1. A spam filter is designed by looking at commonly occurring phrases in spam. Suppose that 80% of email is spam. In 10% of the spam emails, the phrase “free money” is used, whereas this phrase is only used in 1% of non-spam emails. A new email has just arrived, which does mention “free money”. What is the probability that it is spam?

Ans:-

Let's create some notations

S - an email is a spam  
FM - 'free money' phrase is used  
N - an email is normal (not a spam)

We are asked to find  
P(S|FM)P(S|FM)

Notice P(S)+P(N)=1P(S)+P(N)=1 and apply Baysian formula

P(S|FM)=P(FM|S)P(S)P(FM|S)P(S)+P(FM|N)P(N)P(S|FM)=P(FM|S)P(S)P(FM|S)P(S)+P(FM|N)P(N)

where

P(FM|S)=10%, P(FM|N)=1%P(FM|S)=10%, P(FM|N)=1%

Plug all back to the conditional probability, I got

P(S|FM)=4041