Knowledge Discovery and Data Mining (CS 513)

(Assignment 1 – Intro Probability)

Prof. Khasha Dehnad

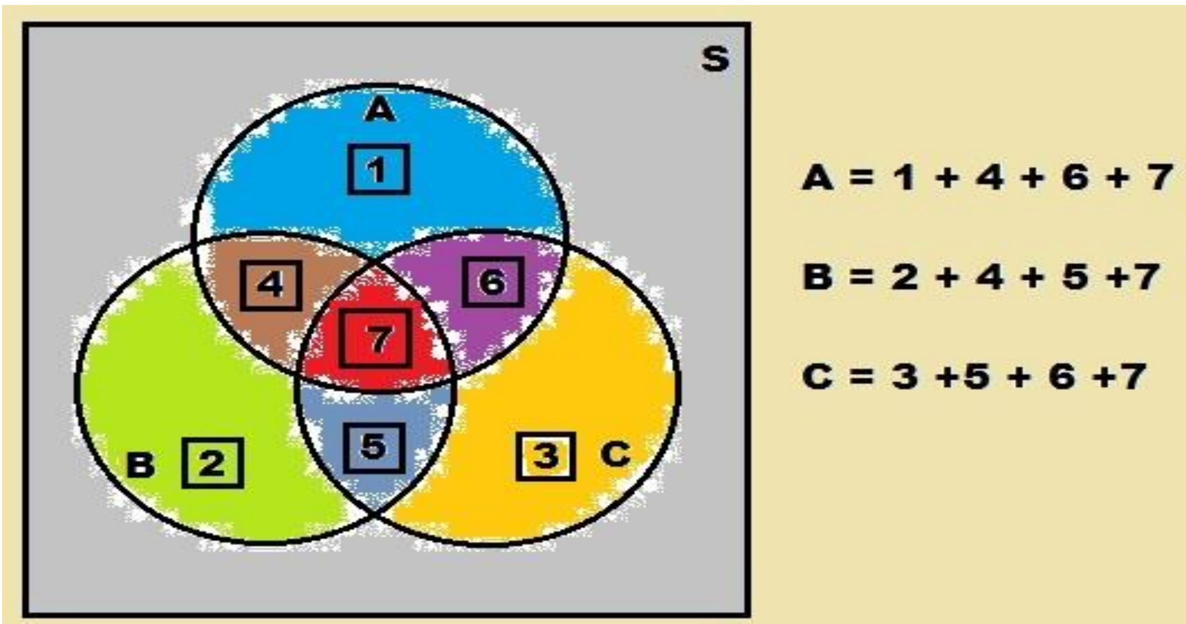
Student Name : **Paras Garg**

Course Section : **CS 513-A**

Homework 1.1 – True or False

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

Solution 1.1 –



(Venn Diagram for Solution 1.1)

