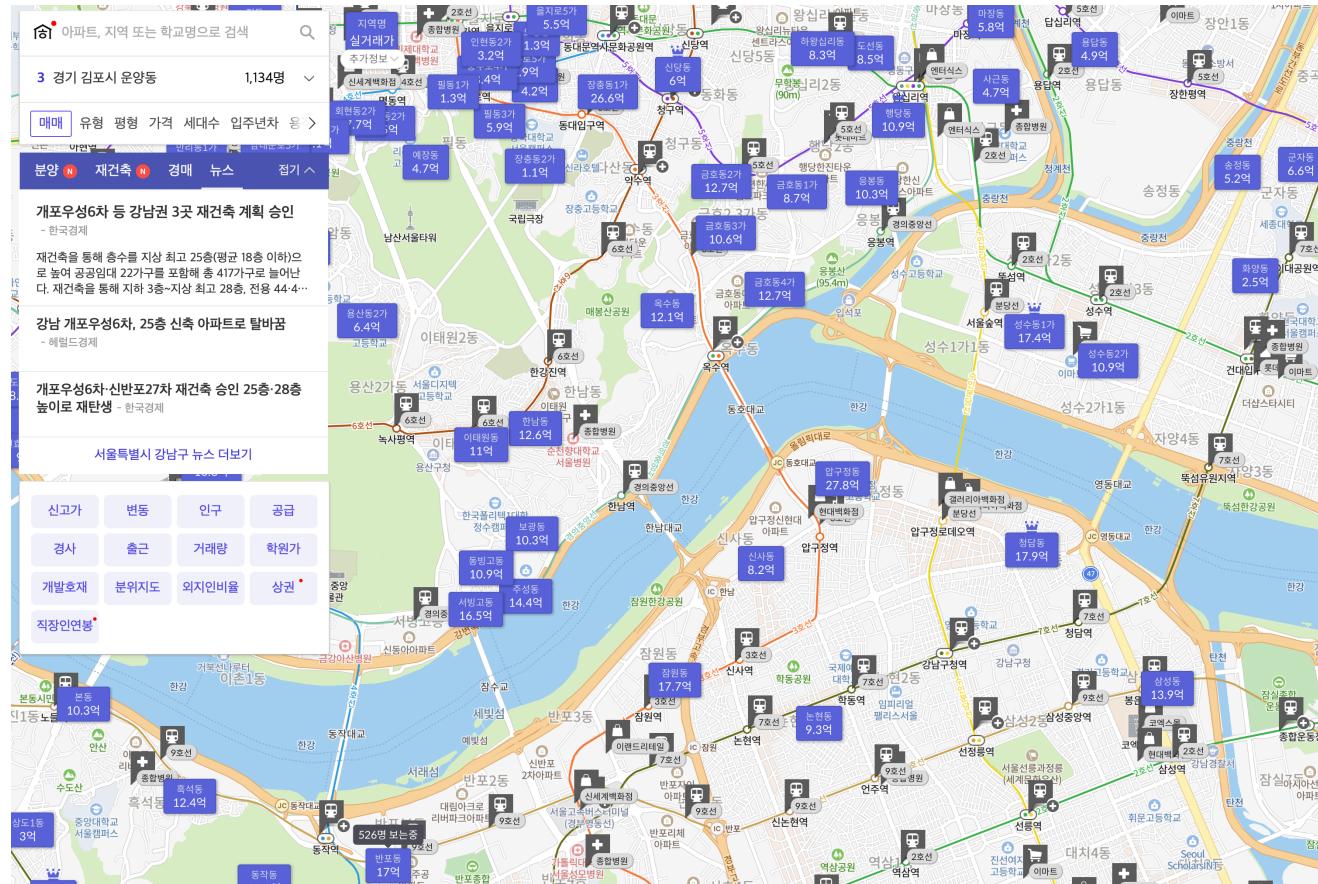


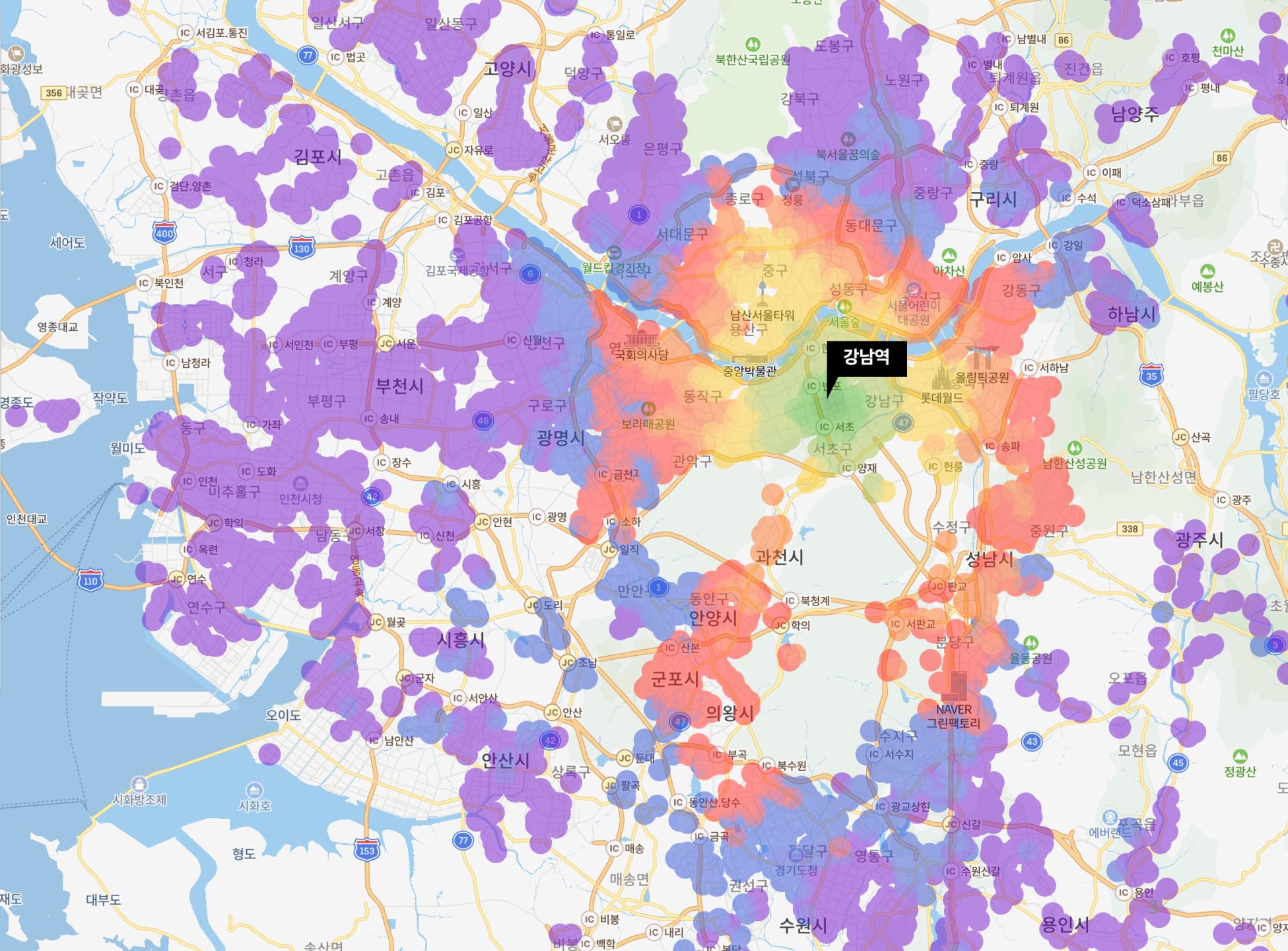
Geospatial Visualization of Approximate weighted Euclidean distance heatmap

