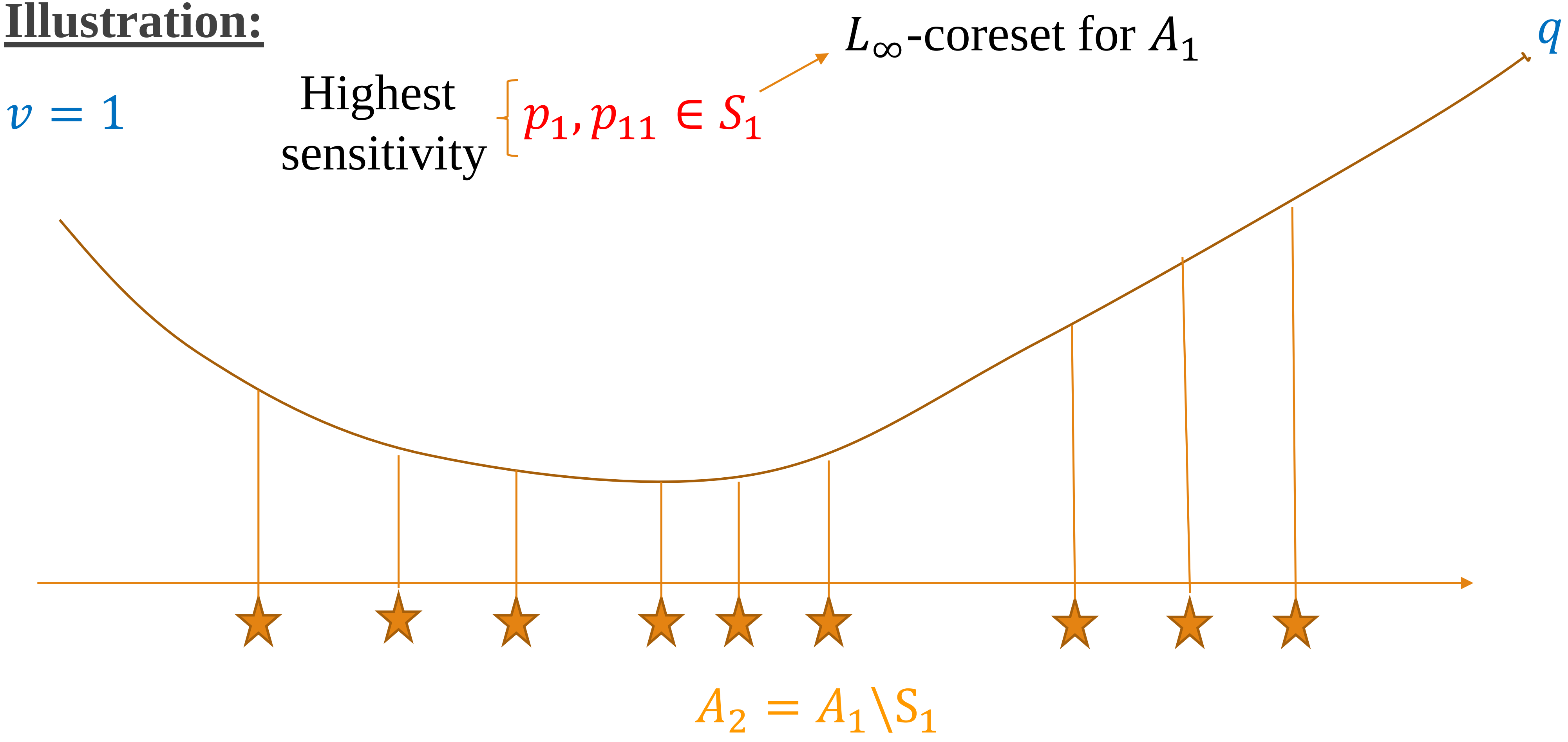
L_∞-coreset → Bound on Total Sensitivity

Illustration:



