MODULE 2

Probabilistic learning

* Events are possible outcomes such as a heads or tails result in a coin flip.
* A Trial is a single opportunity for the event to occur such as a coin flip.

Probability

* The probability of an event can be estimated from observed data by dividing the no. of trials in which an event occurred by the total no. of trials.
* The notation P(A) is used to denote the probability of event A
* The total probability of all possible outcomes of a trial must always be 100 percent.
* Mutually exclusive: when 2 events are mutually exclusive, it means that they cannot both occur at the same time.
* When 2 events are exhaustive, it means that one of them must occur.
* If two events are totally unrelated, they are called independent events.

conditional probability with Bayes' theorem

* The relationship b/w dependent events can be described using Bayes' theorem
* The notation $P(A\backslash B)$ can be read as the probability of event A given that event B occurred. This is known as conditional probability, since the probability of A is dependent on what happened with event B.

$$P(A\backslash B) = \frac{P(B\backslash A)\,P(A)}{P(B)}$$

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

Note:

$$P(A\cap B) = P(A\backslash B)\,P(B)$$
$$= P(B\backslash A)\,P(A)$$

Proof,

$$P(A) = \frac{n(A)}{S}, \quad P(B) = \frac{n(B)}{S}, \quad P(A\cap B) = \frac{n(A\cap B)}{S}$$

$$P(A\backslash B) = \frac{n(A\cap B)}{n(B)}$$

$$= \frac{n(A\cap B)}{S} \cdot \frac{S}{n(B)}$$

$$P(A\backslash B) = P(A\cap B) \cdot P(B)$$

Similarly,

$$P(B\backslash A) = P(A\cap B) \cdot P(A)$$

? A bag contains 3R & 4W balls 2 draws are made without replacement. What is the probability that both are red

A) $P(A) = P(\text{1st ball is red})$

$P(B|A) = P(\text{2nd ball is red without replacement})$

$P(A \cap B) = P(A) \cdot P(B|A)$

$= 3/7 \cdot 2/6$

$= \underline{1/7}$

? 3 baskets are given each containing R & W balls as given below

B_1	B_2	B_3
6R	2R	1R
4W	6W	8W

A basket is selected at random and a ball is drawn from it. the ball is red. Find the probability that the selected basket is B_1

A). E_1: choosing basket 1

E_2: " " 2

E_3: " " 3

~~$P(E_1$~~

$P(E_1) = P(E_2) = P(E_3) = 1/3$

A be the event of drawing a red ball

$P(A|E_1) = 6/10 = 3/5$

$P(A|E_2) = 2/8 = 1/4$

$P(A|E_3) = 1/9$

$$\overline{P(E_1 \cap A) + P(E_2 \cap A) + P(E_3 \cap A)}$$

$$= \frac{P(E_1 \cap A)}{\sum_{i=1}^{3} P(A|E_i) P(E_i)}$$

$$= \frac{P(E_1) \cdot P(A|E_1)}{\sum_{i=1}^{3} P(A|E_i) P(E_i)}$$

$$= \frac{1/3 \cdot 3/5}{1/3 \cdot 3/5 + 1/3 \cdot 1/4 + 1/3 \cdot 1/9}$$

$$= \frac{1/5}{1/5 + \frac{1}{12} + 1/27}$$

$$= \frac{1/5}{\frac{17}{60} + 1/27}$$

$$= \frac{1/5}{\frac{519}{1620}} = \frac{324}{519}$$

b) a basket is chosen at random and 2 balls are drawn without replacement from the same basket if both balls are red then what is the probability that the ball was selected from B_1 under this condition. what is the probability that the B_3 is selected.

