Challenge Project 2 - Bayesian Spam Filtering

1) input: "WATCH ANIME NOW"

- P(WATCH) - P("WATCH" | SPAM) → 0.095
  - P("WATCH" | Not SPAM) → 0.053
- P(ANIME) - P("ANIME" | SPAM) → 0.095
  - P("ANIME" | Not SPAM) → 0.053
- P(NOW) - P("NOW" | SPAM) → 0.095
  P("NOW" | Not SPAM) → 0.053

- P(WATCH ANIME NOW | SPAM) = 0.095 · 0.095 · 0.095 = 0.000857375
- P(WATCH ANIME NOW | Not SPAM) = 0.053 · 0.053 · 0.053 = 0.00014887
- P(WATCH ANIME NOW | SPAM) > P(WATCH ANIME NOW | Not SPAM)
  therefore "watch anime now" is likely to be spam.

2) input: "TAKEOUT AND ANIME AT MY HOUSE"

- P(TAKEOUT) - P(TAKEOUT | SPAM) → 0.0476
  - P(TAKEOUT | Not SPAM) → 0.105
- P(AND) - P(AND | SPAM) → 0.0476
  - P(AND | Not SPAM) → 0.053
- P(ANIME) - P(ANIME | SPAM) → 0.095
  - P(ANIME | Not SPAM) → 0.053
- P(AT) - P(AT | SPAM) → 0.0476
  P(AT | Not SPAM) → 0.105
- P(MY) - P(MY | SPAM) → 0.0476
  - P(MY | Not SPAM) → 0.053
- P(HOUSE) - P(HOUSE | SPAM) → 0.095
  - P(HOUSE | Not SPAM) → 0.105

- P(TAKEOUT AND ANIME AT MY HOUSE | SPAM) = $0.0476^4 \cdot 0.095^2 = 4.63314\,e^{-8}$
- P(TAKEOUT AND ANIME AT MY HOUSE | Not SPAM) = $0.105^3 \cdot 0.053^3 = 1.72344\,e^{-7}$
- P(input | SPAM) < P(input | Not SPAM)
  therefore "takeout and anime at my house" is likely to not be spam.

3) input: "SELL ME YOUR ANIME COLLECTION"

- P(SELL) - P(SELL | SPAM) = 0.095
  - P(SELL | Not SPAM) = 0.105
- P(ME) - P(ME | SPAM) = 0.0476
  - P(ME | Not SPAM) = 0.053
- P(YOUR) - P(YOUR | SPAM) = 0.095
  - P(YOUR | Not SPAM) = 0.053
- P(ANIME) - P(ANIME | SPAM) = 0.095
  - P(ANIME | Not SPAM) = 0.053
- P(COLLECTION) - P(COLLECTION | SPAM) = 0.0476
  - P(COLLECTION | Not SPAM) = 0.053

- $P(\text{SELL ME YOUR ANIME COLLECTION} \mid \text{SPAM}) = 0.095^3 \cdot 0.0476^2 = 1.94261e^{-6}$
- $P(\text{SELL ME YOUR ANIME COLLECTION} \mid \text{Not SPAM}) = 0.053^4 \cdot 0.105 = 8.28501e^{-7}$
- $P(\text{input} \mid \text{SPAM}) < P(\text{input} \mid \text{Not SPAM})$

  therefore "sell me your anime collection" is likely to be spam.