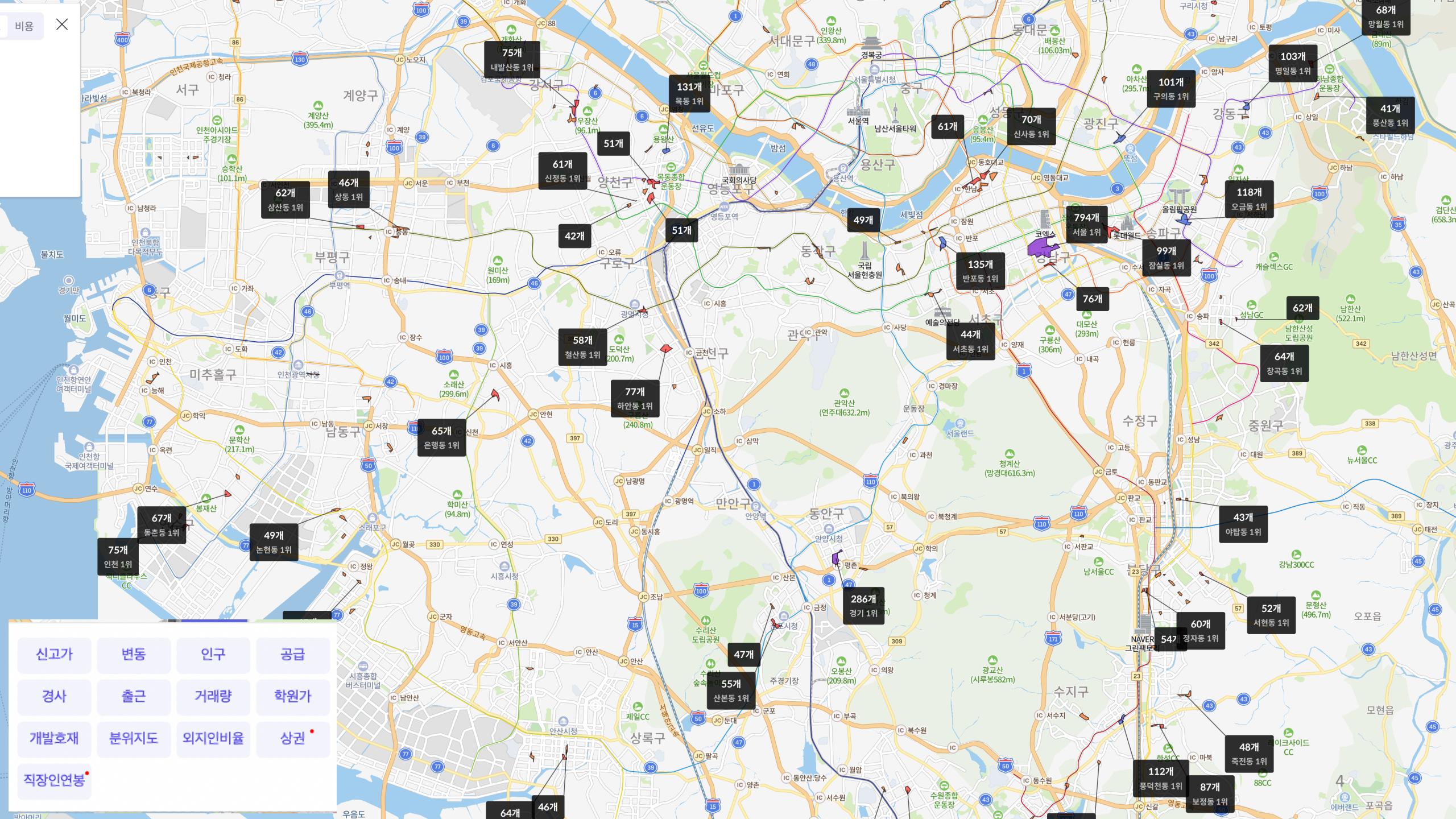
Wooil Kim

Introduction

- Distance from other items matters in location selection problem

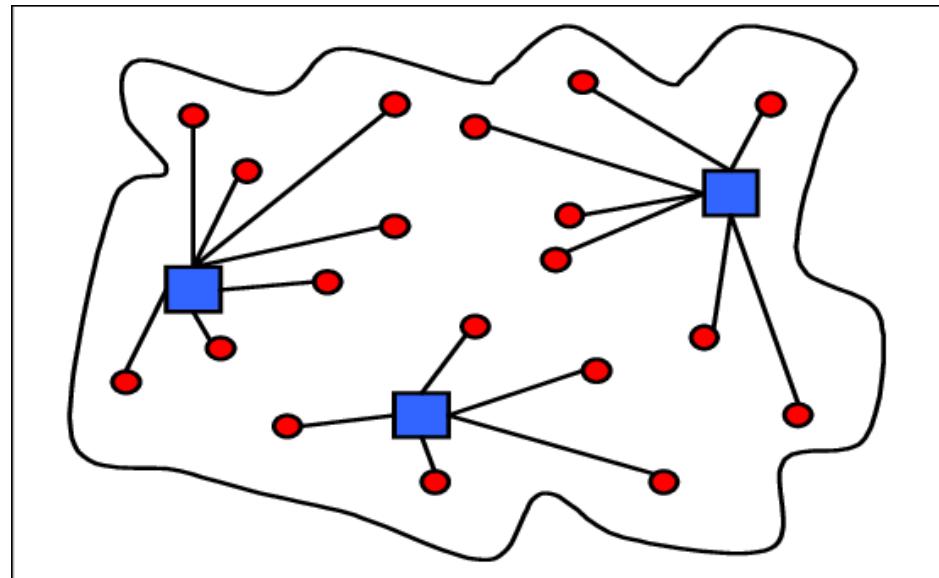






Introduction

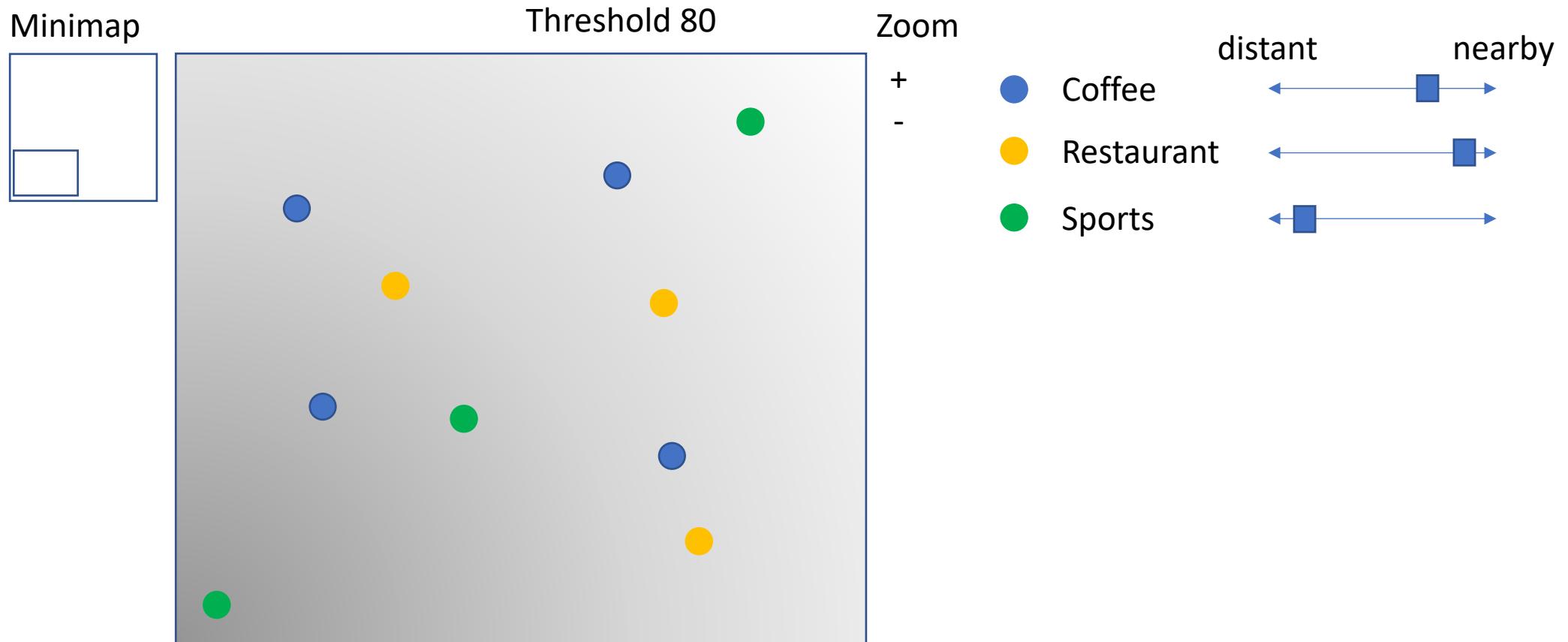
- Facility location problem
 - Facility location problem is to find locations for new facilities such that the conveying cost from facilities to customers is minimized.
 - Proved to be NP-hard



Problem Definition

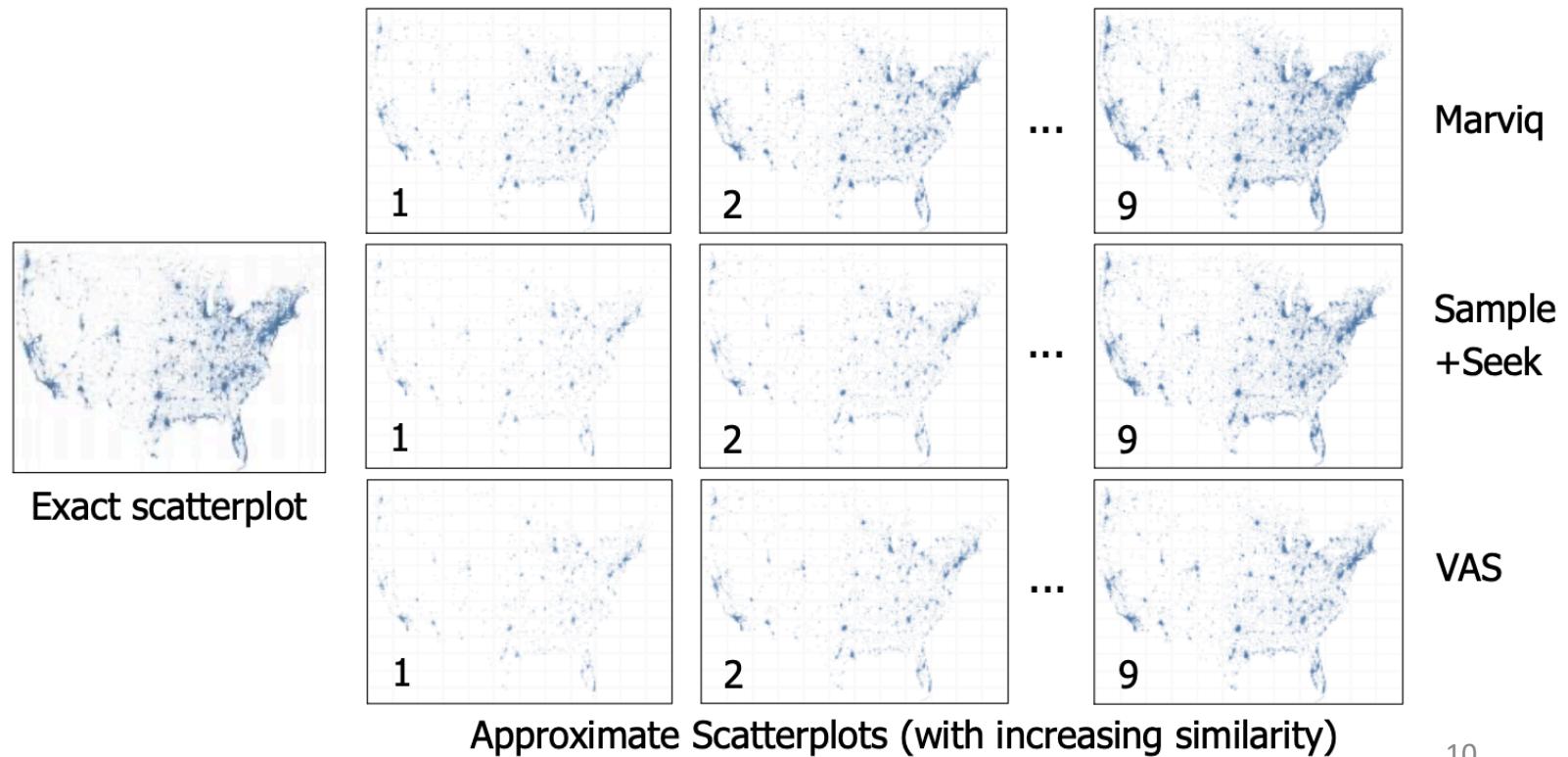
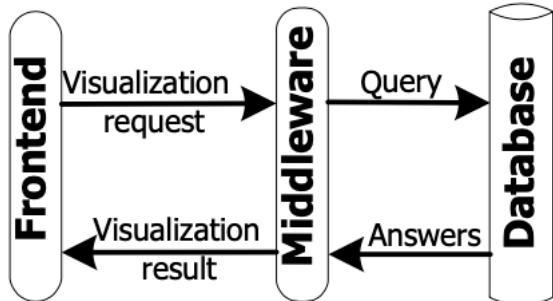
- Given
 - Items (ID, x, y, categories)
- Input
 - Item correlations, visual threshold
- Output
 - For all x and y, sum of positive-dist and negative-dist to the items

