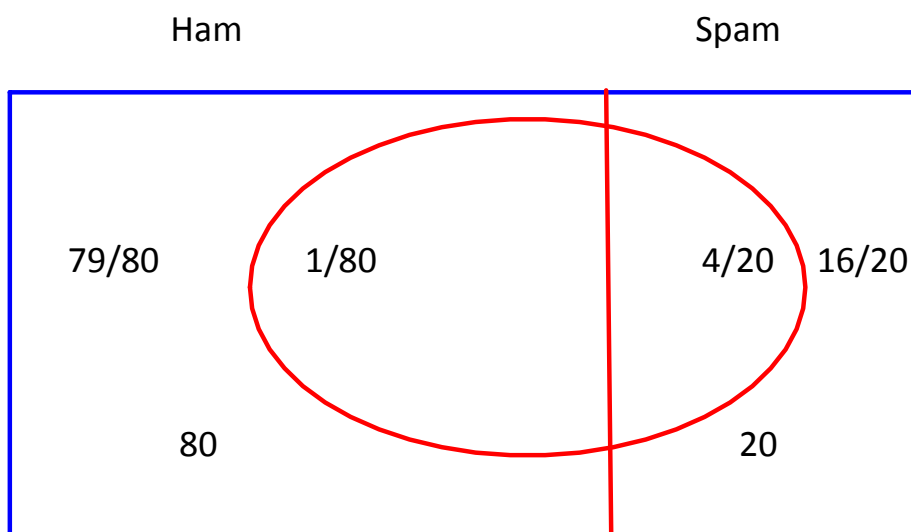


## Application

Whatever the mails we received in our email are divided in two sets i.e Ham and Spams. Whenever a message receives I need to predict/classify it is either Ham or Spam. How?



- ★ We have the **Prior information** as 20% of spam and 80% of Ham.  
Therefore Prior probabilities are  
 $P(\text{Spam}) = 0.2$   
 $P(\text{Ham}) = 0.8$

Assume that the word "Viagra" is in a statement might be in spam email or ham emails. In general Most of the times will be seen in Ham mails and less number of times in spam mails. So, how the word recognition to be categorized in this exercise.

| TYPE  | V. YES | V. NO | TOTAL |
|-------|--------|-------|-------|
| SPAM  | 4      | 16    | 20    |
| HAM   | 1      | 79    | 80    |
| TOTAL | 5      | 95    | 100   |

? How many mails are there which containing word Viagra in spam?

$$P(\text{Viagra/Spam}) = 4/20$$

? How many mails are there which containing word Viagra in Ham?

$$P(\text{Viagra/Ham}) = 1/80$$

? How many mails are there which not containing word Viagra in Spam?

$$P(\text{not Viagra/Spam}) = 16/20$$

? How many mails are there which not containing word Viagra in Ham?

$$P(\text{not Viagra/Ham}) = 79/80$$

? **The above information is already having with us with the past data. Now, what is the chance that when a new mail received which will be going to Spam or Ham?**

? What is the chance or Probability that the word Viagra going to be saved in spam?

$$P(\text{Spam / Viagra}) = ?$$

? What is the chance or Probability that the word Viagra going to be saved in Ham?

$$P(\text{Ham / Viagra}) = ?$$

What is the formulae we have now

$$\text{So, Baye's rule is } P(B_i/A) = \frac{P(A/B_i) \cdot P(B_i)}{\sum P(A/B_i) \cdot P(B_i)}$$

Here Assume that Bi's are Spam and Ham and 'A' is new word received in message.

$$P(B1) = 0.2, P(B2) = 0.8$$

$$\text{Similarly } P(A/B1) \text{ is } P(A/B1) = P(\text{Viagra/Spam}) = 4/20$$

$$\text{Similarly } P(A/B2) \text{ is } P(A/B2) = P(\text{Viagra/Ham}) = 1/80$$

So, either we calculate for  $P(B1/A)$  or  $P(B2/A)$  the denominator remains same.