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?

Answer:

Let S be the event that the email is spam and F be the event that the email contains the phrase "free money". We want to find the conditional probability P(S|F).

By Bayes' theorem, we have:

P(S|F) = P(F|S)P(S) / P(F)

We can compute each of these probabilities:

* P(F|S) = 0.1, as 10% of spam emails contain the phrase "free money".
* P(S) = 0.8, as 80% of all emails are spam.
* P(F) = P(F|S)P(S) + P(F|S^c)P(S^c) = 0.1*0.8 + 0.01*0.2 = 0.082, by the law of total probability and using that the phrase "free money" only appears in 1% of non-spam emails.

Substituting these values, we get:

P(S|F) = 0.1\*0.8 / 0.082 ≈ 0.973

So the probability that the email is spam, given that it contains the phrase "free money", is approximately 0.973 or 97.3%.