Knowledge Discovery and Data Mining (CS 513)

(Midterm)

Prof. Khasha Dehnad

Student Name : **Paras Garg**

Course Section : **CS 513-A**

CWID : **10414982**

Question 1 – For the experiment consisting of a single die toss, let **A = {outcome is <= 3}, B = {1, 2, 5, 6}** and **C = {outcome is odd}.** Please answer each of the following three **True / False** question. Show your work.

1. $(A \cup B)' = A' \cap B'$
2. $(A \cap C)' = A' \cup C'$
3. $(A \cap C)' - B = A' \cup (C - B)$

Solution 1 – Since given,

$$\mathbf{A = \{1, 2, 3\}}$$
$$\mathbf{B = \{1, 2, 5, 6\}}$$
$$\mathbf{C = \{1, 3, 5\}}$$

1. $(A \cup B)' = A' \cap B'$

$$A = \{ 1, 2, 3 \}$$
$$B = \{ 1, 2, 5, 6 \}$$

$$\boldsymbol{A \cup B} = \{1, 2, 3\} \cup \{1, 2, 5, 6\} = \{ \mathbf{1, 2, 3, 5, 6} \}$$
$$(\boldsymbol{A \cup B})' = \{ \mathbf{4} \}$$

$$A = \{ 1, 2, 3 \}, \quad B = \{ 1, 2, 5, 6 \}$$
$$A' = \{ 4, 5, 6 \}, \quad B' = \{ 3, 4 \}$$

$$\boldsymbol{A' \cap B'} = \{\mathbf{4, 5, 6}\} \cap \{\mathbf{3, 4}\} = \{ \mathbf{4} \}$$
$$\boldsymbol{A' \cap B'} = \{ \mathbf{4} \}$$

$$Hence, (\boldsymbol{A \cup B})' = \boldsymbol{A' \cap B'} => \boldsymbol{True}$$

2. $(A \cap C)' = A' \cup C'$

$$A = \{ 1, 2, 3 \}$$
$$C = \{ 1, 3, 5 \}$$

$$\boldsymbol{A \cap C} = \{1, 2, 3\} \cap \{1, 3, 5\} = \{ \mathbf{1, 3} \}$$
$$(\boldsymbol{A \cap C})' = \{ \mathbf{2, 4, 5, 6} \}$$

$$A = \{ 1, 2, 3 \}, \quad C = \{ 1, 3, 5 \}$$
$$A' = \{ 4, 5, 6 \}, \quad C' = \{ 2, 4, 6 \}$$

$$\boldsymbol{A' \cup C'} = \{4, 5, 6\} \cup \{2, 4, 6\} = \{ \mathbf{2, 4, 5, 6} \}$$

$$Hence, (\boldsymbol{A \cap C})' = \boldsymbol{A' \cup C'} => \boldsymbol{True}$$

3. $(A \cap C)' - B = A' \cup (C - B)$

$$A = \{ 1, 2, 3 \}$$
$$B = \{ 1, 2, 5, 6 \}$$
$$C = \{ 1, 3, 5 \}$$

$$A \cap C = \{1, 2, 3\} \cap \{1, 3, 5\} = \{ 1, 3 \}$$
$$(A \cap C)' = \{ 2, 4, 5, 6 \}$$
$$(\boldsymbol{A \cap C})' - \boldsymbol{B} = \{2, 4, 5, 6\} - \{1, 2, 5, 6\} = \{ \mathbf{4} \}$$

$$A = \{ 1, 2, 3 \}$$
$$B = \{ 1, 2, 5, 6 \}$$
$$C = \{ 1, 3, 5 \}$$

$$A' = \{ 4, 5, 6 \}$$
$$(C - B) = \{1, 3, 5\} - \{1, 2, 5, 6\} = \{ 3 \}$$
$$\boldsymbol{A' \cup (C - B)} = \{4, 5, 6\} \cup \{3\} = \{ \mathbf{3, 4, 5, 6} \}$$

$$Hence, (\boldsymbol{A \cap C})' - \boldsymbol{B} \neq \boldsymbol{A' \cup (C - B)} => \boldsymbol{False}$$

Question 2 – Is the following function a proper distance function? Why? Explain your answer.

$$\boldsymbol{d(x,y) = (\sum_i (x_i - y_i)^2)^{3/2}}$$

Hint: Measure the distance between $(0,0), (0,1) and\ (1,1)$.

