CWDID-20027346

CS-513

"Knowledge Discovery And Data Mining"

Homework 01 - Probability

Solutions :-

1. 1) Given:-

Probability of Jerry goes to bank, $P(J) = 20\% = 0.2$
Probability of Susan goes to bank, $P(S) = 30\% = 0.3$
Probability of together at the bank, $P(J \cap S) = 8\% = 0.08$.

a) Probability of Jerry at bank given Susan was at Bank, $P(J/S) = ?$

WKT,

$$P(J/S) = \frac{P(J \cap S)}{P(S)} = \frac{0.08}{0.3} = \underline{0.26667} \text{ or } \underline{26.667\%}$$

b) Probability of Jerry at the bank given Susan was'nt at the bank, $P(J/S') = ?$

$P(S') \rightarrow$ probability of Susan not at bank $= 1 - P(S) = 1 - 0.3$
$= 0.7$

WKT,

$$P(J/S') = \frac{P(J \cap S')}{P(S')}$$

$$P(J \cap S') = P(J) - P(J \cap S) = 0.2 - 0.08 = 0.12$$

Hence,

$$P(J/S') = \frac{P(J \cap S')}{P(S')} = \frac{0.12}{0.7} = \underline{0.171428} \text{ or } \underline{17.1428\%}$$

c) Probability of both at the bank given atleast one of them at the bank, $P(J \cap S / J \cup S) = ?$

$$P(J \cup S) = P(J) + P(S) - P(J \cap S) = 0.2 + 0.3 - 0.08 = 0.42$$

WKT,

$$P(J \cap S / J \cup S) = \frac{P(J \cap S)}{P(J \cup S)} = \frac{0.08}{0.42} = \underline{0.190476 \text{ or } 19.0476\%}$$

1.2) Probability of Harold getting 'B', $P(H) = 80\% = 0.8$
Probability of Sharon getting 'B', $P(S) = 90\% = 0.9$
Probability of at least one getting 'B', $P(H \cup S) = 91\% = 0.91$

So, we know
$$P(H \cap S) = P(H) + P(S) - P(H \cap S) = 0.8 + 0.9 - 0.91 = 0.79$$

a) Probability of only Harold getting 'B', $P(\text{only } H) = ?$

$$P(\text{only } H) = P(H \cup S) - P(S) = 0.91 - 0.9$$
$$= \underline{0.01 \text{ or } 1\%}$$

b) Probability of only Sharon getting 'B', $P(\text{only } S) = ?$

$$P(\text{only } S) = P(H \cup S) - P(H) = 0.91 - 0.8$$
$$= \underline{0.11 \text{ or } 11\%}$$

c) Probability of both not getting a 'B', $P((H \cup S)') = ?$

$$P((H \cup S)') = 1 - P(\text{H and S getting 'B'})$$
$$= 1 - P(H \cup S)$$
$$= 1 - 0.91 = \underline{0.09 \text{ or } 9\%}$$

1.3) Given:
$P(J) = 20\% = 0.2$; $P(S) = 30\%$ or 0.3, $P(J) \times P(S) = 0.2 \times 0.3 = 0.06$
$P(J \cap S) = 8\% = 0.08$

For the two events to be independent, the events individual parameter's product should be equal to the intersection of the two events. So,
$$P(J \cap S) = P(J) \times P(S)$$
$$0.08 \neq 0.06$$
Hence, proved the events "Jerry is at bank" and "Susan is at the bank" are NOT INDEPENDENT.

1.4) Given:-

Event : Rolling two dice

Sample Space - { (1,1), (1,2), (1,3), (1,4), (1,5), (1,6),
(2,1), (2,2), (2,3), (2,4), (2,5), (2,6),
(3,1), (3,2), (3,3), (3,4), (3,5), (3,6),
(4,1), (4,2), (4,3), (4,4), (4,5), (4,6),
(5,1), (5,2), (5,3), (5,4), (5,5), (5,6),
(6,1), (6,2), (6,3), (6,4), (6,5), (6,6) }

a) Event A: the sum is 6; (1,5), (2,4), (3,3), (4,2), (5,1)
Event B : the second dice shows 5; (1,5), (2,5), (3,5), (4,5)
(5,5), (6,5)

$P(A) = \frac{5}{36}$, $P(B) = \frac{6}{36}$

$P(A \cap B)$ = Probability of both A and B occuring is when we get (1,5).

