SMS Spam Classification using Markov Chains and Bayes Theorem

Task Overview

Implement a spam classifier that combines Markov chains with Bayes' theorem to identify spam SMS messages using spam.csv dataset.

Dataset

- File name: spam.csv
- Contains two columns:
    - 'label': marks if message is 'spam' or 'ham'
    - 'message': contains the SMS text

Bayes' Theorem Application

For a message M, we calculate P(spam|M) using Bayes' theorem:

P(spam|M) = P(M|spam)P(spam) / P(M)

Breaking this down:

1. Prior Probability (in log space):

    log P(spam) = log(number of spam messages / total messages)

    log P(ham) = log(number of ham messages / total messages)

2. Likelihood P(M|spam) for message M = [$word_1$, $word_2$, ..., word□]:

    log P(M|spam) = log P($word_1$|spam) + # Initial probability

    log P($word_2$|$word_1$) + # First transition

    log P($word_3$|$word_2$) + # Second transition

    ... # Remaining transitions

$$\log p(s_{1...T}) = \log \pi_{s_1} + \sum_{t=2}^{T} \log A_{s_{t-1},s_t}$$

3. Final Classification:
    - Calculate log P(spam|M) = log P(M|spam) + log P(spam)
    - Calculate log P(ham|M) = log P(M|ham) + log P(ham)
    - Compare: if log P(spam|M) > log P(ham|M), classify as spam

Required Formulas

1. Initial State Probabilities:

$$\hat{\pi}_i = \frac{count(s_1 = i) + \varepsilon}{N + \varepsilon M}$$

2. Transition Probabilities:

$$\hat{A}_{ij} = \frac{count(i \to j) + \varepsilon}{count(i) + \varepsilon M}$$

Implementation Steps

1. Data Processing:
    - Load spam.csv
    - Split into train (80%) and test (20%)
    - Preprocess messages (lowercase, basic cleaning)
2. Training:
    - Calculate log prior probabilities P(spam) and P(ham)
    - Build vocabulary for spam and ham
    - Compute initial state probabilities
    - Compute transition probabilities
3. Prediction: For each message:
    - Calculate log P(M|spam):
        - Add log of initial probability for first word
        - Add log of transition probabilities for each word pair
        - Add log prior P(spam)
    - Calculate log P(M|ham) similarly
    - Compare and classify
4. Evaluation:
    - Calculate accuracy on test set