Nearest neighbor search in high dimensional spaces

Ilya Fedorov

March 2020

Content

- Why do we need to optimize KNN?
- Complexity analysis
- Data structures
- Difficulties in high dimensional spaces
- Approximate NN search: Locality Sensitive Hashing
- Approximate NN search: FlyHash
- Approximate NN search: Hierarchical Navigable Small World

Why do we need to optimize KNN??

- Very large datasets
- Very frequent queries
- Duplicates search
- It seems that brute force algorithm doens't use information gained from calculating previous distances: if $d(a, b)$ is big and $d(a, c)$ is small does it mean that $d(c, b)$ is big?