Mathematical Foundations of the Naive Bayes Rating Classifier

Overview

This document explains the mathematical foundation behind the rating prediction system implemented using a Naive Bayes classifier. The system assigns a rating to a review based on the likelihood of the words appearing in that rating class.

1 Naive Bayes Theorem

The Naive Bayes classifier uses Bayes' Theorem with the "naive" assumption of conditional independence between words:

$$P(r \mid w_1, w_2, \ldots, w_n) = \frac{P(r) \cdot P(w_1, w_2, \ldots, w_n \mid r)}{P(w_1, w_2, \ldots, w_n)}$$

Since $P(w_1, \ldots, w_n)$ is constant across all ratings r, we only consider the numerator for comparison:

$$P(r \mid w_1, w_2, \ldots, w_n) \propto P(r) \cdot P(w_1, w_2, \ldots, w_n \mid r)$$

2 Conditional Independence Assumption

Assuming the words are conditionally independent given the rating r, the joint probability becomes:

$$P(w_1, w_2, \ldots, w_n \mid r) = \prod_{i=1}^{n} P(w_i \mid r)$$

Hence, the posterior becomes:

$$P(r \mid w_1, w_2, \ldots, w_n) \propto P(r) \cdot \prod_{i=1}^{n} P(w_i \mid r)$$

3 Logarithmic Transformation

To prevent numerical underflow from multiplying small probabilities, we take the logarithm:

$$\log P(r \mid w_1, \ldots, w_n) \propto \log P(r) + \sum_{i=1}^{n} \log P(w_i \mid r)$$

4 Parameter Estimation

Prior Probability $P(r)$

The prior probability of a rating is calculated as:

$$P(r) = \frac{\text{Number of reviews with rating } r}{\text{Total number of reviews}}$$

Likelihood $P(w \mid r)$

The likelihood of word w given rating r is:

$$P(w \mid r) = \frac{\text{Count of word } w \text{ in reviews with rating } r + 1}{\text{Total number of words in rating } r + |V|}$$

Where:

- $|V|$ is the vocabulary size.
- $+1$ is used for Laplace smoothing.

5 Prediction

The predicted rating r^* is the one that maximizes the posterior:

$$r^* = \arg\max_r \left[\log P(r) + \sum_{i=1}^{n} \log P(w_i \mid r) \right]$$

6 Implementation Notes

- The classifier is trained independently for each category (e.g., seat comfort, food, etc.).
- Preprocessing includes:
  - Lowercasing
  - Punctuation removal
  - Stopword filtering
- For each input review, the classifier computes the log-probability of each possible rating class and selects the rating with the highest score.

7 Smoothing and Vocabulary

To ensure robustness:

- Laplace smoothing prevents zero-probability issues for unseen words.
- The vocabulary is built from all training reviews and shared across all ratings in a category.