A) A : be the both ball are red

~~P(E₁|A)=~~

$P(A \setminus E_1) = \frac{6C2}{10C2}$

$P(A \setminus E2) = \frac{2C2}{8C2}$

$P(A \setminus E3) = 0$

$$P(E_1 \setminus A) = \frac{P(E_1) \cdot P(A \setminus E_1)}{P(E_1) \cdot P(A \setminus E_1) + P(E2) P(A \setminus E2) + P(E3) P(A \setminus E_3)}$$

$$= \frac{1/3 \cdot \frac{6C2}{10C2}}{1/3 \cdot \frac{6C2}{10C2} + 1/3 \cdot \frac{2C2}{8C2} + 1/3 \cdot 0}$$

$$= \frac{1/9}{\frac{1}{9} + 1/28}$$

$$= \frac{1/9}{37/228}$$

$$= \frac{228}{9 \cdot 37}$$

$$= \frac{228}{333}$$

Bayes' theorem

If $E_1, E2, \ldots E_n$ are mutually exclusive events with probability $P(E_i) \neq 0$ for $i = 1, 2, \ldots n$ Then, for any event 'A' that is a subset of $\sum_{i=1}^{n} E_i$ such that $P(A) > 0$

Then we've,

$$P(E_i \setminus A) = \frac{P(E_i) \cdot P(A \setminus E_i)}{\sum_{i=1}^{n} P(E_i) P(A \setminus E_i)}$$

2. A bag contain 5 balls. 2 balls are drawn and are found to be white. what is the probability that all the balls are being white.

event

A) $E_1 \to$ event that basket contains atleast 2 white ball

$E_2 \to$ " " " " " 3 " "

$E_3 \to$ " " " " " 4 " "

$E_4 \to$ " " " " " 5 " "

$P(A \setminus E_1) = \frac{2C2}{5C2} = 1/10$ $\quad P(A \setminus E_3) = \frac{4C2}{5C2} = 6/10$

$P(A \setminus E_2) = \frac{3C2}{5C2} = 3/10$ $\quad P(A \setminus E_4) = 1$

$P(E_1) = P(E_2) = P(E_3) = P(E_4) = 1/4$

$$P(E_4 \setminus A) = \frac{P(E_4)\, P(A \setminus E_4)}{\Sigma P(E_i)\, P(A \setminus E_i)}$$

$$= \frac{1/4}{1/4 \cdot 1/10 + 3/10 \cdot 1/4 + 1/4 \times 6/10 + 1/4}$$

$$= \frac{1/4}{\frac{1}{40} + 3/40 + 6/40 + 1/4}$$

$$= \frac{1/4}{1/4 + 1/4}$$

$$= 1/4 \cdot 2/4$$

$$= \frac{2}{16} = 1/8$$

2. A bag contain 7 red & 3 black marble. Another bag contains 4 red and 5 black marble. one marble is transport from 1st bag to the 2nd bag and then a

event

marble is taken out of the second bag. If this marble to be red Find the probability that a black marble was found out.

A) ~~$E_1 \to$ red marble transfered~~

~~A = Transfered marble is red.~~

~~$E_2 \to$ black " "~~

~~A) (black marble wrt E_i)~~

7R 3B   4R 5B

~~$P(A|E_1) = 5/10$~~

~~$P(A|E_2) =$~~

~~A: Moved Marble is red~~

E_1: A black marble is transfered

E_2: red marble is transferred.

A: Marble is red

we've to find $P(E_1|A)$?

$P(E_1) = 3/10$ $P(E_2) = 7/10$

$P(A|E_1) = 4/10$ (probability of getting red marble when black marble is transferred)

$P(A|E_2) = 5/10$

$$P(E_1|A) = \frac{P(E_1)\cdot P(A|E_1)}{P(E_1)\cdot P(A|E_1) + P(E_2)\cdot P(A|E_2)}$$

$$= \frac{3/10 \cdot 4/10}{3/10 \cdot 4/10 + 7/10 \cdot 5/10}$$

$$= \frac{\frac{12}{100}}{\frac{12}{100} + \frac{35}{100}}$$

$$= \frac{12}{47}$$

2. The members of a consulting firm hire cars or vehicles from rental agencies X, Y, Z as 60%, 30% and 10% respectively. If 9%, 20% and 6% of cars from agencies X Y and Z turnup and if a rental car delivered to the firm doesn't turnup. what is the probability that the vehicle came from agency Y

A) $E_1 \to$ vehicle came from agency X.

