Introduction

Yufei Tao

ITEE
University of Queensland

Multidimensional Data

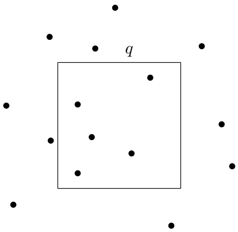
We will consider d-dimensional space $\mathbb{R}^d$, where d is the dimensionality, and $\mathbb{R}$ represents the real domain.

Multidimensional data, in general, refer to geometric objects in $\mathbb{R}^d$, for example, points, line segments, rectangles, circles, triangles, etc.

This course is dedicated to studying algorithms for solving a series of problems that are fundamental in analyzing such data. Next, we will give an overview of some of these problems.

Orthogonal Range Reporting (Window Queries)

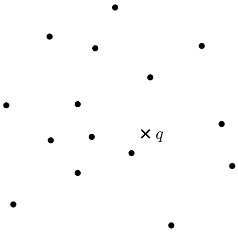
We have a set S of points in d-dimensional space $\mathbb{R}^d$. A query specifies an orthogonal rectangle $q = [x_1, y_1] \times [x_2, y_2] \times ... \times [x_d, y_d]$, and returns $S \cap q$, namely, all the points of S that fall in q.



Think 1: Applications?
Think 2: How would you process such queries?
Think 3: What if S contains other types of geometric objects such as line segments, rectangles, etc.?

Nearest Neighbor Queries

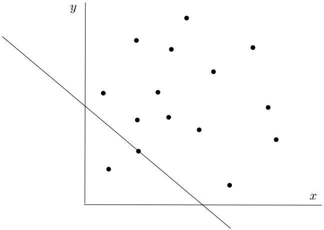
We have a set S of points in d-dimensional space $\mathbb{R}^d$. A query specifies a point q and an integer k, and returns the k points in S that are closest to q, in terms of Euclidean distance.



Think 1: Applications?
Think 2: How would you process such queries?

Top-k Preference Queries

We have a set S of points in d-dimensional space $\mathbb{R}^d$. A query specifies an integer k, and a linear function $f : \mathbb{R}^d \to \mathbb{R}$ with the form $f(p) = \sum_{i=1}^{d} c_i \cdot p[i]$, where $p[i]$ represents the i-th coordinate of a point, and c_i is a constant. The query returns the k points $p \in S$ whose $f(p)$ values are the largest.



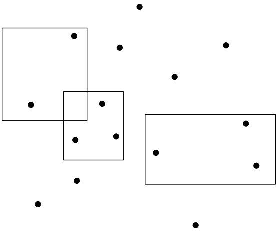
Think 1: What are the points returned by the query with $k = 2$ and $f(p) = p[1] + p[2]$?

Think 2: Applications?

Think 3: How would you process such queries?

Spatial Join

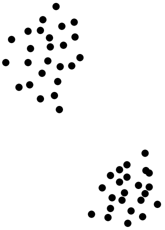
We have a set P of points and a set R of of rectangles all in the same d-dimensional space $\mathbb{R}^d$. The goal is to report all $(p, r) \in P \times R$ such that point p falls in r.



Think 1: What is the result in the above example?
Think 2: Applications?
Think 3: How would you process such queries?

Clustering

We have a set P of points in d-dimensional space $\mathbb{R}^d$. The goal is to partition P into groups each of which is called a cluster, such that points in the same cluster are "closely related", while points in different clusters are not.



Intuitively there are three clusters in the above example.

Think: Applications?

In this course, we will discuss efficient data structures and algorithms for solving problems like the above. These techniques can be easily implemented, and have been proved to be highly efficient in practice.