Approximate nearest neighbor search

Tianpei Xie

Sep. 3rd., 2022

Contents

1 Definitions

- Given a dataset $\mathcal{D}$ with n points and d dimensions and a query point q in the same space as the dataset, the *goal* of ***c-ANN search*** (where $c > 1$ is an approximation ratio) is to return points $o \in \mathcal{D}$ such that $\text{dist}(o, q) \leq c \times \text{dist}(o^*, q)$, where o^* is the true nearest neighbor of q in $\mathcal{D}$ and dist is the distance between the two points. Similarily, ***c-k-ANN search*** aims at returning top-k points such that $\text{dist}(o_i, q) \leq c \times \text{dist}(o_i^*, q)$, where $1 \leq i \leq k$. [Jafari et al., 2021]

References

Omid Jafari, Preeti Maurya, Parth Nagarkar, Khandker Mushfiqul Islam, and Chidambaram Crushev. A survey on locality sensitive hashing algorithms and their applications. *arXiv preprint arXiv:2102.08942*, 2021.