Introduction to Naïve Bayes

- Kathirmani Sukumar

Bayes’ Theorem

In probability theory and statistics, **Bayes’ theorem** (alternatively **Bayes’ law** or **Bayes' rule**) describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

(source https://en.wikipedia.org/wiki/Bayes%27_theorem)

Bayes’ Theorem

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

P(A|B) = Probability of A given that B happens

P(A) = Probability of A

P(B|A) = Probability of B given that A happens

P(B) = Probability of B

$$P(Rain|Cloud) = \frac{P(Rain)\,P(Cloud|Rain)}{P(Cloud)}$$

Spam Filtering

Email Subject	Label (Spam / Not Spam)
Great offer ends today	Spam
Your twitter Account is ready	Not Spam
Your account has won 100 crores	Spam
Get expert opinion for your retirement. Offer ends today	Spam
Payment Acknowledgement	Not Spam
Congratulations. Your coupon has won Ipad today	Spam
You won lottery worth 10 crores	???

Bayes’ Theorem

$$P(Spam|You\ won\ \ lottery\ worth\ 10\ crores) = \frac{P(You\ won\ \ lottery\ worth\ 10\ crores|Spam)\ * P(Spam)}{P(You\ won\ lottery\ worth\ 10\ crores)}$$

$$P(Not\ Spam|You\ won\ lottery\ worth\ 10\ crores) = \frac{P(You\ won\ a\ lottery\ worth\ 10\ crores|Not\ Spam)\ * P(Not\ Spam)}{P(You\ won\ lottery\ worth\ 10\ crores)}$$

Being Naive

$P($You won lottery worth 10 crores | Spam)

↓

P(you | Spam)

*

P(won | Spam)

*

P(lottery | Spam)

*

P(worth | Spam)

*

P(crores | Spam)

(ignoring common words and numbers)

$P($You won lottery worth 10 crores | Not Spam)

↓

P(you | Not Spam)

*

P(won | Not Spam)

*

P(lottery | Not Spam)

*

P(worth | Not Spam)

*

P(crores | Not Spam)

(ignoring common words and numbers)

Count of words

Number of words (including duplicates) under spam category: 22 words

Great-2, offer-2, ends-2, today-3, your-3, account-1, won-1, crores-1, get-1, expert-1, opinion-1, retirement-1, congratulations-1, coupon-1, ipad-1,

Number of words (including duplicates) under not spam: 6 words (ignoring *is*)

(your-1, twitter-1, account-1, ready-1, payment-1, acknowledgement-1)

Count of words

Number of unique words under spam category: 15 words (ignoring *has, 100, for, is*)

(*great, offer, ends, today, your, account, won, crores, get, expert, opinion, retirement, congratulations, coupon, ipad)*

Number of unique words under Not Spam category: 6 words (ignoring *is*)

(*your, twitter, account, ready, payment acknowledge)*

Total number of unique words: 19 words (ignoring *has, 100, for, is*)

(*great, offer, ends, today, your, account, won, crores, get, expert, opinion, retirement, congratulations, coupon, ipad, twitter, ready, payment, acknowledgement)*

Being Naive

$$P(word|spam) = \frac{No. of\ times\ the\ word\ appeared\ in\ spam\ rows + 1}{Total\ no. of\ words\ in\ spam\ rows + No. of\ unique\ word\ in\ all\ rows}$$

P(you | Spam) = (0 + 1) / (22 + 19)

P(won | Spam) = (2 + 1) / (22 + 19)

P(lottery | Spam) = (0 + 1) / (22 + 19)

P(worth | Spam) = (0 + 1) / (22 + 19)

P(crores | Spam) = (1+1)/(22 + 19)

P(you won lottery worth crores | Spam) = $5.17e^{-8}$

P(you | Not Spam) = (0 + 1) / (6 + 19) = 0.04

P(won | Not Spam) = (0 + 1) / (6 + 19) = 0.04

P(lottery | Not Spam) = (0 + 1) / (6 + 19) = 0.04

P(worth | Not Spam) = (0 + 1) / (6 + 19) = 0.04

P(crores | Not Spam) = (0 + 1)/(6 + 19) = 0.04

P(you won lottery worth crores | Not Spam) = $1.02e^{-7}$

Bayes' Theorem

$$P(Spam|You\ won\ \ lottery\ worth\ 10\ crores) = \frac{P(You\ won\ \ lottery\ worth\ 10\ crores|Spam)\ \ * P(Spam)}{P(You\ won\ lottery\ worth\ 10\ crores)}$$

$$P(Not\ Spam|You\ won\ lottery\ worth\ 10\ crores) = \frac{P(You\ won\ a\ lottery\ worth\ 10\ crores|Not\ Spam)\ \ * P(Not\ Spam)}{P(You\ won\ lottery\ worth\ 10\ crores)}$$

$P(Spam|You\ won\ \ lottery\ worth\ 10\ crores)$ = 5.17e^{-8} * *(4/6)* = 3.452e^{-8}

$P(Not\ Spam|You\ won\ lottery\ worth\ 10\ crores)$= 1.02$e^{-7}$ * *(2/6)* = 3.413e^{-8}

Probability of spam is higher. Hence using Naïve Bayes, we can categorize the new sentence as Spam

PS: P(You won lottery worth 10 crores) is ignored in denominator since it is common for both the calculations