System UI



Related Works

- Marviq: Quality-Aware Geospatial Visualization of Range-Selection Queries Using Materialization (2020 SIGMOD)

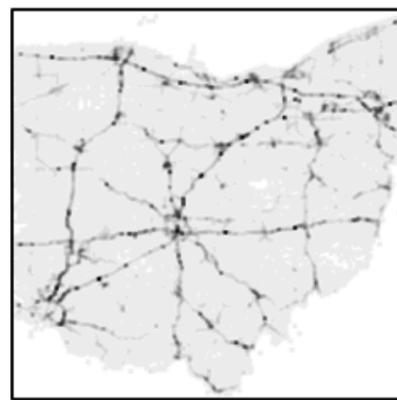


Related Works

- STULL: Unbiased Online Sampling for Visual Exploration of Large Spatiotemporal Data (2000 IEEE VIS)



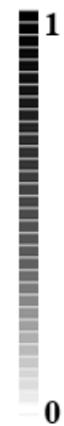
STULL at 0.3% Data



Exact Heatmap with 100% Data

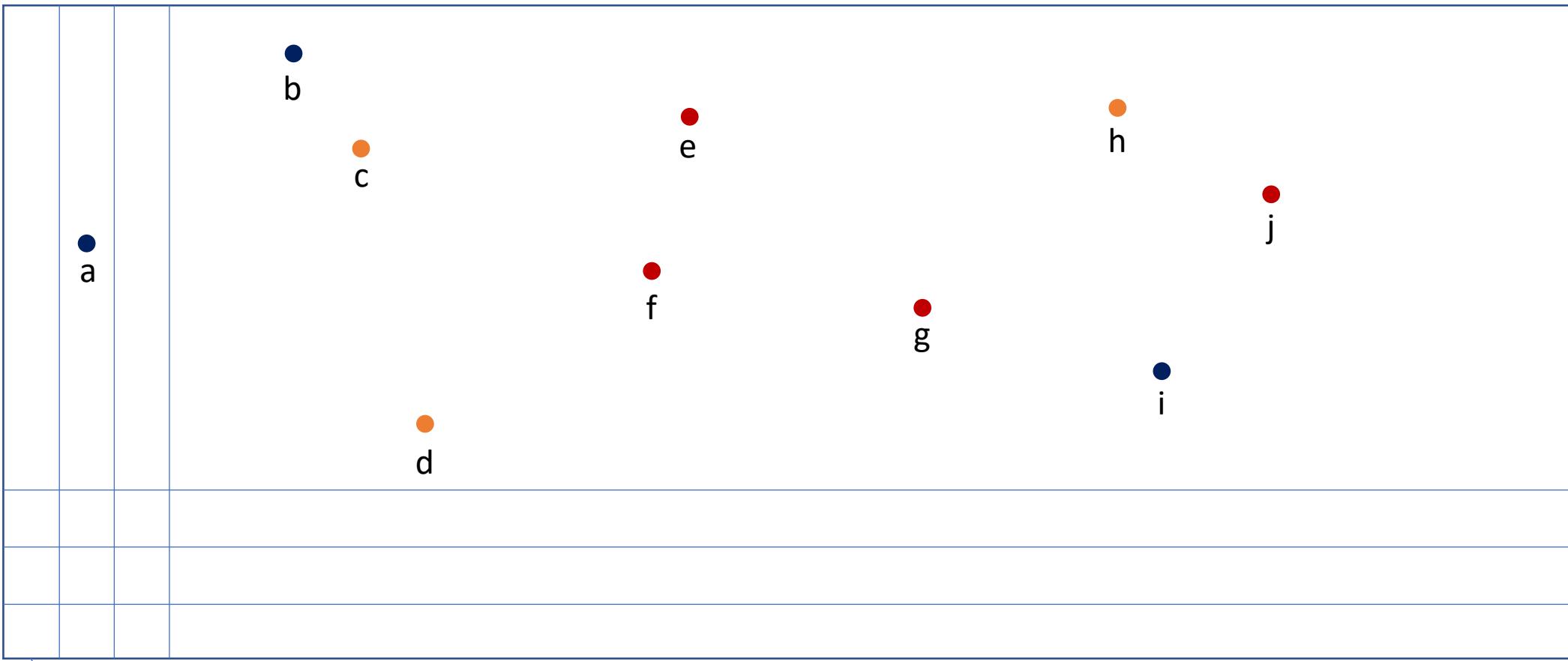


STORM at 0.3% Data



Exact Visualization

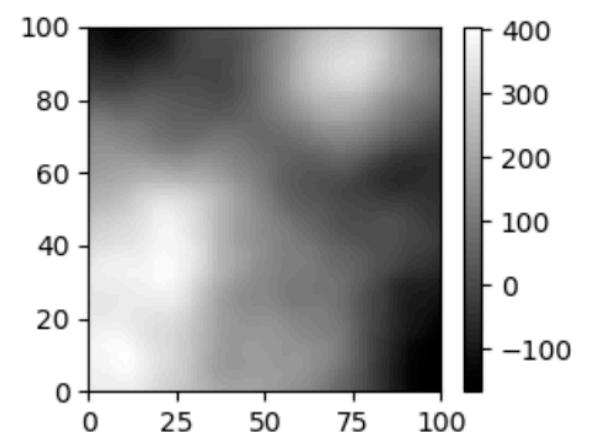
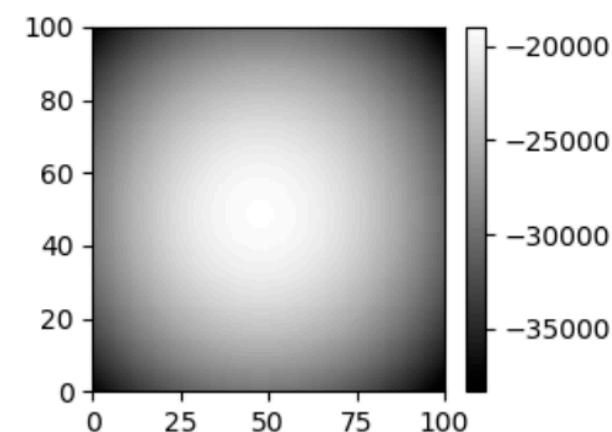
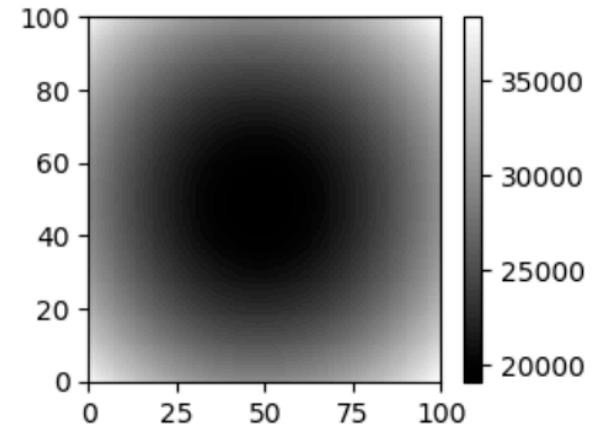
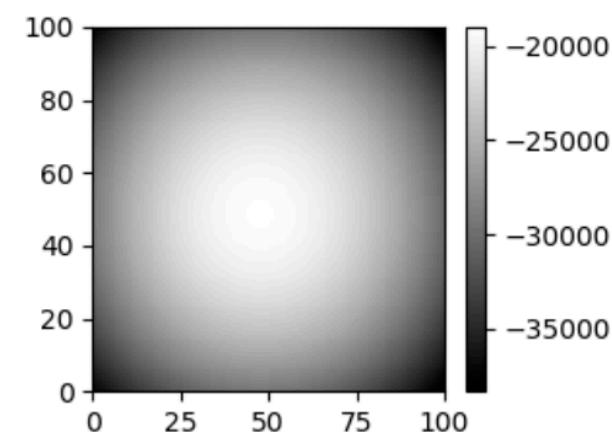
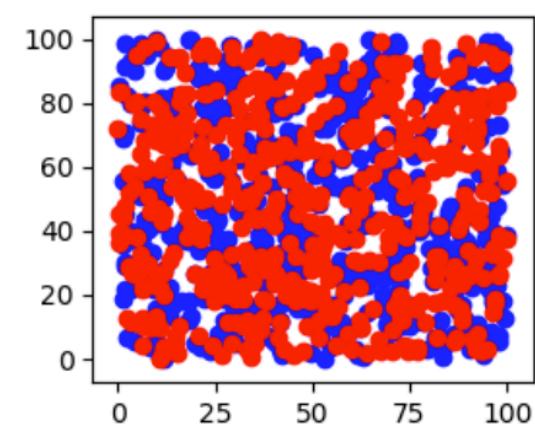
Weights [-7, 5, 10]