$E_2 \to$ " " " " Y.

$E_3 \to$ " " " " Z

$P(E_1) = 60\% = \frac{6}{10}$

$P(E_2) = 30\% = \frac{3}{10}$

$P(E_3) = 10\% = 1/10$

A $\to$ rental car delivered to the firm doesn't turnup.

$P(A \setminus E_1) = 100 - 9\%$

$= 91\% = \frac{91}{100}$

$P(A \setminus E_2) = 89/100$

$P(A \setminus E_3) = \frac{94}{100}$

$91 \times 6 = 5460$

$$P(E_2 \setminus A) = \frac{P(E_2) \cdot P(A \setminus E_2)}{\sum_{i=1}^{3} P(E_i) P(A \setminus E_i)}$$

$$= \frac{\frac{30}{100} \cdot \frac{80}{100}}{\frac{30}{100} \cdot \frac{80}{100} + \frac{60}{100} \cdot \frac{91}{100} + \frac{10}{100} \cdot \frac{94}{100}}$$

$$\frac{240}{240+5460+940}$$

$$= \frac{240}{6640}$$

$$= \underline{\underline{0.0361}}$$

2. In a factory machine X and Y produces springs of same kind. Of this production machine x & y produce 5% & 10% defective spring respectively. Machine X & Y produce 40% 60% total o/p of the factory. Once a spring selected at random found to be defective. What is the probability that the defective spring was produced by machine x.

A) $E_1 \to$ spring produced by machine x

$E_2 \to$ " " " " y

$A \to$ selected spring is defective

$P(E_1) = \frac{40}{100}$ $\quad P(A|E_1) = \frac{5}{100}$

$P(E_2) = \frac{60}{100}$ $\quad P(A|E_2) = \frac{10}{100}$

$$P(E_1|A) = \frac{P(E_1) \cdot P(A|E_1)}{\sum_{i=1}^{2} P(E_i) P(A|E_i)}$$

$$= \frac{\frac{40}{100} \cdot 5/100}{\frac{40}{100} \cdot 5/100 + \frac{60}{100} \cdot \frac{10}{100}}$$

$$= \frac{200}{200+600} = \frac{\cancel{200}}{\cancel{800}} = \underline{\underline{1/4}}$$

2. Assume that the word 'offer' occurs in 80% of spam messages. Also let us assume offer occurs in 10% of desired mail. If 30% of the received emails are considered as spam and received a new message contains offer. what is the probability that it is a spam.

A) $E_1 \rightarrow$ spam message

$E_2 \rightarrow$ desired but spam

$A \rightarrow$ recieved message contains 'offer'

$P(E_1) = 30/100 \quad P(E_2) = 70/100$

$P(A \backslash E_1) = 80/100 \quad P(A \backslash E_2) = 10/100$

$$P(E_1 \backslash A) = \frac{P(E_1) \cdot P(A \backslash E_1)}{\Sigma P(E_i) P(A \backslash E_i)}$$

$$= \frac{\frac{30}{100} \cdot 80/100}{70/100 \times 10/100 + \frac{30}{100} \cdot 80/100}$$