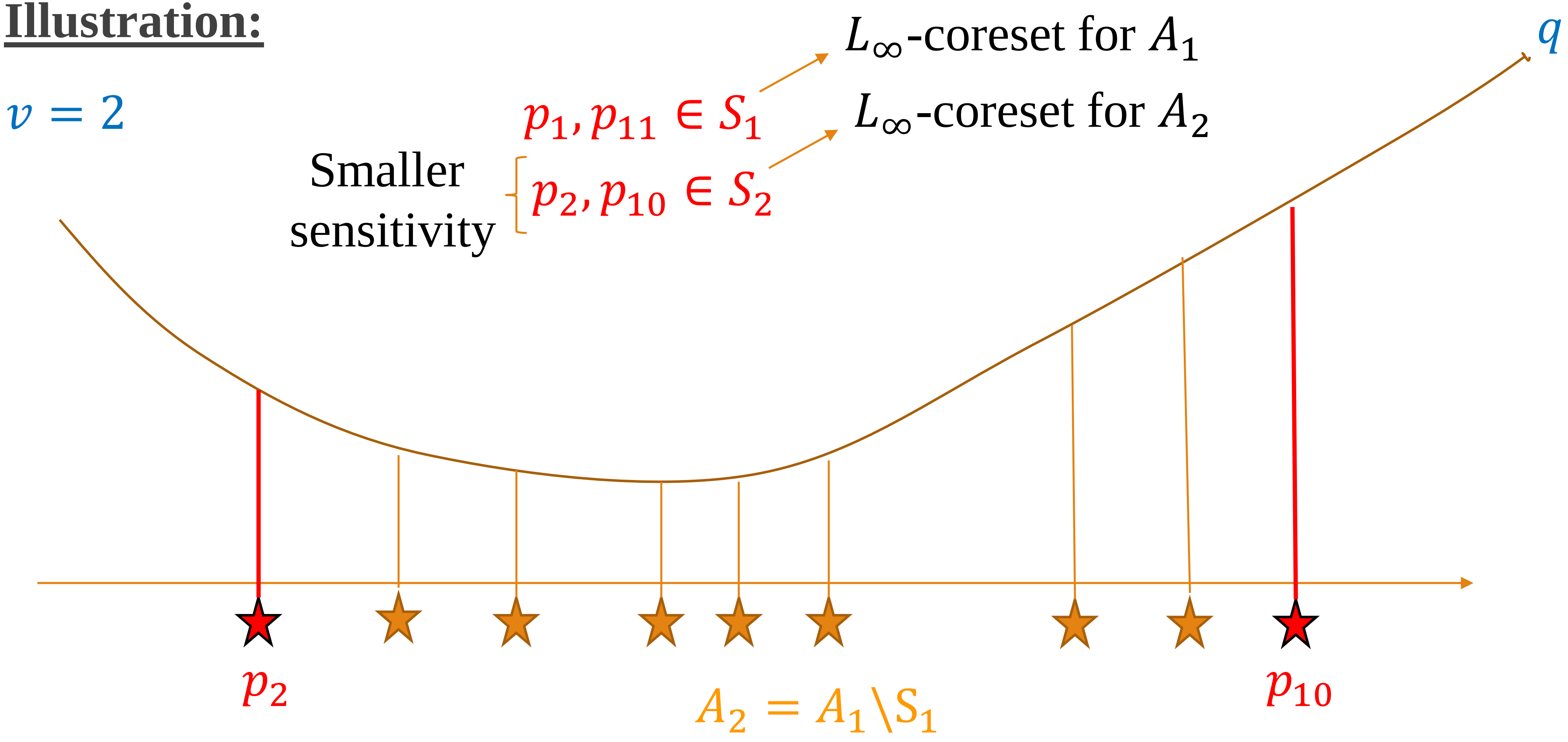
L_∞-coreset → Bound on Total Sensitivity

Illustration:



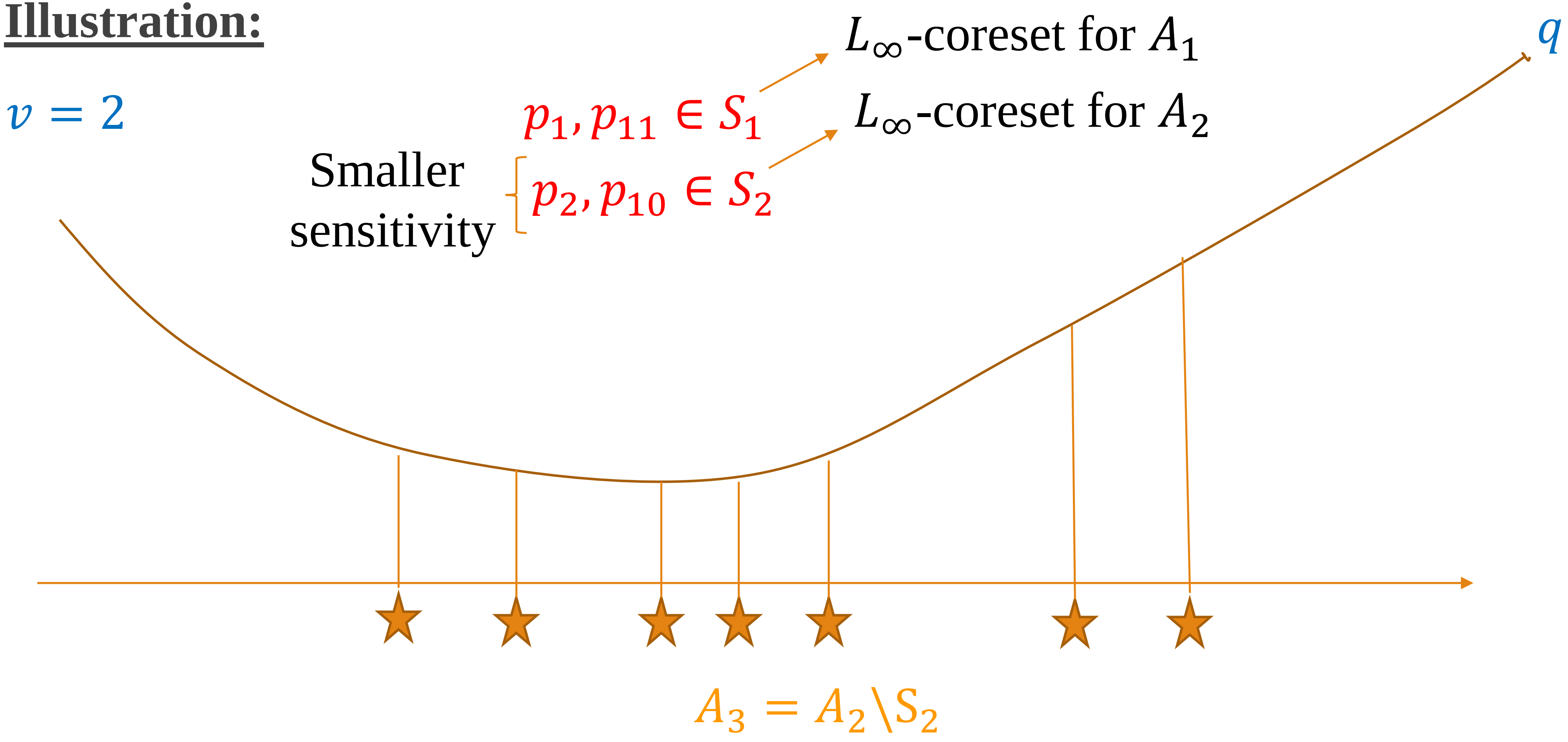
L_∞-coreset → Bound on Total Sensitivity

Illustration:



L_∞-coreset → Bound on Total Sensitivity

Illustration:



L_∞-coreset → Bound on Total Sensitivity

Illustration:

$v = t + 1$