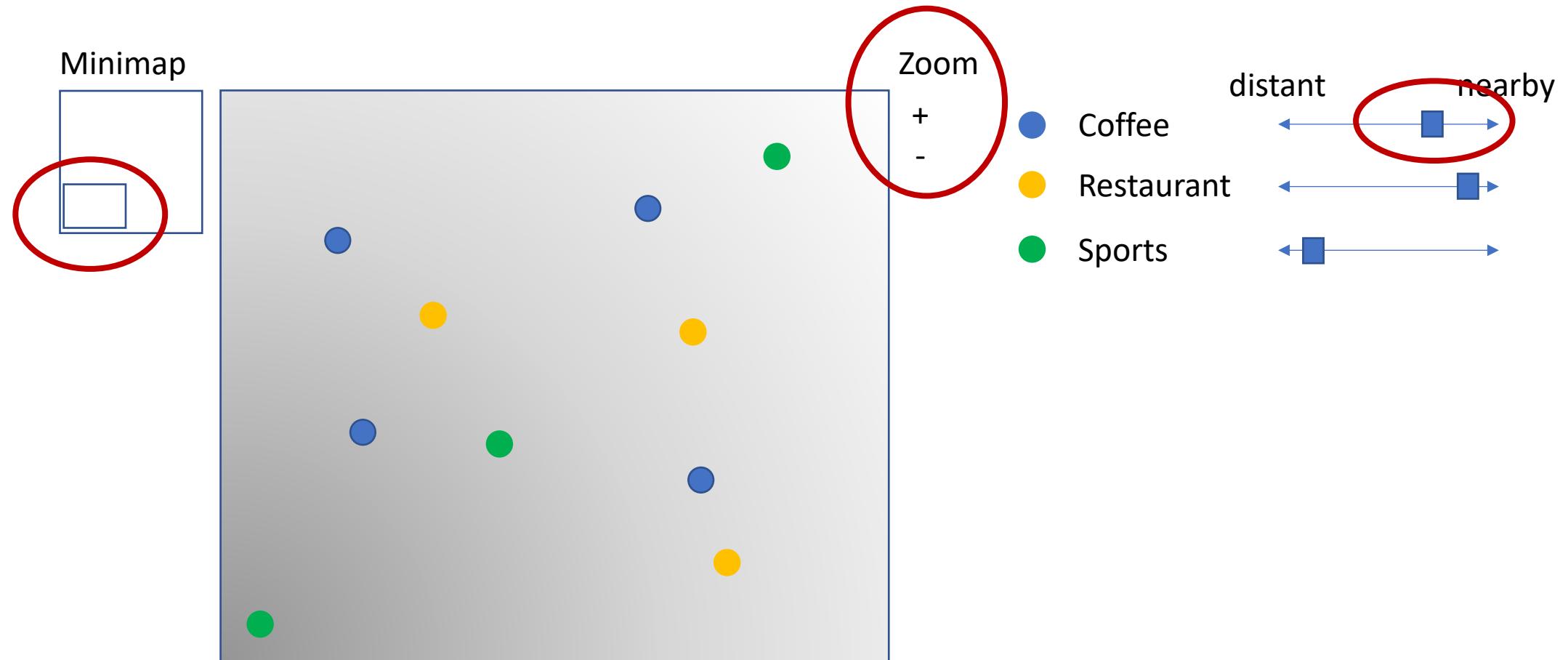
$$10 * \text{dist}((0,0),abi) + 5 * \text{dist}((0,0),cdh) - 7 * \text{dist}((0,0),efgj)$$

Processing time

Number of points	Size of space	Processing time (s)
500	100 * 100	16
100	300 * 300	33
300	300 * 300	94
500	300 * 300	153
100	500 * 500	93
300	500 * 500	256
500	500 * 500	425
100,000	100,000 * 100,000	3.4billion (40,000 days)



System UI



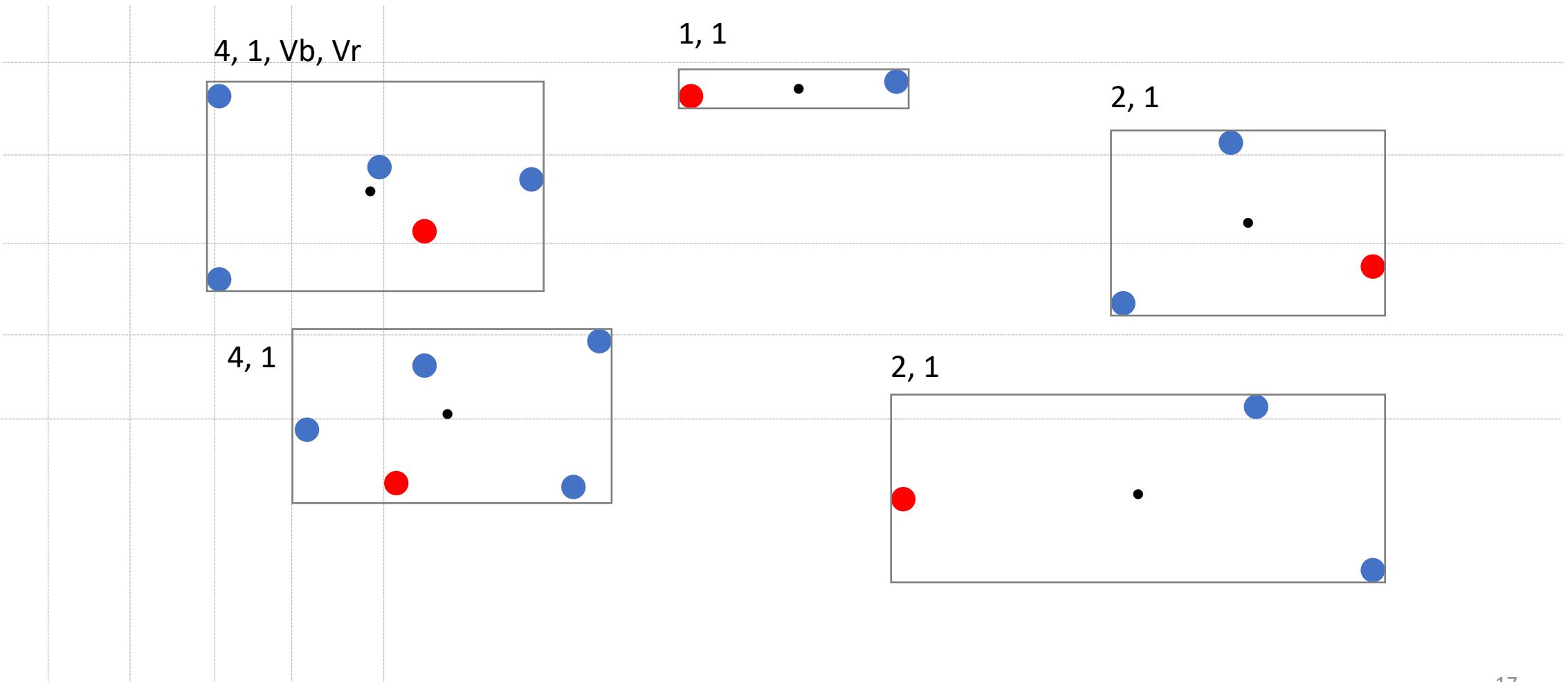
Quality Function

- $E_{x,y} = V_{exact}(x, y) - V_{aprx}(x, y)$
- $Q = \frac{\sum E_{x,y}}{|X| * |Y|} * \frac{1}{\max_{x,y} E_{x,y}}$