$$= \frac{240}{700 + 240}$$

$$= \frac{24}{94} = \frac{12}{47}$$

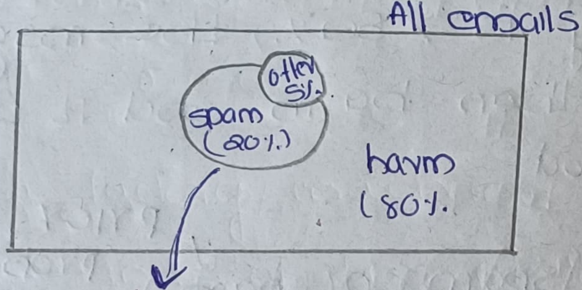
Joint probability

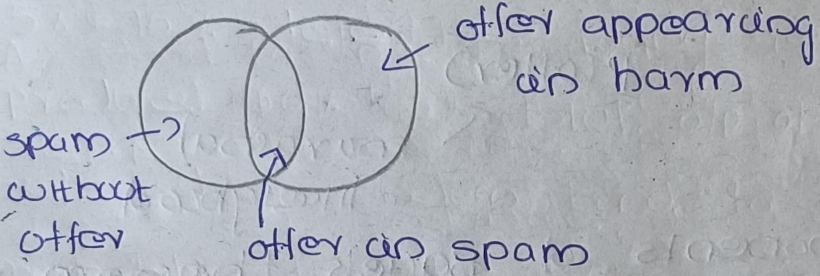
* consider the following venn diagram,

* offer circle doesnot completely fill the spam circle, nor is it completely contained by the spam circle

15

This implies that not all spam messages contain the word offer, and not every email with the word offer is spam.





* calculating P(spam ∩ offer) depends on joint probability of the two events or how the probability of one event is related to the probability of the other.

* To understand how Baye's theorem works in practice, suppose that you were tasked with guessing the probability that an incoming email was spam. without any additional evidence, the most reasonable guess would be the probability that any prior message was spam(R₁ %.20). This estimate is known as the prior probability.

* suppose you were told that the incoming msg uses the term offer. The probability that the word offer was used in previous spam msg is called the likelihood and the probability that offer appeared in any msg at all is known as the marginal likelihood

* By applying Bayes' theorem to this evidence, we can compute a posterior probability that measures how likely the msg is to be spam.
* If the posterior probability is > 50%, the msg is more likely to be spam than harm, and it should be filtered.

likelihood

prior probability

$$P(\text{spam} \mid \text{offer}) = \frac{P(\text{offer} \mid \text{spam})\, P(\text{spam})}{P(\text{offer})}$$

posterior probability

marginal likelihood

* To calculate the components of Bayes theorem, we must construct a frequency table that records the no. of times offer appeared in spam and harm msgs.
* The frequency table can then be used to construct a likelihood table.

*

Frequency	offer: yes	offer: No	Total
spam	4	16	20
harm	1	79	80
Total	5	95	100

Likelihood	offer: yes	offer: No	Total
spam	4/20	16/20	20
harm	1/80	79/80	80
Total	5/100	95/100	100

The naive Bayes algorithm

* The naive Bayes (NB) algorithm describes a simple application using Bayes' theorem for classification.
* The naive Bayes algorithm is named as such bcoz it

makes a couple of 'naive' assumptions about the data.

* In particular, naive Bayes assumes that all of the features in the dataset are equally important & independent. These assumptions are rarely true in most of the real-world applications.
* For example, if you were attempting to identify spam by monitoring email messages, it is almost certainly true that some features will be more imp than others.
* Naive Bayes algorith is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.

The naive Bayes Classification

eg:
The naive Bayes learner is trained by constructing a likelihood table for the appearance of 4 words (w_1, w_2, w_3 and w_4) for 100 emails

Likelihood	offer(w_1) yes	offer(w_1) NO	Money(w_2) yes	Money(w_2) NO	Loan(w_3) yes	Loan(w_3) NO	unsubscribe(w_4) yes	unsubscribe(w_4) NO
spam	4/20	16/20	10/20	10/20	0/20	20/20	12/20	8/20
ham	1/80	79/80	14/80	66/80	8/80	72/80	23/80	51/80
Total	5/100	95/100	24/100	76/100	8/80	92/100	35/100	65/100

calculate the probability that the message is a spam given that offer = yes, money = No, loan = No, unsubscribed = yes.