1. $A - B = A \cap B'$

$$A = 1 + 4 + 6 + 7$$
$$B = 2 + 4 + 5 + 7$$

$$A - B = 1 + 6$$

$$A = 1 + 4 + 6 + 7$$
$$B = 2 + 4 + 5 + 7$$
$$B' = 1 + 3 + 6$$

$$A \cap B = 1 + 6$$

$$Hence, \boldsymbol{A - B = A \cap B'} => True$$

2. $(A \cup B) - C = A \cup (B - C)$

$$A = 1 + 4 + 6 + 7$$
$$B = 2 + 4 + 5 + 7$$
$$C = 3 + 5 + 6 + 7$$

$$A \cup B = 1 + 2 + 4 + 5 + 6 + 7$$
$$(A \cup B) - C = 1 + 2 + 4$$

$$A = 1 + 4 + 6 + 7$$
$$B = 2 + 4 + 5 + 7$$
$$C = 3 + 5 + 6 + 7$$

$$B - C = 2 + 4$$
$$A \cup (B - C) = 1 + 2 + 6 + 7$$

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

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

$$A = 1 + 4 + 6 + 7$$
$$B = 2 + 4 + 5 + 7$$
$$C = 3 + 5 + 6 + 7$$

$$A \cup B = 1 + 2 + 4 + 5 + 6 + 7$$
$$(A \cup B) \cap C = 5 + 6 + 7$$

$$A = 1 + 4 + 6 + 7$$
$$B = 2 + 4 + 5 + 7$$
$$C = 3 + 5 + 6 + 7$$

$$B \cap C = 5 + 7$$
$$A \cup (B \cap C) = 1 + 4 + 5 + 6 + 7$$

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

4. $A - (A' \cap B) = A$

$$A = 1 + 4 + 6 + 7$$
$$A' = 2 + 3 + 5$$
$$B = 2 + 4 + 5 + 7$$

$$A' \cap B = 2 + 5$$
$$A - (A' \cap B) = 1 + 4 + 6 + 7$$

$$A = 1 + 4 + 6 + 7$$

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

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

$$A = 1 + 4 + 6 + 7$$
$$B = 2 + 4 + 5 + 7$$
$$B' = 1 + 3 + 6$$
$$C = 3 + 5 + 6 + 7$$

$$A \cap B' = 6$$
$$A \cap B' \cap C = 6$$

$$A = 1 + 4 + 6 + 7$$
$$B = 2 + 4 + 5 + 7$$
$$C = 3 + 5 + 6 + 7$$

$$A - B = 1 + 6$$
$$(A - B) \cap C = 6$$

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

Homework 1.2 – Juan is playing the following game: he rolls two dice. If they sum up to 7 he loses a dollar. If they sum up to 2, he wins 2 dollars. Otherwise, he doesn’t win or lose. After playing this game for a long time, what shall happen? Why?

Solution 1.2 –

Probability (Sum = 2) = 1 / 36 = 0.0278
Probability (Sum = 7) = 6 / 36 = 0.1667
Probability (Sum = other) = 29 / 36 = 0.8056

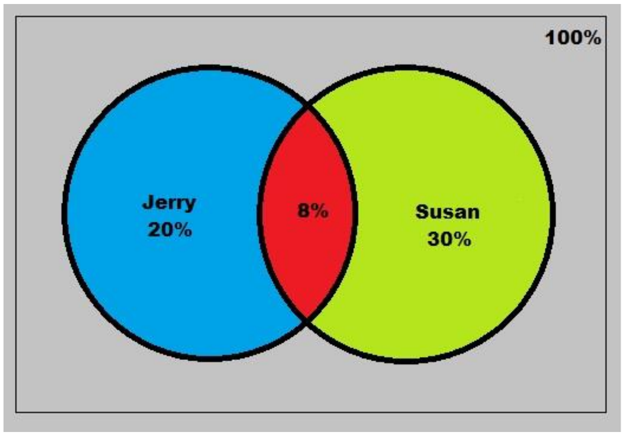
In a long run winning/losing status = P(2) * 2 - P(7) * 1 + P(O) * 0
= 0.0278 * 2 - 0.1667 * 1 + 0.8056 * 0
= 0.0556 – 0.1667
= - 0.1111
= - 0.11 (approx.)

After playing this game for a long time, Jaun will **lose 11 Cents** as an average on per roll.

Homework 1.3 – Jerry and Susan have a joint bank account. Jerry goes to the bank 20% of the days. Susan goes there 30% of the days. Together they are at the bank 8% of the days.

a. Susan was at the bank last Monday. What's the probability that Jerry was there too?
b. Last Friday, Susan wasn't at the bank. What's the probability that Jerry was there?
c. Last Wednesday at least one of them was at the bank. What is the probability that both of them were there?

Solution 1.3 –



(Venn Diagram of Jerry and Susan)

	Susan at Bank	Susan not at Bank
Jerry at Bank	8%	12%
Jerry not at Bank	22%	58%

$$\begin{aligned}
&\boldsymbol{P\ (Jerry \cap Susan) = 8\%}\\
&\boldsymbol{P\ (Jerry - Susan) = 20\% - 8\% = 12\%}\\
&\boldsymbol{P\ (Susan - Jerry) = 30\% - 8\% = 22\%}\\
&\boldsymbol{P\ (Jerry' \cup Susan'^{)} = 100\% - (Jerry \cup Susan)}\\
&\qquad\qquad\qquad\boldsymbol{= 100\% - \{P(J) + P(S) - P(J \cap S)\}}\\
&\qquad\qquad\qquad\boldsymbol{= 100\% - \{20\% + 30\% - 8\%\}}\\
&\qquad\qquad\qquad\boldsymbol{= 100\% - 42\%}\\
&\qquad\qquad\qquad\boldsymbol{= 58\%}
\end{aligned}$$

a. $\boldsymbol{P\ (Jerry \mid Susan)} = \frac{P(Jerry \cap Susan)}{P(Susan)} = \frac{8\%}{30\%} = \boldsymbol{26.66\%}$

