Exam 1 Written

Benny Chen

February 22, 2023

Question 1

For each attribute given, classify its type as:

- discrete or continuous AND
- qualitative or quantitative AND
- nominal, ordinal, interval, or ratio

Indicate your reasoning if you think there may be some ambiguity in some cases.
Example: Age in years.
Answer: Discrete, quantitative, ratio.

(a) Daily user traffic volume at YouTube.com (i.e., number of daily visitors who visited the Web site).

(b) Air pressure of a car/bicycle tire (in psi).

(c) Credit card number.

(a) Answer: Discrete, quantitative, ratio.

(b) Answer: Continuous, quantitative, ratio.

(c) Answer: Discrete, qualitative, nominal.

Question 2

Consider the following set of frequent 2-itemsets: {p, q}, {p, r}, {p, s}, {p, t}, {q, r}, {q, t}, {r, s}, {s, t}

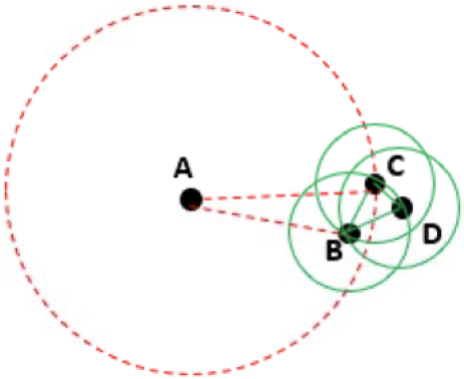
(a) List all the candidate 3-itemsets produced during the candidate generation step of the Apriori algorithm.

(b) List all the candidate 3-itemsets that survive the pruning step of the Apriori algorithm.

(a) Answer: {p, q, r}, {p, q, s}, {p, q, t}, {p, r, s}, {p, s, t}, {p, r, t}, {q, r, t}, {q, r, s}, {q, s, t}, {r, s, t}

(b) Answer: {p, q, r}, {p, q, s}, {p, q, t}, {p, r, s}, {p, s, t}, {p, r, t}, {q, r, t}, {r, s, t}
Removed: {q, r, s}, {q, s, t}

Question 3



This question aims at finding the local outlier factor (LOF) for the data points (a) A and (b) C from above figure. Suppose $k = 2$. We know that: B, C are the two nearest neighbors to A; B and D are the two nearest neighbors to C. We also know the given distances: $d(A, B) = 4$, $d(A, C) = 5$, $d(B, C) = 1.5$, $d(C, D) = 1$, $d(B, D) = 1.2$.

(a) LOF (A)

(b) LOF (C)

A $k=2$

B and C closest

1 d(AB) = 4

2 d(AC) = 5

k Nearest neighbor		average RD	LRD
A 5 = AC	set AB AC	$d_2(A) = 5$	$\frac{1}{5}$
B 1.5 = BC	set BD BC	$d_2(B) = 1.5$	$\frac{1}{1.5}$
C 1.5 = BC	set CD BC	$d_2(C) = 1.5$	$\frac{1}{1.5}$
D 1.2 = BD	set CD BD	$d_2(D) = 1.2$	$\frac{1}{1.2}$

RD AB max (1.5, 4)

RD AC max (1.5, 5)

RD BD max (1.2, 1.2)

RD BC max (1.5, 1.5)

RD CD max (1, 1.2)

RD BC max (1.5, 1.5)

RD CD max (1, 1.5)

RD BD max (1.2, 1.5)

LOF: $\frac{\frac{lrd_2B}{lrd_2A} + \frac{lrd_2C}{lrd_2A}}{2}$

$= \frac{\frac{1.5}{\frac{1}{5}} + \frac{1.5}{\frac{1}{5}}}{2} =$

3.333333

C k=2

B and D closest

1 d(CD) = 1.2

2 d(BC) = 1.5

k Nearest neighbor

A 5 = AC set AB AC

B 1.5 = BC set BD BC

C 1.5 = BC set CD BC

D 1.2 = BD set CD BD

RD AB max(1.5, 4)

RD AC max(1.5, 5)

RD BD max(1.2, 1.2)

RD BC max(1.5, 1.5)

RD CD max(1, 1.2)

RD BC max(1.5, 1.5)

RD CD max(1, 1.5)

RD BD max(1.2, 1.5)

	average RD	LRD
A	$d_2(A) = 5$	$\frac{1}{5}$
B	$d_2(B) = 1.5$	$\frac{1}{1.5}$
C	$d_2(C) = 1.5$	$\frac{1}{1.5}$
D	$d_2(D) = 1.2$	$\frac{1}{1.2}$

$$\frac{\frac{1}{1.2}}{\frac{1}{1.5}} + \frac{\frac{1}{1.5}}{\frac{1}{1.5}} = \frac{}{2}$$

1.125