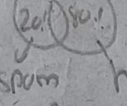
A) $P(s|w_1 \cap \neg w_2 \cap \neg w_3 \cap w_4) = \dfrac{P(s).P(w_1 \cap \neg w_2 \cap \neg w_3 \cap w_4 | s)}{\underbrace{P(s).P(w_1 \cap \neg w_2 \cap \neg w_3 \cap w_4 | s)}_{T_1} + \underbrace{P(h).P(w_1 \cap \neg w_2 \cap \neg w_3 \cap w_4 | h)}_{T_2}}$

$= \dfrac{T_1}{T_1 + T_2}$

$T_1 = \overset{P(s)}{\cancel{4/20}} \cdot P(w_1|s).P(\neg w_2|s)\,P(\neg w_3|s)\,P(w_4|s)$

$= \cancel{4/20}$:

($P(S)$ = prior probability



$P(S) = \frac{20}{100}$)

$= \frac{\cancel{20}}{\cancel{100}_{25}} \cdot \frac{4}{\cancel{20}} \times \frac{10}{20} \times \frac{\cancel{20}}{\cancel{20}} \times \frac{\overset{6}{\cancel{12}}}{20}$

$= \dfrac{\overset{3}{\cancel{6}}}{\underset{10}{\cancel{20}} \cdot 25} = \dfrac{3}{250}$

$T_2 = P(H) \cdot P(w_1|H).P(\neg w_2|H)\,P(\neg w_3|H)\,P.(w_4|H)$

$= \frac{80}{100} \cdot \frac{1}{80} \cdot \frac{66}{\underset{10}{\cancel{80}}} \cdot \frac{\overset{9}{\cancel{72}}}{80} \cdot \frac{23}{80}$

$= \dfrac{13662}{6400000}$

$= 0.002$

Desired probability $= \dfrac{0.012}{0.002 + 0.012} = \underline{\underline{0.857}}$

The Laplace estimator

consider the above problem,

If we've to find the $P(s|w_1 \cap w_2 \cap w_3 \cap w_4)$

clearly likelihood of spam is,

$(4/20) \times (10/20) \times (0/20) \cdot (12/20) \cdot (20/100) = 0$

$\therefore P(spam)$ is $\frac{0}{0+0.009} = 0$

& $P(harm)$ is $\frac{0.00005}{0+0.00005} = 1$.

* This results suggest that the msg is spam with 0/ and harm with 100% probability. It is ∴ very likely that the msg has been incorrectly classified.
* This problem might arise if an event never occurs for one or more levels of the class.
* A solution to this problem involves using something called the Laplace estimator,
* the laplace estimator essentially adds a small no to each of the counts in the frequency table, which ~~esti~~ ensures that each feature has a nonzero probability of occurring with each class.
* Typically, laplace estimator is set to 1, which ensures that each class-feature combination is found in the data atleast once.

20/12/2021

? Given the following data of set of patients can the dr conclude that the person having chills mild head ache and without running nose has fever flue. use naive basian theorem.

chil

chills (A)	Running Nose (B)	head ache (C)	Fever D	has flue E
yes ✓	NO ✓	✓ mild	yes ✓	NO
yes	yes	NO	NO	yes
yes	NO	strong	yes	yes
NO	yes	mild	yes	yes
NO	NO	NO	NO	NO
NO	yes	strong	yes	yes
NO	yes	strong	NO	NO
yes	yes	Mild	yes	yes

$$P(\text{Flue} \mid \underbrace{\text{chills}=y,\ RN=N,\ HA=\text{mild},\ \text{Fever}=y}_{T_1})$$

$$= \frac{P(F) \cdot P(\text{chills}=y \mid \text{Flue}) \cdot P(RN=N \mid \text{Flue}) \cdot P(HA=\text{Mild} \mid \text{Flue}) \cdot P(\text{Fever}=y \mid \text{Flue})}{P(T_1 + T_2)}$$

$$T_2 = P(\text{Flue}=\text{No} \mid P(\text{chills}=y,\ RN=N, \qquad))$$

$$P(\text{Flue} \mid \text{chills}=y,\ \text{heac}=N,\ \text{Fever}=y)$$

$$= \frac{5/8\,(3/5 \cdot 2/5 \cdot 1/8 \cdot 4/8)}{(5/8 \cdot 3/5 \cdot 2/5 \cdot 1/5 \cdot 4/5) + (}$$