Solution 2 – Any distance function can only be a proper distance function if it follows the following properties:-

1. $d(x,y) \geq 0\ \&\ d(x,y) = 0, if\ and\ only\ if\ x = y$
2. $d(x,y) = d(y,x)$
3. $d(x,z) \leq d(x,y) + d(y,z)$

Now, let us consider the three points in a coordinate system as $x(0,0),\ y(0,1)\ and\ z(1,1)$.

Case 1: Calculating distance between $\boldsymbol{x\ and\ y}$ using above function

$$\boldsymbol{d(x,y) = (\sum_i (x_i - y_i)^2)^{3/2} = ((0-0)^2 + (0-1)^2)^{3/2} = ((0)^2 + (1)^2)^{3/2} = (0+1)^{3/2} = (1)^{3/2} = 1}$$

$$\boldsymbol{d(x,y) = 1}$$

Case 2: Calculating distance between $\boldsymbol{x\ and\ z}$ using above function

$$\boldsymbol{d(x,z) = (\sum_i (x_i - z_i)^2)^{3/2} = ((0-1)^2 + (0-1)^2)^{3/2} = ((1)^2 + (1)^2)^{3/2} = (1+1)^{3/2} = (2)^{3/2} = 2\sqrt{2}}$$

$$\boldsymbol{d(x,z) = 2\sqrt{2}}$$

Case 3: Calculating distance between $\boldsymbol{y\ and\ x}$ using above function

$$\boldsymbol{d(y,x) = (\sum_i (y_i - x_i)^2)^{3/2} = ((0-0)^2 + (1-0)^2)^{3/2} = ((0)^2 + (1)^2)^{3/2} = (0+1)^{3/2} = (1)^{3/2} = 1}$$

$$\boldsymbol{d(y,x) = 1}$$

Case 4: Calculating distance between $\boldsymbol{y\ and\ z}$ using above function

$$\boldsymbol{d(y,z) = (\sum_i (y_i - z_i)^2)^{3/2} = ((0-1)^2 + (1-1)^2)^{3/2} = ((1)^2 + (0)^2)^{3/2} = (1+0)^{3/2} = (1)^{3/2} = 1}$$

$$\boldsymbol{d(y,z) = 1}$$

Case 5: Calculating distance between $\boldsymbol{z\ and\ x}$ using above function

$$\boldsymbol{d(z,x) = (\sum_i (z_i - x_i)^2)^{3/2} = ((1-0)^2 + (1-0)^2)^{3/2} = ((1)^2 + (1)^2)^{3/2} = (1+1)^{3/2} = (2)^{3/2} = 2\sqrt{2}}$$

$$\boldsymbol{d(z,x) = 2\sqrt{2}}$$

Case 6: Calculating distance between $\boldsymbol{z\ and\ y}$ using above function

$$\boldsymbol{d(z,y) = (\sum_i (z_i - y_i)^2)^{3/2} = ((1-0)^2 + (1-1)^2)^{3/2} = ((1)^2 + (0)^2)^{3/2} = (1+0)^{3/2} = (1)^{3/2} = 1}$$

$$\boldsymbol{d(z,y) = 1}$$

Checking the validity of the distance function properties on the distance values calculated by the given function.

Property 1: $d(x,y) \geq 0\ \&\ d(x,y) = 0, if\ and\ only\ if\ x = y$

$$d(x,y) \geq 0, d(x,z) \geq 0, d(y,x) \geq 0, d(y,z) \geq 0, d(z,x) \geq 0\ and\ d(z,y) \geq 0$$

For all cases property 1 satisfied

Property 2: $d(x,y) = d(y,x)$

$$d(x,y) = d(y,x) = 1, \quad \text{from case 1 and case 3}$$
$$d(y,z) = d(z,y) = 1, \quad \text{from case 4 and case 6}$$
$$d(x,z) = d(z,x) = 2\sqrt{2}, \quad \text{from case 2 and case 5}$$

For all cases property 2 satisfied

Property 3: $d(x,z) \leq d(x,y) + d(y,z)$

From case 1, case 2 and case 4

$$d(x,z) \leq d(x,y) + d(y,z)$$
$$2\sqrt{2} \leq 1 + 1$$
$$2\sqrt{2} \leq 2$$

From case 3, case 5 and case 6

$$d(z,x) \leq d(z,y) + d(y,x)$$
$$2\sqrt{2} \leq 1 + 1$$
$$2\sqrt{2} \leq 2$$

$2\sqrt{2} \neq 2 \; and \; 2\sqrt{2} > 2,$
Condition failed.
For both cases property 3 failed

As per the above calculations and observations, the cases satisfies the property 1 and the property 2 of distance function but do not satisfy the property 3. Therefore, **the given function is not a proper distance function.**

Question 4 – A telecommunications company is concerned about the number of customers leaving their business (Churn = True). Using past data, an analyst has prepared the table below. Using the table below, calculate the following probabilities.

