HW1 solution

(3 points each question – total 24 points for Q1)

1(a). 2

1(b). Euclidean: 2; city block distance: 4

1(c). Euclidean: supremum: 3

1(d). Euclidean: supremum: 3 city block: 10

Q2: as long as they answer something reasonable (all give 10 points)

(Q3: 7.5 points each – 15 points total)

3: the sampling method in (a) is stratified random sampling, which enables to have a sample population that best represents the entire population in study. (b) may lead to a sample that can not fully represent the population.

(Q4: 16 points in total)

4(a). SMC = 7/10 (3points). Jaccard = 2/5 (3points).

4(b). Jaccard. Since we want to find the shared genes that is represented as 1. (3points).

4(c). Since two organisms of the same species share a large amount of the same genes, we would like to use SMC to find the differences among the genes. (7points).

(Q5: 12’ in total)

5(a) s({e}) = 8/10 (2’). s({b,d}) = 2/10 (2’). s({b,d,e}) = 2/10 (2’)

5(b) c({b,d} -> {e}) = 2/2 = 1. (2’) c({e}->{b,d}) = 2/8 (2’)

No, it’s not symmetric. (2’)

Q6 (total 3+10’)

6(a). 3^5 – 2^6 + 1 (1’)

6(b). given the graph as below, the maximum size of frequent itemsets is 4. (10’)



6(c). the itemsets {a,e} = 6 (1’)

6(d). {b}->{c} and {c}->{b} (1’)

Q7 (6+10’)

7(a) (10’)  
Diagram

Description automatically generated

7(b). 16/ 32 (including the null set) (3’)

7(c). 11/32 (3’)