Proposed Approximate Visualization Method

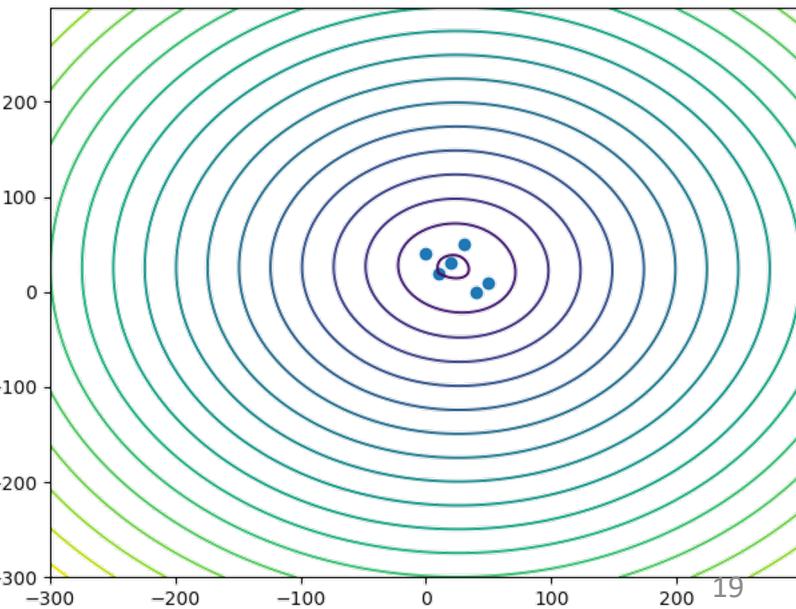
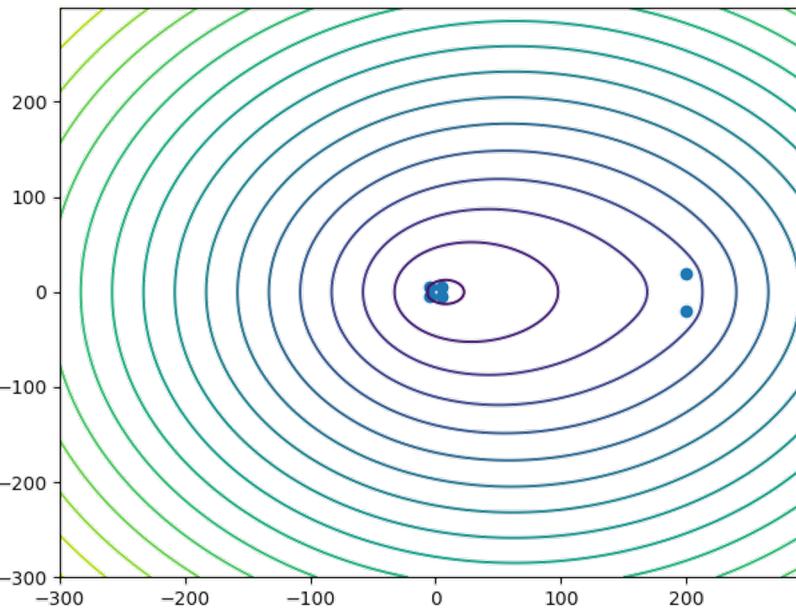
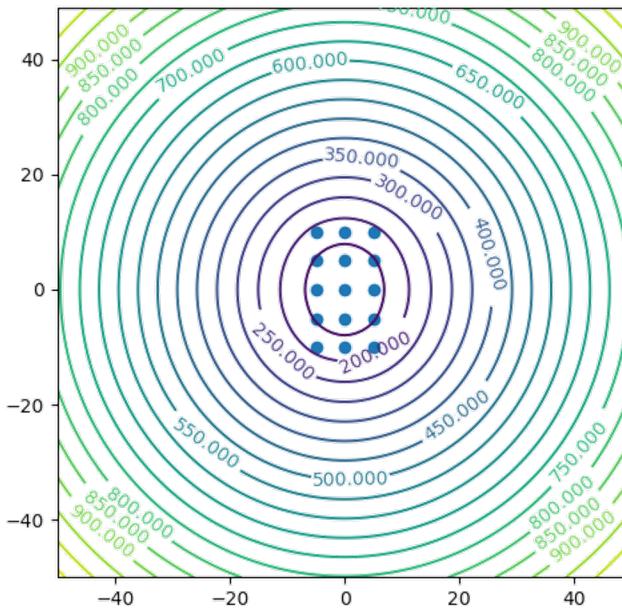
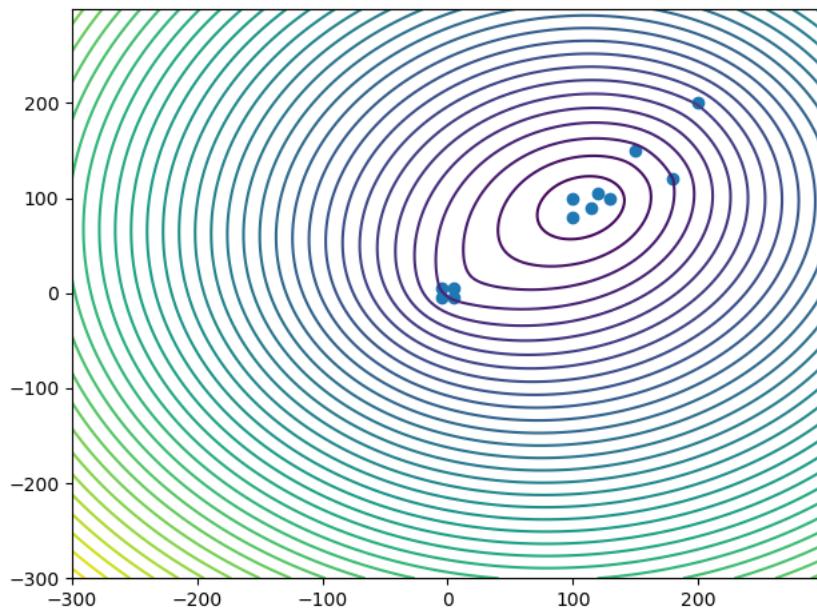
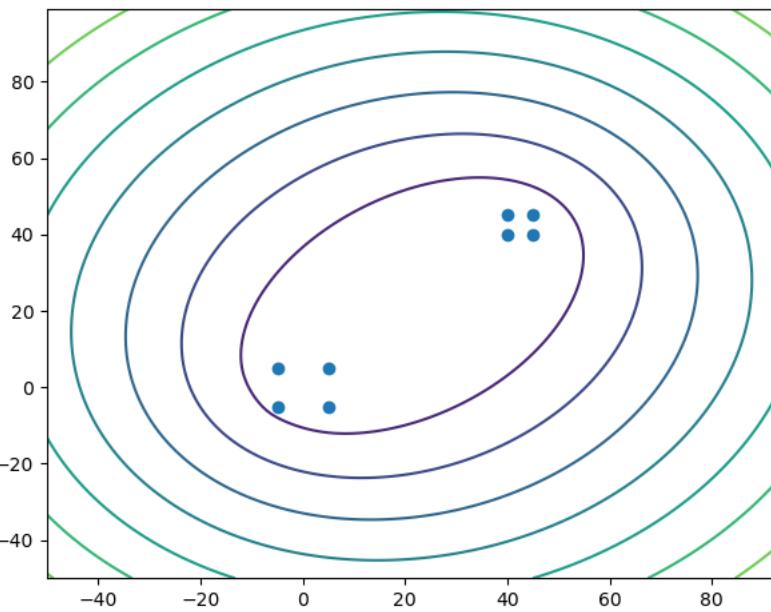
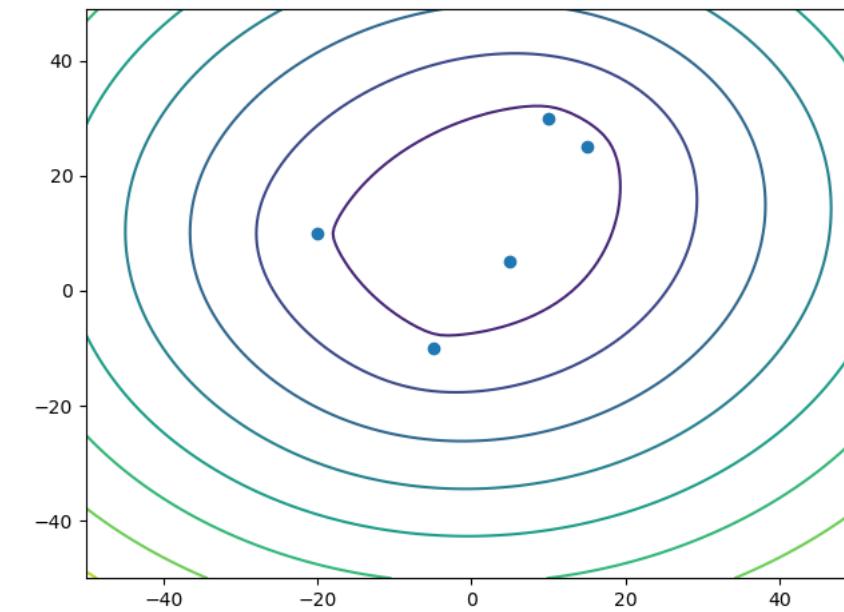
- Spatial index (Grid, R-tree, quad tree)
- Weight boundary
- Regression
- Data Sampling

Spatial index

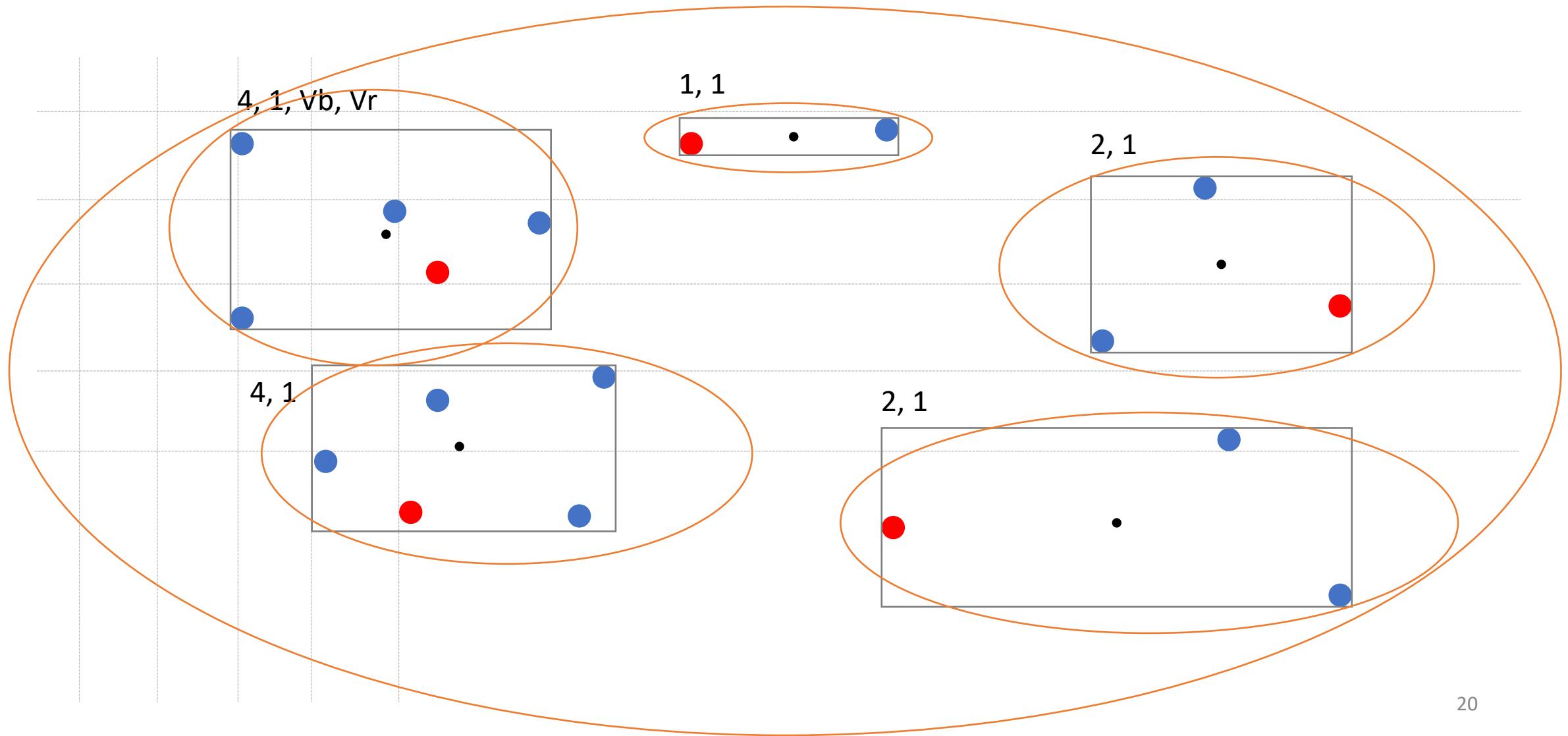


Observations

- Observation 1. 임의의 점의 집합 내부에서의 거리의 합이 최소가 되는 지점은 내부에 있다.
- Observation 2. 점 집합 외부에 타원모양이 형성되고 나면 그이후에는 닮음 꼴로(homomorphism) 등고선이 그려진다.
- Observation 3. 값은 달라지지만 등고선의 모양에 영향이 없게 점을 대체할 수 있다.