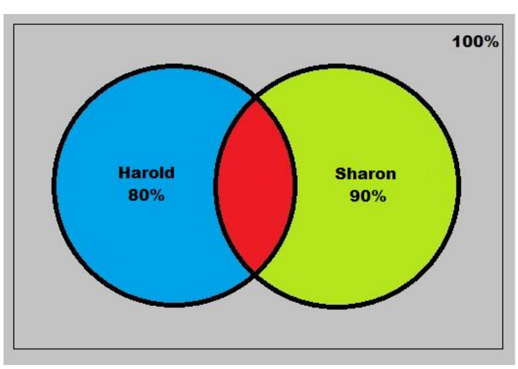
b. $\boldsymbol{P\ (Jerry \mid Susan')} = \frac{P(Jerry \cap\ Susan\prime)}{P(Susan')} = \frac{12\%}{70\%} = \boldsymbol{17.14\%}$

c. $\boldsymbol{P\ (Jerry \cap Susan \mid Jerry \cup Susan)} = \frac{P((Jerry \cap Susan) \cap (Jerry \cap Susan))}{P(Jerry \cup Susan)} = \frac{P(Jerry \cap Susan)}{P(Jerry \cup Susan)} = \frac{8\%}{42\%} = \boldsymbol{19.04\%}$

Homework 1.4 – Harold and Mary are studying for a test. Harold's chances of getting "B" are 80%. Sharon's chances of getting a "B" are 90%. The probability of at least one of them getting a "B" is 91%.

a. What is the probability that only Harold gets a "B""?
b. What is the probability that only Sharon gets a "B"?
c. What is the probability that both won't get a "B"?

Solution 1.4 –



$$\begin{aligned}
&\boldsymbol{P(Harold \cup Sharon) = 91\%}\\
&\boldsymbol{\big(P\ (Harold \cup Sharon)\big) = P\ (H) +\ P(S) -\ P(H \cap S)}\\
&\boldsymbol{\big(P\ (Harold \cap Sharon)\big) = P\ (H) +\ P(S) -\ P(H \cup S)}\\
&\qquad\qquad\qquad\boldsymbol{= 80\% + 90\% - 91\%}\\
&\qquad\qquad\qquad\boldsymbol{= 1\%}
\end{aligned}$$

a. $\boldsymbol{P(Harold)} = P(Harold) - P(Harold \cap Sharon) = 80\% - 79\% = \mathbf{1}\%$

b. $\boldsymbol{P(Sharon)} = P(Sharon) - P(Sharon \cap Harold) = 90\% - 79\% = \mathbf{11}\%$

c. $\boldsymbol{P(Harold' \cup Sharon')} = 100\% - \ P(Harold \cup Sharon) = 100\% - 1\% = \mathbf{9}\%$

Homework 1.5 – Jerry and Susan have a joint bank account. Jerry goes to the bank 20% of the days. Susan goes there 30% of the days. Together they are at the bank 8% of the days. Are the events “Jerry is at the bank” and “Susan is at the bank” independent?

Solution 1.5 – The events “Jerry is at the bank” and “Susan is at the bank” are **not independent** as the probability of occurrence of another event.

Given:

$$\boldsymbol{P\ (Jerry \cap Susan) = 8}\%$$

According to Independent Probability:

$$\begin{aligned}\boldsymbol{P\ (Jerry \cap Susan)} &= \boldsymbol{P(Jerry) * P(Susan)} \\ &= 20\% * 30\% \\ &= \mathbf{6}\%\end{aligned}$$

Homework 1.6 – You roll 2 dice.

a. Are the events “the sum is 6” and “the second die shows 5” independent?
b. Are the events “the sum is 7” and “the first die shows 5” independent?

Solution 1.6 –

a.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The Sum is 6

Second die shows 5

$$\boldsymbol{P\ (Sum\ is\ 6)} = \frac{\mathbf{5}}{\mathbf{36}}$$

$$\boldsymbol{P\ (Second\ die\ shows\ 5)} = \frac{6}{36} = \frac{\mathbf{1}}{\mathbf{6}}$$

According to Joint Probability =>

$$\boldsymbol{P\ (Sum\ is\ 6 \cap Second\ die\ shows\ 5)} = \frac{\mathbf{1}}{\mathbf{36}}$$

According to Independent Probability =>

$$\boldsymbol{P\ (Sum\ is\ 6 \cap Second\ die\ shows\ 5)} = P(Sum\ is\ 6) * P(Second\ die\ shows\ 5) = \frac{\mathbf{1}}{\mathbf{6}} * \frac{\mathbf{5}}{\mathbf{36}}$$

