

Clusterability

1 Problem Statement

1.1 Notation

For any set $B \subset X$, we denote $c(B)$ as the center of B which is defined as the average of points in B . Radius of the set B is defined as $r(B) = \max_{x \in B} |x - c(B)|$.

Definition 1 (Niceness assumption). Given a set \mathcal{X} , we say that a partition of \mathcal{X} , C_1, \dots, C_k is (λ, ν) -nice if the following conditions hold. There exist sets $B_1, \dots, B_k \subset \mathcal{X}$ such that for every $i \in [k]$, there exists $j_i \in [k]$ such that $B_i \subset C_{j_i}$ and

- **Separation:** For all $i, j \in [k]$, $|c(B_i) - c(B_j)| \geq \nu \cdot \max\{r(B_i), r(B_j)\}$
- **Sparse Noise:** For any ball $B \subset \mathcal{X}$ for which $r(B) \leq \lambda \cdot \max_{i \in [k]} r(B_i)$, $|B \cap \{X \setminus \cup_{i \in [k]} B_i\}| \leq \min_{i \in [k]} |B_i|$