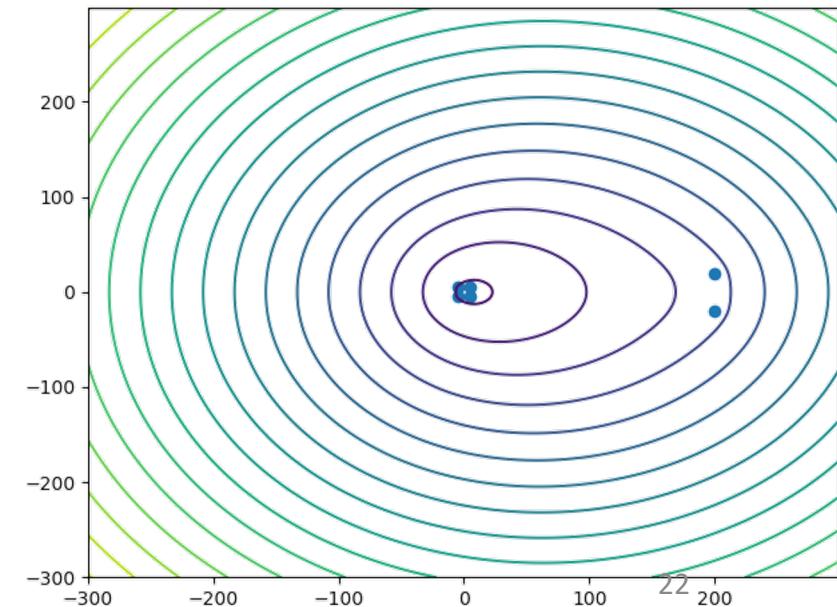
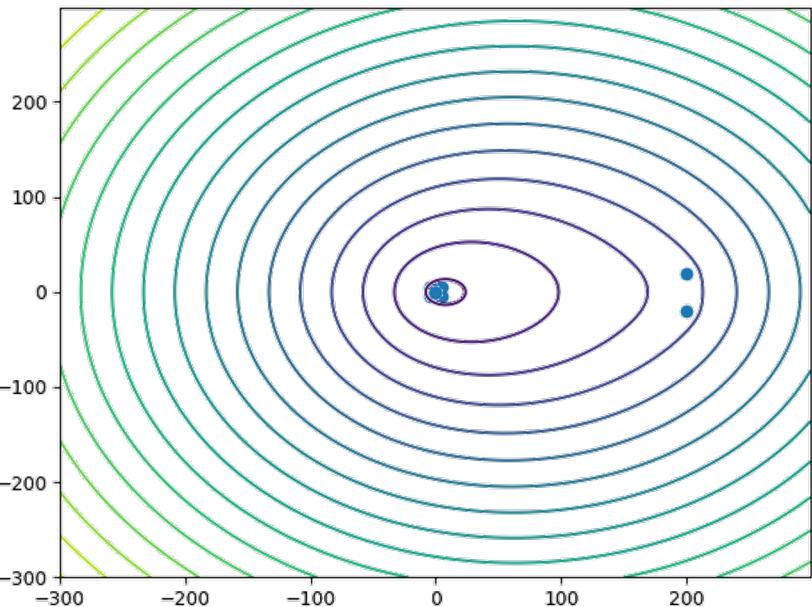
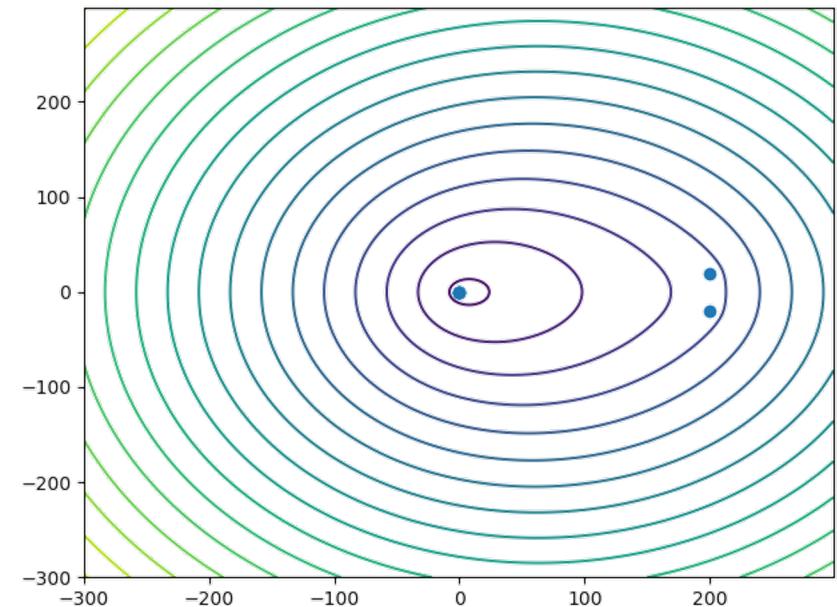
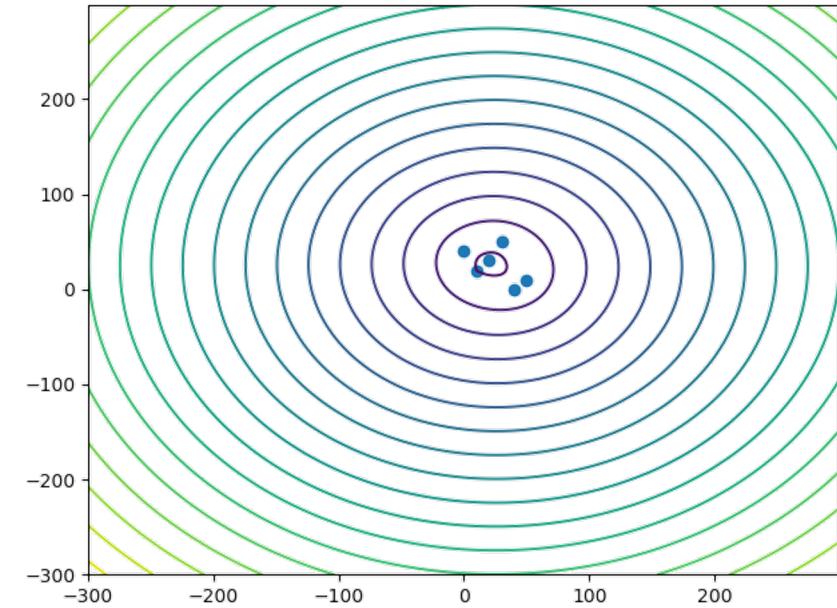
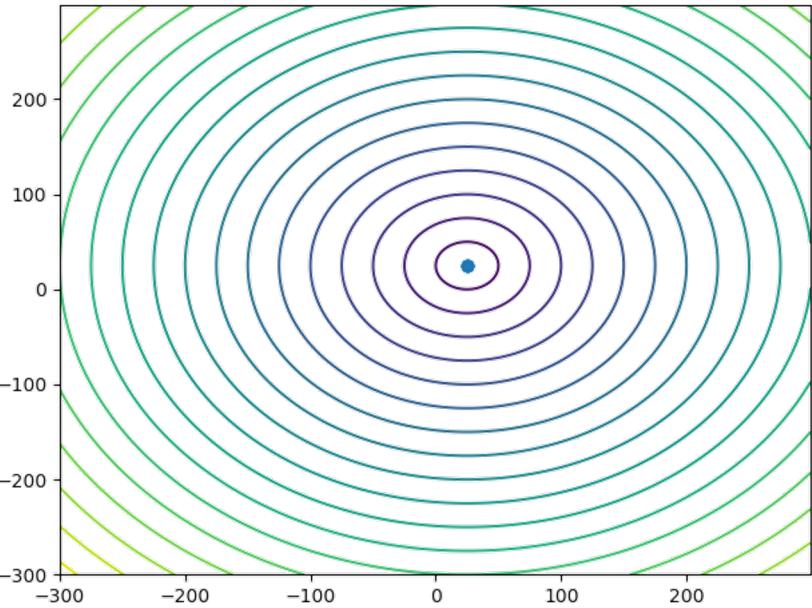
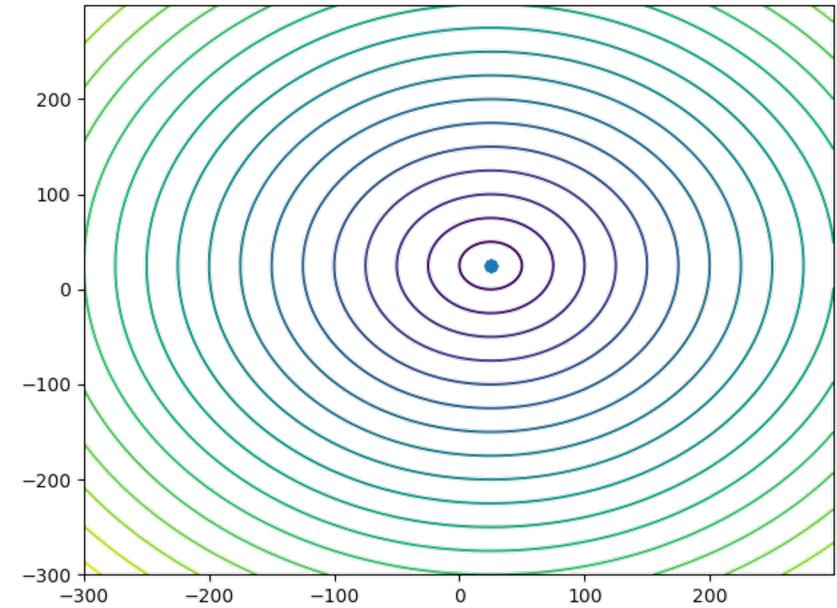