The above equations are not equal, implies, **these two events are not independent.**

b.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Sum is 7

First die shows 5

$$\boldsymbol{P}\ (\boldsymbol{Sum\ is\ 7}) = \frac{6}{36} = \frac{\mathbf{1}}{\mathbf{6}}$$

$$\boldsymbol{P}\ (\boldsymbol{First\ die\ shows\ 5}) = \frac{6}{36} = \frac{\mathbf{1}}{\mathbf{6}}$$

According to Joint Probability =>

$$\boldsymbol{P}\ (\boldsymbol{Sum\ is\ 7} \cap \boldsymbol{First\ die\ shows\ 5}) = \frac{\mathbf{1}}{\mathbf{36}}$$

According to Independent Probability =>

$$\boldsymbol{P}\ (\boldsymbol{Sum\ is\ 7} \cap \boldsymbol{First\ die\ shows\ 5}) = P(Sum\ is\ 7) * P(First\ die\ shows\ 5) = \frac{1}{6} * \frac{1}{6} = \frac{\mathbf{1}}{\mathbf{36}}$$

The above equations are equal, implies, **these two events are independent.**

Homework 1.7 – An oil company is considering drilling in either TX, AK and NJ. The company may operate in only one state. There is a 60 % chance the company will choose TX and 10% chance of NJ. There is a 30 % chance of finding oil in TX, 20% in AK, and 10% in NJ.

a. What's the probability of finding oil?
b. The company decides to drill and found oil. What is the probability that they drilled in TX?

Solution 1.7 –

	TX	**AK**	**NJ**	
Oil	**18%**	**6%**	**1%**	**25%**
No Oil	**42%**	**24%**	**9%**	**75%**
	60%	**30%**	**10%**	**100%**

The probability of finding oil in TX ($P\ (Oil\mid TX) = 30\%$) and probability of choosing TX ($P(TX) = 60\%$),

$$\boldsymbol{P}\ (\boldsymbol{Oil} \mid \boldsymbol{TX}) = \frac{P(Oil \cap TX)}{P(TX)},$$

$$\boldsymbol{P}\ (\boldsymbol{Oil} \cap \boldsymbol{TX}) = P(Oil \mid TX) * P(TX) = 30\% * 60\% = \mathbf{18}\%,$$

Similarly for$P\ (Oil \mid NJ) = 10\%$),

$$\boldsymbol{P\ (Oil \cap NJ)} = P(Oil \mid NJ) * P(NJ) = 10\% * 10\% = \mathbf{1\%},$$

Similarly for$P\ (Oil \mid AK) = 100\% - (60\% + 10\%)) = 30\%$,

$$\boldsymbol{P\ (Oil \cap AK)} = P(Oil \mid AK) * P(AK) = 30\% * 20\% = \mathbf{6}\%$$

a. $\boldsymbol{P\ (Oil)} = 18\% + 6\% + 1\% = \mathbf{25}\%$

b. $\boldsymbol{P\ (TX \mid Oil)} = \frac{P(TX \cap Oil)}{P(Oil)} = \frac{18\%}{25\%} = \mathbf{72}\%$

Homework 1.8 – A company is considering an investment. The outcomes can be follows – Success: 20%, Average: 50% and Failure: 30%. The company decides to hire a specialist. His advice is either Yes or No. From past experience, the company knows that - $P\ (Yes \mid Success) = 0.9, P(Yes \mid Average) = 0.2, and\ P(Yes \mid Failure) = 0.1$. The specialist said Yes. What is the probability for success?

Solution 1.8 –

	Success	**Average**	**Failure**
Yes	**18%**	**10%**	**3%**
No	**2%**	**40%**	**27%**
	20%	**50%**	**30%**

Given,

$$\boldsymbol{P\ (Yes \mid Success)} = 0.9 = \frac{P(Yes \cap Success)}{P(Success)}$$

$$\boldsymbol{P\ (Yes \cap Success)} = P(Yes \mid Success) * P(Success) = 0.9 * 0.2 = 0.18 = \mathbf{18}\%,$$

Similarly,

$$\boldsymbol{P\ (Yes \cap Average)} = P(Yes \mid Average) * P(Average) = 0.2 * 0.5 = 0.10 = \mathbf{10}\%,$$

$$\boldsymbol{P\ (Yes \cap Failure)} = P(Yes \mid Failure) * P(Failure) = 0.1 * 0.3 = 0.03 = \mathbf{3}\%$$

Therefore,

$$\boldsymbol{P\ (Success \mid Yes)} = \frac{P(Success \cap Yes)}{P(Yes)} = \frac{18\%}{31\%} = \mathbf{58.06}\%$$