ASSIGNMENT-1B

Category	Email text
Not spam	"Hi there, how are you?"
Not spam	"Meeting at 3PM 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"
?	"claim your free prize"

=> Prior probabilities

$$P(spam) = \frac{3}{6} = \frac{1}{2}$$

$$P(not\ spam) = \frac{3}{6} = \frac{1}{2}$$

1.

1.1. Total unique word in spam = 14

Total unique word in not spam = 14

Total vocabulary size = 28

• Using Laplace smoothing ("Free meeting tomorrow")

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

from formula:

$$P(w|spam) = \frac{count(w\ in\ spam)+1}{total\ word\ in\ spam + vocabulary\ size}$$

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

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

For $P(\text{spam}/\text{free}, \text{meeting}, \text{tomorrow})$

$= P(\text{spam}) \times P(\text{free}/\text{spam}) \times P(\text{meeting}/\text{spam}) \times$
$$P(\text{tomorrow}/\text{spam})$$

$$= \frac{1}{2} \times \frac{2}{42} \times \frac{1}{42} \times \frac{1}{42}$$

$$\approx 0.0000135$$

- $P(\text{free}/\text{not spam}) = \frac{0+1}{14+28} = \frac{1}{42}$

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

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

$\therefore P(\text{not spam}/\text{free}, \text{meeting}, \text{tomorrow})$

$= P(\text{not spam}) \times P(\text{free}/\text{not spam}) \times$
$$P(\text{meeting}/\text{not spam}) \times P(\text{tomorrow}/\text{not spam})$$

$$= \frac{1}{2} \times \frac{1}{42} \times \frac{2}{42} \times \frac{2}{42}$$

$$\approx 0.000027$$

Since, $P(\text{spam}/\text{free, meeting tomorrow}) <$
$\qquad P(\text{not spam}/\text{free, meeting, tomorrow})$

∴ This is "not spam".

Normalization (not spam) $= \dfrac{0.000027}{0.000027 + 0.0000135} \times 100\%$

$\approx 67.5\%$ not spam

Also,

Normalization (spam) $= \dfrac{0.0000135}{0.0000135 + 0.000027} \times 100\%$

$\approx 32.5\%$ spam

2. For email "claim, your, free, prize"

Using laplace smoothing

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

$P(\text{your}/\text{spam}) = \dfrac{1+1}{14+28} = \dfrac{2}{42}$

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

$P(\text{prize}/\text{spam}) = \dfrac{1+1}{14+28} = \dfrac{2}{42}$

Also,

$P(\text{claim}/\text{not spam}) = \dfrac{0+1}{14+28} = \dfrac{1}{42}$

$$P(\text{your}/\text{not spam}) = \frac{0+1}{14+28} = \frac{1}{42}$$

$$P(\text{free}/\text{not spam}) = \frac{0+1}{14+28} = \frac{1}{42}$$

$$P(\text{prize}/\text{not spam}) = \frac{0+1}{14+28} = \frac{1}{42}$$

Now,

$$P(\text{spam}/\text{claim, your, free, prize})$$

$$= P(\text{spam}) \times P(\text{claim}/\text{spam}) \times P(\text{your}/\text{spam}) \times P(\text{free}/\text{spam}) \times P(\text{prize}/\text{spam})$$

$$= \frac{1}{2} \times \frac{2}{42} \times \frac{2}{42} \times \frac{2}{42} \times \frac{2}{42}$$

$$\approx 0.0000025$$

And,

$$P(\text{not spam}/\text{claim, your, free, prize})$$

$$= P(\text{not spam}) \times P(\text{claim}/\text{not spam}) \times P(\text{your}/\text{not spam}) \times P(\text{free}/\text{not spam}) \times P(\text{prize}/\text{not spam})$$

$$= \frac{1}{42} \times \frac{1}{42} \times \frac{1}{42} \times \frac{1}{42} \times \frac{1}{42}$$

$$\approx 0.00000016$$

Since, P(spam/claim, your, free, prize) >
                    P(not spam/claim, your, free prize)

So, this email is "spam".

∴ Normalization (spam) = $\dfrac{0.0000026}{0.0000026 + 0.0000016} \times 100\%$

$\approx 90.5\%$ Spam

Also,

Normalization (not spam) = $\dfrac{0.0000016}{0.0000016 + 0.0000026} \times 100\%$

$\approx 5.7\%$ not spam.

Hence,

"Free meeting tomorrow" → not spam

"claim your free prize" ⟹ spam.