1. Define Experiment, Sample space, Outcome and Event.

2. What is probability and explain different types of probability?

3. In loan defaulters older people make up only 1.4%. Now the probability that someone defaults on a loan is 0.184, Find the probability of default on loan knowing that he is an old person. Older people make up only 0.8%

4. Define Bayes theorem and write the formulae.

5. Solve the below problem using Bayes theorem:

Spam Assassin works by having users train the system. It looks for patterns in the words in emails marked as spam by the user.

For example, it may have learned that the word “free” appears in 30% of the mails marked as spam, i.e., P(Free | Spam) = 0.30. Assuming 1% of non-spam mail includes the word “free” and 50% of all mails received by the user are spam, find the probability that a mail is spam if the word “free” appears in it.

1) <u>Experiment</u> : It is a procedure that can be repeated infinite no. of times and has a well-defined set of possible out comes.

<u>Random Experiment</u>: Experiment we conduct where the output is uncertain.

<u>Sample space</u>: This is the space that has all the possible outcomes for the experiment.

<u>Event</u>: These are the favourable outcomes in the experiment. It is a subset of sample space.

2) <u>Probability</u>: The extent to which an event is likely to occur.

$$P(E) = \frac{n(\text{favourable cases})}{n(\text{total cases})}$$

<u>Different types of probability</u>:

1) Theoritical probability
2) Experimental probability
3) Axiomatic probability

3) P(older people | loan default) = 1.4%

= 0.014

P(loan default) = 0.184

P(older people) = 0.8%

= 0.008

P(loan default | older people) = ??

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$ (Bayer's theorem)

∴ P(loan default | older people)

$$= \frac{P(\text{older people} \mid \text{loan default}) \cdot P(\text{loan default})}{P(\text{older people})}$$

$$= \frac{0.014 \times 0.184}{0.008}$$

= 0.322

4) <u>Bayes theorem</u>: If A and B are two events, then the formula for the Bayes theorem is given by:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad \text{where } P(B) \neq 0$$

where P(A|B) is the probability of condition when event A is occuring while event B has already occured.

<u>Bayes theorem Derivation</u>

Bayes theorem can be derived for events and random variables using the definition of conditional probability.

from the definition of conditional probability.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{where } P(B) \neq 0$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \quad \text{where } P(A) \neq 0$$

Here, the joint probability $P(A \cap B)$ of both events A and B being true such that

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

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

$$\therefore\ P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \quad \text{where } P(B) \neq 0$$

5) $P(\text{Free}|\text{spam}) = 0.30$

$P(\text{Free}|\text{non-spam}) = 1\% = 0.01$

$P(\text{spam}) = 50\% = 0.5$ $\quad \therefore\ P(\text{non-spam}) = 0.5$

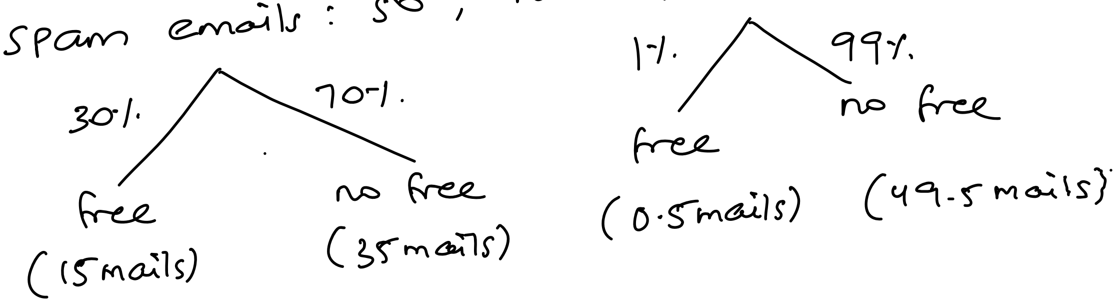
$P(\text{spam}|\text{free}) = ??$

<u>method 1 to find the answer:</u>

$$P(\text{spam}|\text{free}) = \frac{P(\text{free}|\text{spam}) \cdot P(\text{spam})}{P(\text{free})}$$

Total emails : 100

spam emails : 50 , non-spam emails = 50



$\therefore$ total free emails = 15.5

$$p(\text{free}) = \frac{15.5}{100} = 0.155$$

$$p(\text{spam} | \text{free} = \frac{0.3 \times 0.5}{0.155} = 0.96774$$

method 2 to find answer:

formula:

$$p(\text{spam} | \text{free})$$

$$= \frac{p(\text{spam}) * p(\text{free} | \text{spam})}{p(\text{free} | \text{spam}) * p(\text{spam}) + p(\text{free} | \text{no-spam}) * p(\text{no-spam})}$$

$$= \frac{0.5 \times 0.3}{0.3 \times 0.5 + 0.01 \times 0.5}$$

$$= \frac{0.15}{0.15 + 0.005}$$

$$= 0.96774$$