Spatial index

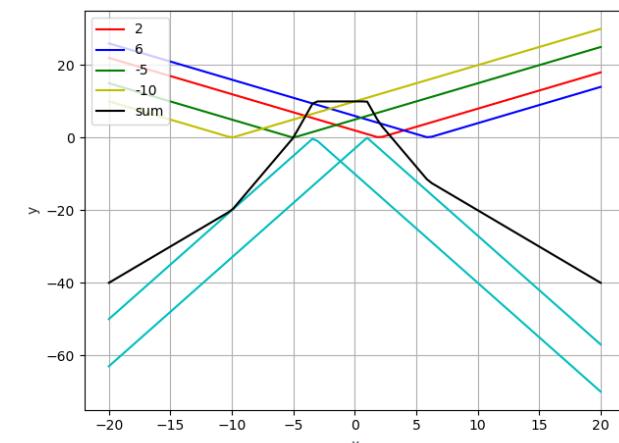
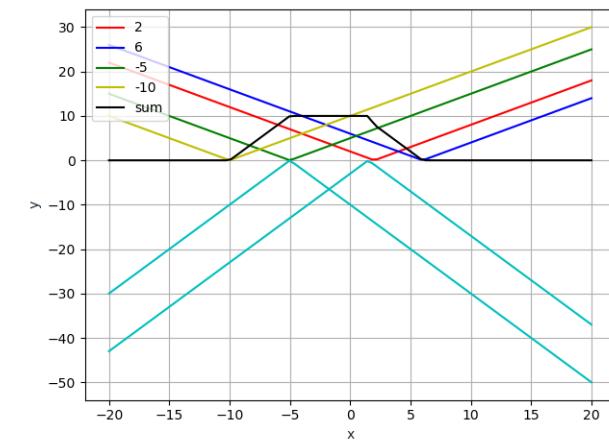
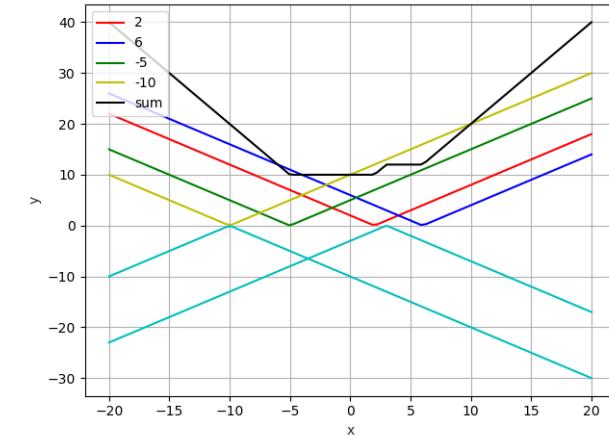
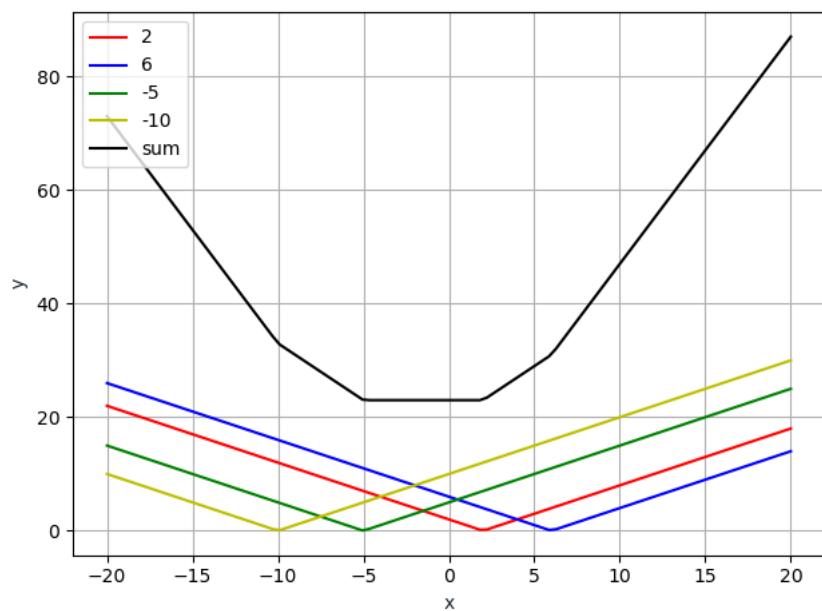


Observations

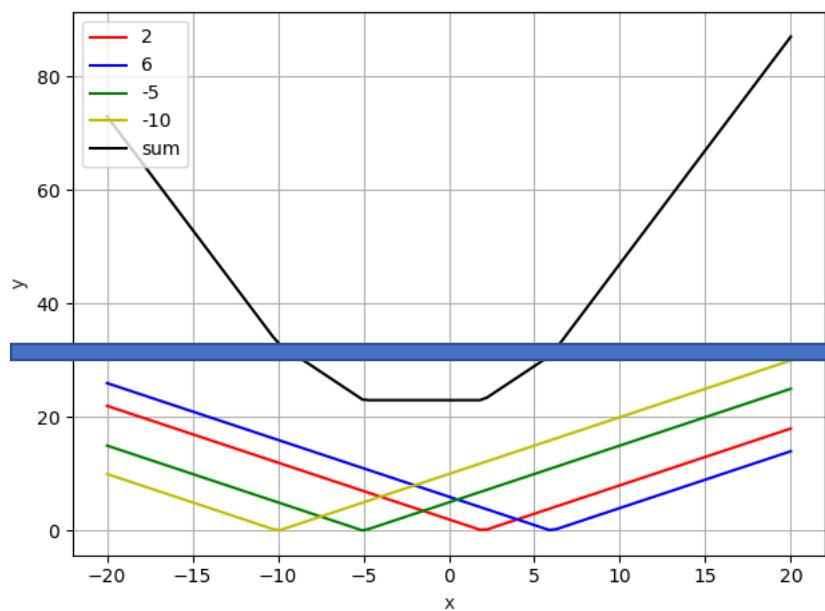
- Observation 1. 임의의 점의 집합 내부에서의 거리의 합이 최소가 되는 지점은 내부에 있다.
- Observation 2. 점 집합 외부에 타원모양이 형성되고 나면 그이후에는 닮음 꼴로(homomorphism) 등고선이 그려진다.
- Observation 3. 값은 달라지지만 등고선의 모양에 영향이 없게 점을 대체할 수 있다.



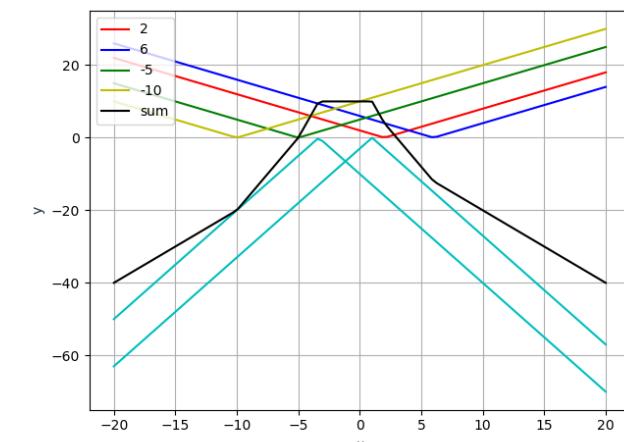
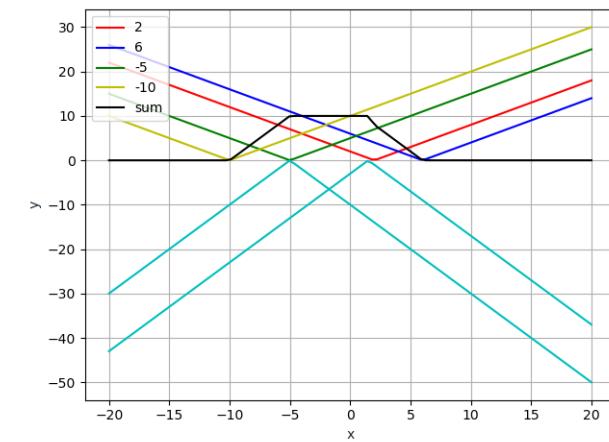
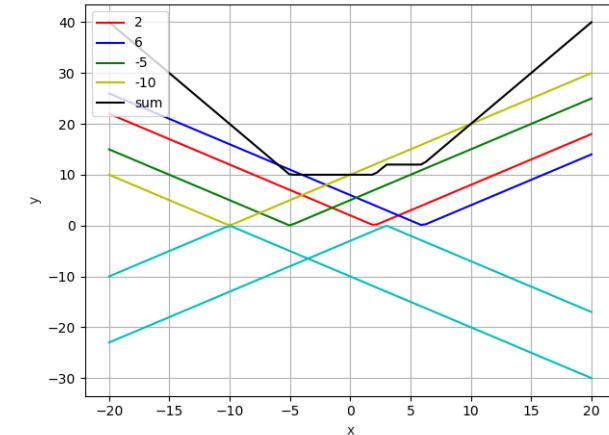
Weight boundary