Hence,

$P(A \cap B) = \frac{1}{36}.$

To check, if the events are independent of each other, it must prove, $P(A \cap B) = P(A) \times P(B)$

So,

LHS: $P(A \cap B) = 1/36$

RHS: $P(A) \times P(B) = 5/36 \times 6/36 = 5/216.$

Hence, LHS $\neq$ RHS, so the events are "<u>not</u> <u>independent</u>"

b) Event A: the sum is 7; (1,6), (2,5), (3,4), (4,3), (5,2), (6,1)
Event B: the first die shows 5; (5,1), (5,2), (5,3), (5,4), (5,5), (5,6)

$P(A) = \frac{6}{36} = \frac{1}{6}$; $P(B) = \frac{6}{36} = \frac{1}{6}$

$P(A \cap B)$ = Probability of both A and B occuring is when we get (5,2)

Hence, $P(A \cap B) = \frac{1}{36}$

P.T.O

To check, if the events are independent of each other, it must prove, $P(A \cap B) = P(A) \times P(B)$.

So,

LHS: $P(A \cap B) = 1/36$

RHS: $P(A) \times P(B) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

Hence, proved LHS = RHS, which means events are "independent".

1.5) Given:

Probability of Company at TX, $P(TX) = 60\% = 0.6$

Probability of Company at NJ, $P(NJ) = 10\% = 0.1$

By probability axioms,

$$P(TX) + P(NJ) + P(AK) = 1$$

$$P(AK) = 1 - P(TX) - P(NJ)$$

$$= 1 - 0.6 - 0.1 = 0.3$$

Probability of Company at AK, $P(AK) = 0.3$ or 30%

Probability of find oil given state is TX $= 30\% = 0.3 = P(O/TX)$

Probability of finding oil given state is AK, $20\% = 0.2 = P(O/AK)$

Probability of finding oil given state is NJ, $10\% = 0.1 = P(O/NJ)$

a) Probability of finding oil, $P(O) = ?$

$$P(O) = P(TX) \times P(O/TX) + P(AK)\,P(O/AK) + P(NJ)\,P(O/NJ)$$

$$= (0.6 * 0.3) + (0.3 * 0.2) + (0.1 * 0.1)$$

$$= 0.18 + 0.06 + 0.01$$

$$= 0.25 \text{ or } 25\%$$

b) Probability of drilling on TX given they found oil.

$$P(TX/O) = \frac{P(O/TX) \times P(TX)}{P(O)} = \frac{0.3 \times 0.6}{0.25} = \frac{0.18}{0.25} = 0.72 \text{ or } 72\%$$

1.6) Given:

Survived

AGE	1st	2nd	3rd	Crew	Sub Total
Adult	197	94	151	212	654
Child	6	24	27	-	57
Sub Total	203	118	178	212	711

Not Survived

AGE	1st	2nd	3rd	Crew	Sub Total
Adult	122	167	476	673	1438
Child	-	-	52	-	52
Sub Total	122	167	528	673	1490

Total

AGE	1st	2nd	3rd	Crew	Sub Total
Adult	319	261	627	885	2092
Child	6	24	79	-	109
Sub-Total	325	285	706	885	2201

a) The probability that a passanger did not survive.

$$P(\text{not survived}) = \frac{\text{No. of passenger din't survive}}{\text{Total number of passanger}} = \frac{1490}{2201}$$

$= 0.676965$ or 67.6965%

b) The probability that a passanger was staying in first class.

$$P(1^{st}\text{ class}) = \frac{\text{No. of passanger in } 1^{st}\text{ Class}}{\text{Total number of passanger}} = \frac{325}{2201} = 0.14766 \text{ or } 14.766\%$$

c) Probability of passenger staying in first class given passenger survived.