[5 → 5 persons have flue in table., out of which how many people have chills, headache ......]

no. of persons have no Flue = 3

$P(A|E=N) = 1/3$

$P(B=N|E=N) = 2/3$

$P(C=M|E=N) = 1/3$

$P(D=Y|E=N) = 1/3$

$P(E=N) = 3/8.$

$P(\text{Flue}=\text{No}|T_1) = 3/8 \cdot 1/3 \cdot 2/3 \cdot 1/3 \cdot 1/3$ $\quad 12 \times 9 = 108$

$= 3/8 \cdot 2/9 \times 9$

$= \frac{6}{8 \times 9 \times 9} = \frac{2}{\cancel{8}_4 \times 3 \times 9} = \frac{1}{12 \times 9} = \underline{\underline{1/108}}$

required probability

$$= \frac{5/8 \cdot \frac{6 \times 4}{5 \times 5 \times 5 \times 5} \cancel{8 \times 8 \times 8 \times 8}}{5/8 \cdot \frac{6 \times 4}{\cancel{8} \times 5 \times 5 \times 5} + 1/108}$$

$$= \frac{\cancel{5}/\cancel{8} \cdot \frac{\cancel{24}^{3}}{25 \times 5}}{1/8 \cdot \frac{24}{25 \times 5} + 1/108}$$

$$= \frac{\frac{3}{125}}{\frac{3}{125} + 1/108} = \frac{\frac{3}{125}}{\frac{429}{125 \cdot 108}}$$

$$= \frac{3 \times \cancel{125} \cdot 108}{\cancel{125} \cdot \cancel{108}\ 429}$$

$$= \frac{3 \times 108}{429} =$$

$\begin{array}{r} 324 \\ 3\ 2 \\ \hline 304 + \\ 125 \\ \hline 429 \end{array}$

2 using naive baye algorithm determine whether a red domestic SUV car is stolen or not using following data

Example	A color	B Type	C origin	D stolen
1	Red	sports	Domestic	y
2	Red	sports	Domestic	N
3	Red	sports	Domestic	y
4	yellow	sports	Domestic	N
5	yellow	sports	imported	y
6	yellow	SUV	imported	N
7	yellow	SUV	Imported	y
8	yellow	SUV	Domestic	N
9	Red	SUV	Imported	N
10	Red	sports	imported	y

$P(\text{stolen} = y \mid \text{color} = \text{red}, \text{Type} = \text{SUV}, \text{origin} = \text{domestic}) = ?$

$P(D=y) = 5/10$

$P(A=R \mid D=y) = 3/5$

$P(B=SUV \mid D=y) = 1/5$

$P(C=Dom \mid D=y) = 2/5$

$P(D=N) = 5/10$

$P(A=R \mid D=N) = 2/5$

$P(B=SUV \mid D=N) = 3/5$

$P(C=Dom \mid D=N) = 3/5$

$$= \frac{5/10 \cdot 3/5 \cdot 1/5 \cdot 2/5}{(5/10 \cdot 3/5 \cdot 1/5 \cdot 2/5) + (5/10 \cdot 2/5 \; 3/5 \cdot 3/5}$$

$$= \frac{5 \cdot 3 \cdot 2}{(5 \cdot 3 \cdot 2) + 5 \cdot 2 \cdot 3 \cdot 3)}$$

$$= \frac{30}{30+90} = \frac{3\cancel{0}}{12\cancel{0}} = 3/12 = \underline{\underline{1/4}}$$

Numerical Features on Bayes

* For eg:
* consider the spam email problem,

sir, Dear → This is not considered as features, some significant words such as unsubscribed, offer ... etc are called Feature.

* categorical feature (p(offer included or not ?))

when time is considered with categorical feature, It is difficult to construct frequency table.

∴ we divide 24 hr duration in to some parces.

~~these time intervals~~ are called Bins.

* Time divided into different small interval called Bin.
* Discretization of numerical feature is called binning or, discretization of values called binning.