category	Email text
Not Spam	"Hi there, how are you?"
Not spam	"Meeting at 3 PM tomorrow"
Not spam	"Please send the report"
spam	"win a free prize now!"
spam	"claim your discount today"
spam	"Limited time offer: click here"
?	"free meeting tomorrow" (To classify)
?	"claim your free prize" (To classify)

Sol^n ① Total unique words in spam: 14

Total unique words in notspam: 14

Vocabulary Size = 28

using smoothing:

Now, To classify "free meeting tomorrow"

we need to ~~do~~ find;

$$P(\text{free}|\text{spam}) = \frac{1+1}{14+28} = \frac{2}{42}$$

$$P(\text{meeting}|\text{spam}) = \frac{0+1}{14+28} = \frac{1}{42}$$

$$P(\overset{\text{tomorrow}}{\cancel{\text{meeting}}}\,|\,\text{spam}) = \frac{0+1}{14+28} = \frac{1}{42}$$

$$P(\text{free}|\text{Notspam}) = \frac{0+1}{14+28} = \frac{1}{42}$$

$$P(\text{meeting}|\text{Notspam}) = \frac{1+1}{14+28} = \frac{2}{42}$$

$$P(\text{tomorrow}\,|\,\text{notspam}) = \frac{1+1}{14+28} = \frac{2}{42}$$

Prior probabilities

$P(spam) = \frac{3}{6} = 0.5$

$P(notspam) = \frac{3}{6} = 0.5$

{(3 not spam and 3 spam emails, total 6)}

Now,

$P(spam|free, meeting, tomorrow) \propto P(spam) \times P(free|spam) \times P(meeting|spam) \times P(tomorrow|spam)$

$\propto 0.5 \times \frac{2}{42} \times \frac{1}{42} \times \frac{1}{42}$

≈ 0.00001349

$P(notspam|free, meeting, tomorrow) \propto P(notspam) \times P(free|notspam) \times P(meeting|notspam) \times P(tomorrow|notspam)$

$\approx 0.5 \times \frac{2}{42} \times \frac{1}{42} \times \frac{2}{42}$

≈ 0.00002699

since P(notspam) > P(spam), the email is not spam.

$\text{Normalization} = \frac{0.00002699}{0.00002699 + 0.00001349} \times 100\%$

$= \frac{0.00002699}{0.00004048} \times 100\% = 66.67\%$ not spam

② for email "claim your free prize"

using smoothing

$P(\text{claim} \mid \text{spam}) = \frac{1+1}{14+28} = \frac{2}{42}$

$P(\text{your} \mid \text{spam}) = \frac{0+1}{14+28} = \frac{1}{42}$

$P(\text{free} \mid \text{spam}) = \frac{1+1}{14+28} = \frac{2}{42}$

$P(\text{Prize} \mid \text{spam}) = \frac{1+1}{14+28} = \frac{2}{42}$

$P(\text{claim} \mid \text{notspam}) = \frac{0+1}{14+28} = \frac{1}{42}$

$P(\text{your} \mid \text{not spam}) = \frac{1+1}{14+28} = \frac{2}{42}$

$P(\text{free} \mid \text{notspam}) = \frac{0+1}{14+28} = \frac{1}{42}$

$P(\text{Prize} \mid \text{spam}) = \frac{0+1}{14+28} = \frac{1}{42}$

Now,

$P(\text{spam} \mid \text{claim, your, free, prize}) \propto P(\text{spam}) \times P(\text{claim} \mid \text{spam}) \times P(\text{your} \mid \text{spam}) \times P(\text{free} \mid \text{spam}) \times P(\text{Prize} \mid \text{spam})$

$\approx 0.5 \times \frac{2}{42} \times \frac{1}{42} \times \frac{2}{42} \times \frac{2}{42}$

$\approx$ ~~0.00000026~~

$\approx \frac{4}{3111696} \approx 0.0000012855$

P(notspam|claim, your, free, prize) ∝ P(notspam), P(claim|notspam) × P(your|notspam) × P(free|not spam) × P(Prize|not spam

$$\approx 0.5 \times \frac{1}{42} \times \frac{1}{42} \times \frac{2}{42} \times \frac{1}{42}$$

$$= \frac{1}{3111696}$$

$$\approx 0.0000003213$$

Since P(spam) > P(notspam), the email "claim your free prize" is spam

$$\text{Normalize} = \frac{0.0000012855}{0.0000012855 + 0.0000003213}$$

$$= \frac{0.0000012855}{0.0000016068}$$

$$= 0.800 \times 100$$

$$= 80.0\% \quad \underline{\underline{\text{spam}}}$$