	International Plan	Voice Plan	Churn FALSE	Churn TRUE	Row Total
	no	no	1,878	302	2,180
	no	yes	786	44	830
Sub-Total			2,664	346	3,010
	yes	no	130	101	231
	yes	yes	56	36	92
Sub-Total			186	137	323
GrandTotal			2,850	483	3,333

1. $P\,(Churn = True)$

$$\boldsymbol{P\,(Churn = True)} = \frac{302 + 44 + 101 + 36}{3333} = \frac{483}{3333} = \mathbf{14.49\%}$$

2. $P\,(Churn = False)$

$$\boldsymbol{P\,(Churn = False)} = \frac{1878 + 786 + 130 + 56}{3333} = \frac{2850}{3333} = \mathbf{85.51\%}$$

3. $P\,(International\ Plan = Yes)$

$$\boldsymbol{P(\ International\ Plan = Yes)} = \frac{130 + 101 + 56 + 36}{3333} = \frac{231 + 92}{3333} = \frac{323}{3333} = \mathbf{9.69\%}$$

4. $P\,(Voice\ Plan = Yes)$

$$\boldsymbol{P(\ Voice\ Plan = Yes)} = \frac{786 + 44 + 56 + 36}{3333} = \frac{830 + 92}{3333} = \frac{922}{3333} = \mathbf{27.66\%}$$

5. $P\,(Internatial\ Plan = Yes,\ Voice\ Plan = Yes)$

$$\boldsymbol{P(International\ Plan = Yes\ \cap\ Voice\ Plan = Yes)} = \frac{56 + 36}{3333} = \frac{92}{3333} = \mathbf{2.76\%}$$

6. Are "Voice Plan" and "International Plan" independent?

   According to independent probability =>

   $$\boldsymbol{P(International\ Plan = Yes \cap\ Voice\ Plan = Yes) = P(International\ Plan = Yes) * P(Voice\ Pkan = Yes)}$$

   $$\boldsymbol{P(International\ Plan = Yes \cap\ Voice\ Plan = Yes)} = \frac{92}{3333} = \mathbf{2.76\%},$$

   $$\boldsymbol{P(International\ Plan = Yes) * P(Voice\ Pkan = Yes)} = \frac{323}{3333} * \frac{922}{3333} = \mathbf{2.68\%}$$

   The values on left side and right side is not equal implies condition for independent probability failed, so we can consider it as the "voice plan" and "international plan" **are not independent or are dependent**.

7. $P\ (Internatial\ Plan = Yes,\ Voice\ Plan = Yes \mid Chrun = True)$

   $$\frac{\boldsymbol{P((Internatial\ Plan = Yes\ \cap\ Voice\ Plan = Yes) \cap (Churn = True))}}{\boldsymbol{P(Churn = True)}} = \frac{\frac{36}{3333}}{\frac{483}{3333}} = \frac{36}{483} = \mathbf{7.45\%}$$

8. $P\ (Internatial\ Plan = Yes,\ Voice\ Plan = Yes \mid Chrun = False)$

   $$\frac{\boldsymbol{P((Internatial\ Plan = Yes\ \cap\ Voice\ Plan = Yes) \cap (Churn = False))}}{\boldsymbol{P(Churn = False)}} = \frac{\frac{56}{3333}}{\frac{2850}{3333}} = \frac{56}{2850} = \mathbf{1.96\%}$$

9. $P\ (Chrun = False \mid Internatial\ Plan = Yes,\ Voice\ Plan = Yes)$

   $$\frac{\boldsymbol{P((Churn = False) \cap (Internatial\ Plan = Yes \cap\ Voice\ Plan = Yes))}}{\boldsymbol{P(Internatial\ Plan = Yes \cap\ Voice\ Plan = Yes)}} = \frac{\frac{56}{3333}}{\frac{92}{3333}} = \frac{56}{92} = \mathbf{60.87\%}$$