Date / /
Page

Assignment - 1B

category	Email text
Not Spam	"Hi there how are you?"
Not Spam	"Meeting at 3 PM tommorow"
Not Spam	"Please send the report"
Spam	"Win a free prize now!"
Spam	"Claim your discount today"
Spam	"limite time offer click here"
?	"Free meeting tomorrow"
?	"claim your free prize"

Prior probabilities

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

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

Total unique word in spam = 14

Total unique word in not spam = 14

Total vocabulary size = 28

using laplace smoothing ("Free meeting tomorrow")

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

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

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

From Formula :-

$P(w|\text{spam}) = \frac{\text{Count}(w \text{ in spam})+1}{\text{Total word in spam} + \text{Vocabulary size}}$

P(spam | free, meeting, tommrow)

= P(spam) x P(free|spam) x P(meeting|spam) x P(tommrow|spam)

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

≈ 0.0000185

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

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

P(tommrow | not spam) = $\frac{1+1}{14+28} = \frac{2}{42}$

P(not spam | free, meeting, tommrow)

= P(not spam) x P(free|not spam) x P(meeting|not spam) x P(tommrow|not spam)

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

≈ 0.000027

Since p(spam|free, meeting, tommrow) < P(not spam | free, meeting, tommrow) this email is 'not spam'

Normalization (not spam) = $\frac{0.000027}{0.000027+0.000013} \times 100\% \approx 67.5\%$ not spam

Normalization (spam) = $\frac{0.000013}{0.000013 + 0.000027} \times 100\%$

$\approx$ 32.5% spam

2 For email "claim your free prize"

Using laplace smoothing

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

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

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

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

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

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

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

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

Now,

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

$\propto P(\text{spam}) \times P(\text{claim} | \text{spam}) \times P(\text{your} | \text{spam}) \times P(\text{free} | \text{spam}) \times P(\text{prize} | \text{spam})$

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

≈ 0.0000025

P(not spam | claim, your, free, prize)

$= P(\text{not spam}) \times P(\text{not claim} | \text{not spam}) \times P(\text{your} | \text{not spam}) \times P(\text{Free} | \text{not spam}) \times P(\text{prize} | \text{not spam})$

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

$= 0.000000016$

Since $P(\overset{\text{span}}{\cancel{\text{claim}}} | \text{claim, your, free, prize}) > P(\text{not spam} | \text{claim, your, free, prize})$ so this email is "spam".

$\text{Normalization (Spam)} = \frac{0.0000026}{0.0000026 + 0.000000016}$

$\approx 90.5\%$ spam

$\text{Normalization (not spam)} = \frac{0.000000016}{0.000000016 + 0.0000026}$

$\approx 5.7\%$ not spam

∴ "Free meeting tommrow" → not spam

∴ "claim your free prize" → spam