$$P(1^{st}\text{ class} / \text{survived}) = \frac{P(1^{st}\text{class} \cap \text{survived})}{P(\text{survived})} = \frac{0.0922}{0.3231} = 0.28536 \text{ or } 28.536\%$$

$$P(1^{st}\text{ Class} \cap \text{survived}) = \frac{\text{No. of passenger survived and in } 1^{st}\text{ class}}{\text{Total passengers}}$$

$$= \frac{203}{2201} = 0.0922$$

$$P(\text{survived}) = 1 - P(\text{not survived}) = 1 - 0.6769 = 0.3231$$

d) We know that,

$P(1^{st}\text{ class}) = 0.14766$ $P(\cancel{\text{not}}\text{ survival}) = 0.3231$

$P(1^{st}\text{ class} \cap \text{survived}) = 0.0922$

The two events to be independent, the events individual parameter's product should be equal to intersection of two events.

$$P(1^{st}\text{ class} \cap \text{survived}) = P(1^{st}) \times P(\text{survived})$$

LHS:- $P(1^{st}\text{class} \cap \text{survived}) = 0.0922$

RHS:- $P(1^{st}\text{ Class}) \times P(\text{survived}) = 0.14766 \times 0.3231 = 0.047$

Hence, proved the events (passanger in 1^{st} class) and event (passenger is survived) are "not independent".

e) The probability that the passenger was in first class and is a child given passenger survived,

$$P(\text{passenger in } 1^{st}\text{ Class being Child} / \text{survived}) = \frac{P(1^{st}\text{ class} \cap \text{Child} \cap \text{survived})}{P(\text{survived})}$$

$$= \frac{6}{711} = 0.008438 \text{ or } 0.8438\%$$

f) The probability that the passenger was a ~~the~~ adult given passenger survived,

$$P(\text{adult} \mid \text{survived}) = \frac{P(\text{adult} \cap \text{survived})}{P(\text{survived})} = \frac{0.29713}{0.3231} = \underline{0.91962} \text{ or } \underline{91.962\%}$$

$$P(\text{adult} \cap \text{survived}) = \frac{654}{2201} = 0.29713$$

g) 1. Probability of adult, 1^{st} class given survived
2. Probability of child, 1^{st} class given survived.

1. $P(\text{adult} \cap 1^{st} \text{ class})$ should be equal to $P(\text{adult}) \times P(1^{st} \text{ class})$ if they are independent.

$$P(\text{adult} \cap 1^{st} \text{ class}) = \frac{197}{711} = 0.277074$$

$$P(\text{adult}) \times P(1^{st} \text{ class}) = \frac{654}{711} \times \frac{203}{711} = 0.262624$$

As they aren't equal, the events are "<u>not independent</u>" given survived

2. $P(\text{child} \cap 1^{st} \text{ class})$ should be equal to $P(\text{child}) \times P(1^{st} \text{ class})$ if they are independent

$$P(\text{child} \cap 1^{st} \text{ class}) = \frac{6}{711} = 0.00843$$

$$P(\text{child}) \times P(1^{st} \text{ class}) = \frac{57}{711} \times \frac{203}{711} = 0.22889.$$

As they aren't equal, the events are "<u>not independent</u>" given survived.

1.7) Given:

- In total, 1000 human generated documents and 1000 AI generated document.
- app misclassified 70 human generated as AI generated
- 30 AI generated as human generated

Confusion matrix:

	AI generated	Human generated	Total
Predicted as AI generated	TP 970	FP 70	1040
Predicted as human generated	FN 30	TN 930	960
Total	1000	1000	2000

* Accuracy $= \frac{TP + TN}{TP+FP+TN+FN} = \frac{970+930}{970+70+930+30} = \frac{1900}{2000} = 0.95$ or 95%

* Precision $= \frac{TP}{TP+FP} = \frac{970}{970+70} = \frac{970}{1040} = 0.93269$ or 93.269%

* Recall $= \frac{TP}{TP+FN} = \frac{970}{970+30} = \frac{970}{1000} = 0.97$ or 97%

* F1 score $= \frac{2 * P * R}{P+R} = \frac{2 \times 0.9326 \times 0.97}{0.9326+0.97} = \frac{1.80924}{1.9026} = 0.95093$ or 95.093%

- AMRUTHA KANAKATTE RAVISHANKAR