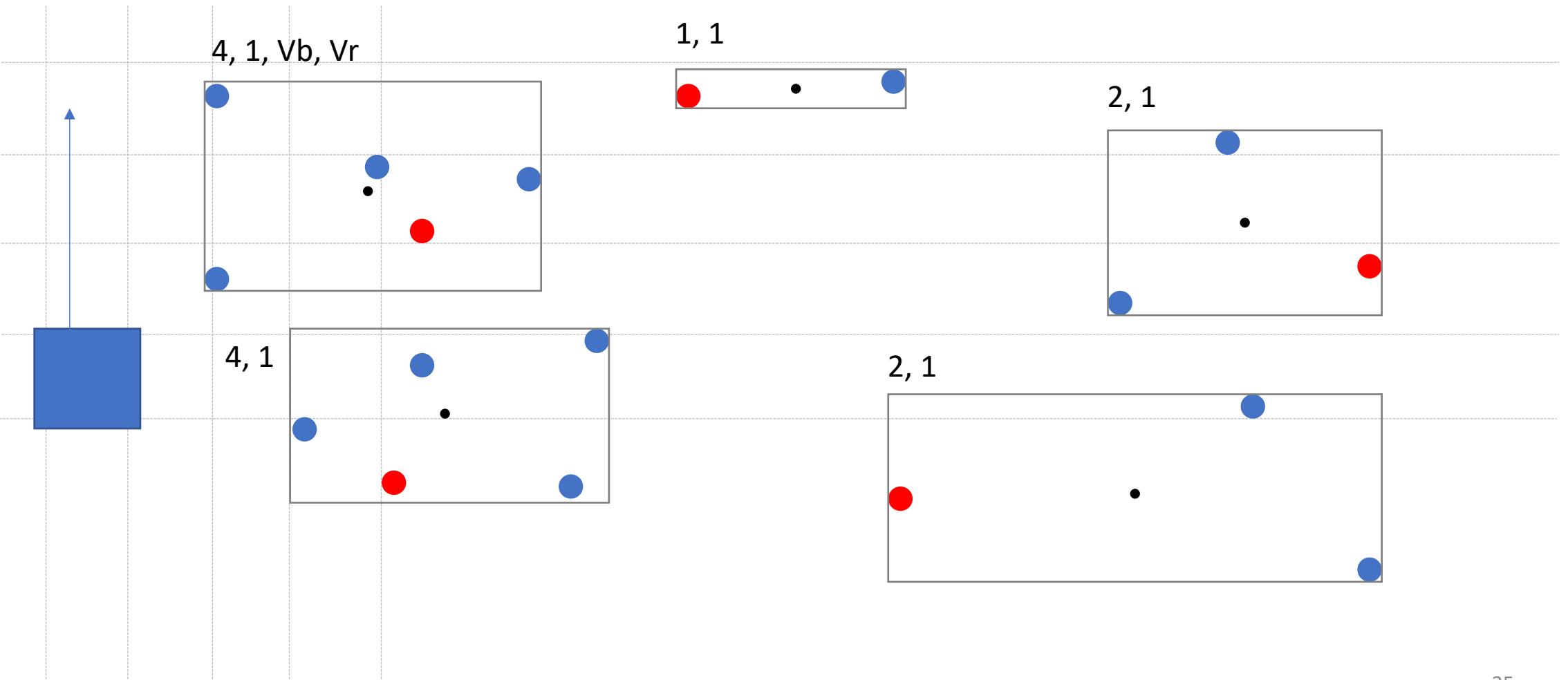
Weight boundary



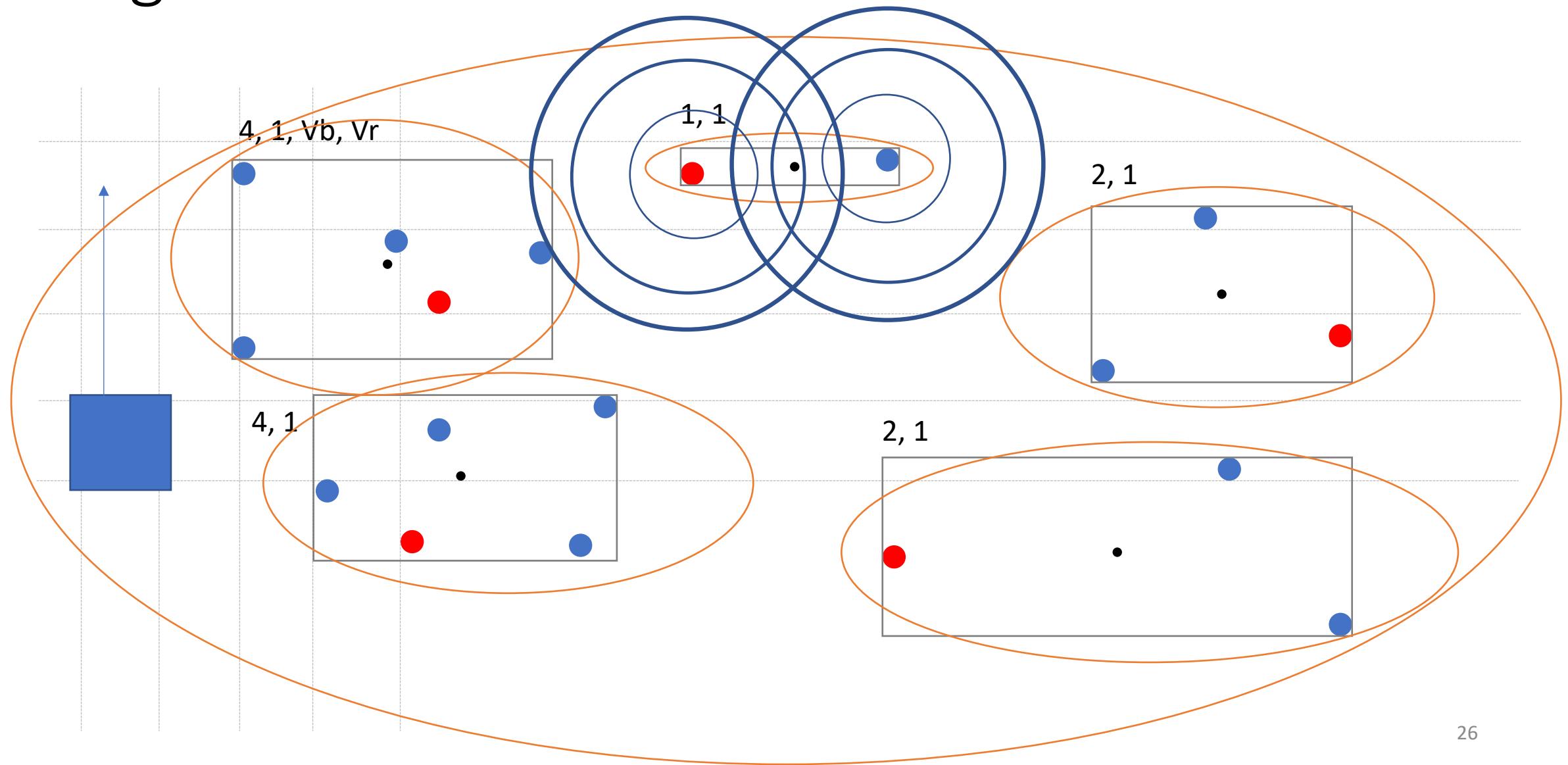
threshold



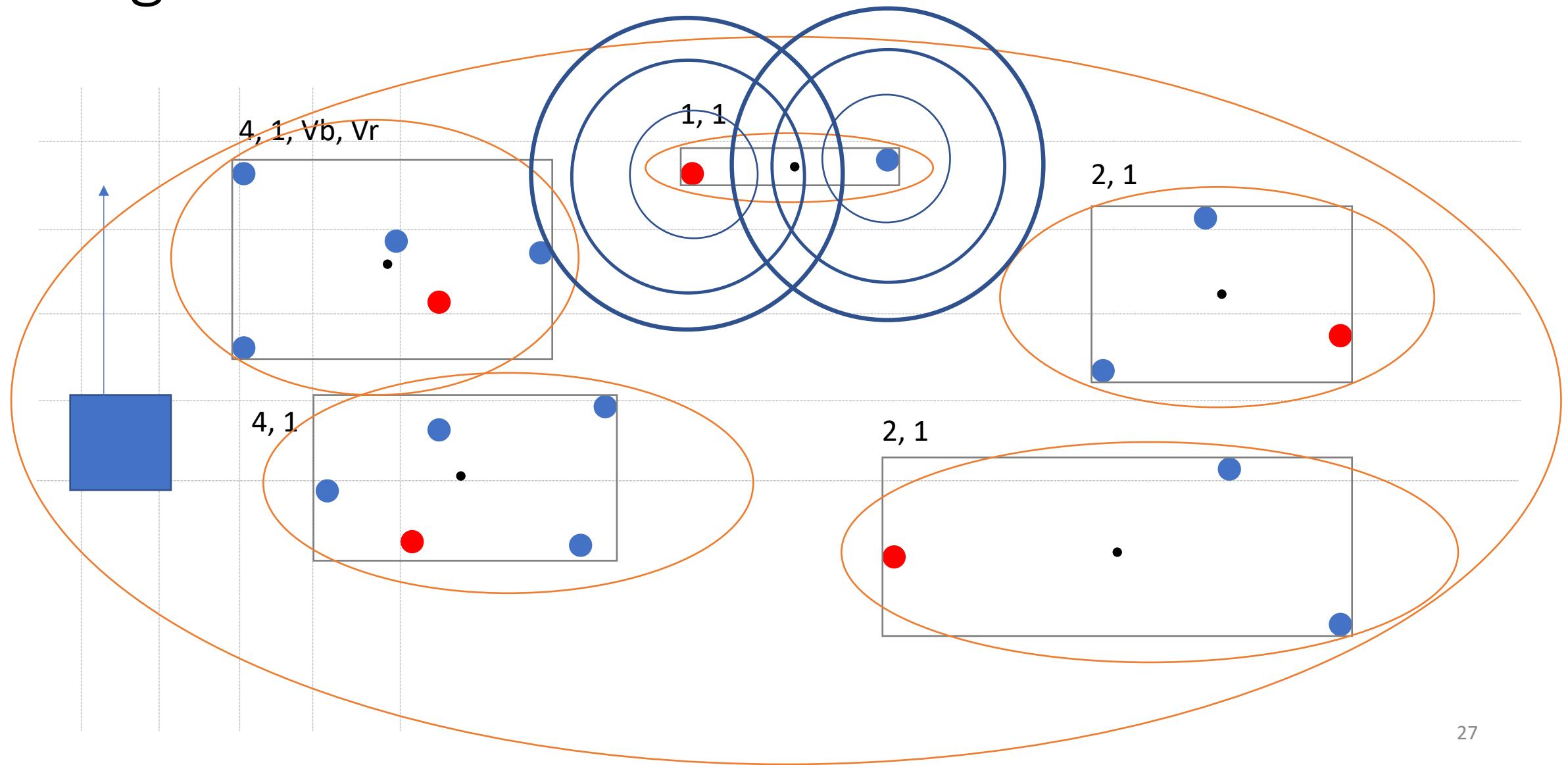
Regression



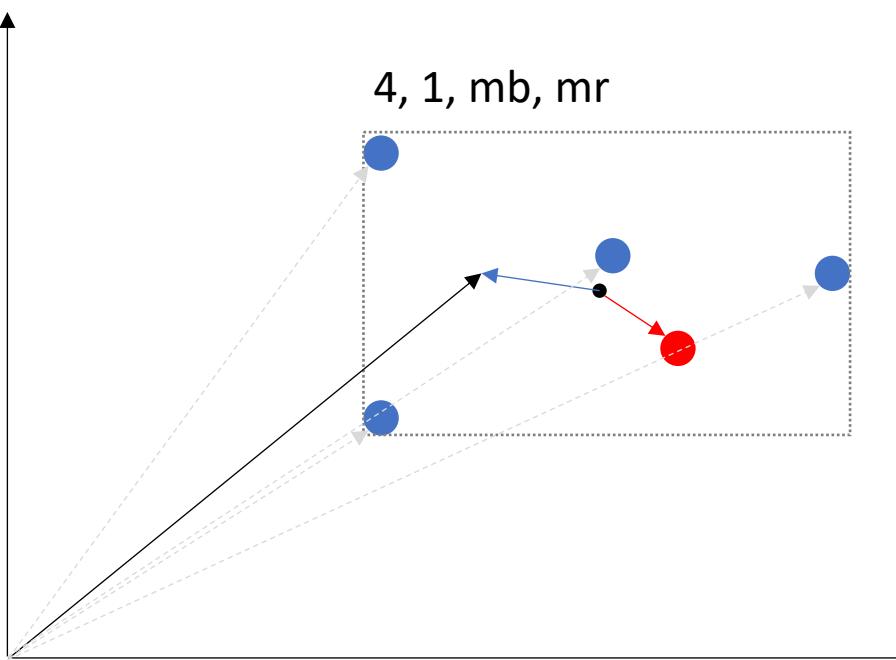
Regression



Regression



Spatial index



$$x_1 + x_2 + x_3 + x_4$$

Threshold ($2x_1x_2 + x_2x_3 + \dots$)

$$4 * \sqrt{\left(\frac{x_1 + x_2 + x_3 + x_4}{4}\right)^2}$$

Parallel processing

- Independent computation
 - Item categories
 - Region

Evaluation

- User Study
 - Analysis result comparison between exact visualization & approximate visualization
- Processing time
 - Exact visualization & approximate visualization

Implementation

- Synthetic dataset with categories (Zipf, Uniform, Normal)
- Multilevel grid, R-tree with optional parameters
- Accuracy estimation without exact value comparison
- Regression computation
- Redundant computation when weights vary

Contributions

- Quality preserving approximate visualization method
 - Modified spatial index
 - Weight boundary
 - Regression
- Metric for approximate visualization
- Parallel processing

ToDo

- Formulate & prove Lemma

Thanks