Laplacian Smoothing

We usually compute the probability of a word given a class as follows:

$$P(\mathrm{w_i} \mid \text{class}) = \frac{\text{freq}(\mathrm{w_i}, \text{class})}{\mathrm{N_{class}}} \qquad \text{class} \in \{\text{Positive}, \text{Negative}\}$$

However, if a word does not appear in the training, then it automatically gets a probability of 0, to fix this we add smoothing as follows

$$P(\mathrm{w_i} \mid \text{class}) = \frac{\text{freq}(\mathrm{w_i}, \text{class})+1}{(\mathrm{N_{class}}+\mathrm{V})}$$

Note that we added a 1 in the numerator, and since there are V words to normalize, we add V in the denominator.

N_{class}: frequency of all words in class

V: number of unique words in vocabulary