**CS 49995 & CS 63016 ST: Big Data Analytics**

**Homework 5**

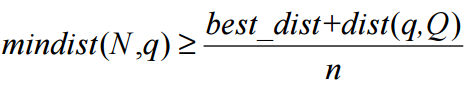
**Instructor:** Xiang Lian

**Due Date:** Please refer to the course website

1. Read the lecture slides about NN pruning heuristics, and write the pseudo code of computing the minimum distance, *mindist*(*q*, *E*), between a query point *q* and an MBR node *E* in *d*-dimensional space. [20 points]

2. Read the lecture slides about NN pruning heuristics, and write the pseudo code of computing the maximum distance, *maxdist*(*q*, *E*), between a query point *q* and an MBR node *E* in *d*-dimensional space. [20 points]

3. Read the lecture slides about the pruning conditions for Simple Point Method (SPM) of GNN queries, and prove the pruning condition below (**Hint:** *you may need to read the original paper of GNN and use the triangle inequality for the proof*) [30 points]



4. Given an aR-tree (as mentioned in the reference below) *I* and a hyperrectangular query region *Q*, please write the pseudo code of retrieving the number of objects in *Q* by traversing the aR-tree. [30 points]

I. Lazaridis and S. Mehrotra. Progressive Approximate Aggregate Queries with a Multi-Resolution Tree Structure. In *SIGMOD*, 2001.

**Bonus Question [20 extra points]**

5. Find one paper (in SIGMOD, PVLDB, and ICDE) related to big data analytics (indexing or querying) you are interested in.

* Provide the reference of this paper;
* Briefly describe the problem and solutions of this paper; and
* Explain the main contributions of this paper.

**Submission**

Submit an electronic copy of your homework solution to the [Blackboard](https://learn.kent.edu/).