

INFS 4203 / 7203 Data Mining

Tutorial 4: Clustering

T4-Q1: Calculating distance

Calculate the L1, L2, L-infinite norms of two points: $(1, 0, 5)$ and $(2, 4, 9)$

T4-Q2: AGNES algorithm

Suppose the data mining task is to use agglomerative clustering to group measurements of the variable age.

- Age = {18, 28, 22, 33, 40, 48}

Commute the step-by-step agglomerative grouping using:

- a. Single linkage
- b. Complete linkage
- c. Average linkage

with **Manhattan distance**

T4-Q3: k -means and SSE calculation

Suppose the data mining task is to cluster the following measurements

Age = {18, 22, 25, 42, 27, 43, 33, 35, 56, 28} into **three** groups. For initial centroid of {18, 27, 35}:

- a. Use the k -means algorithm to show the clustering procedures step by step;
- b. Calculate the final SSE using **Manhattan distance**