# Naive Bayes Introduction

To build a classifier, we will first start by creating conditional probabilities given the following table:

**Positive tweets**

I am happy because I am learning NLP
I am happy, not sad.

**Negative tweets**

I am sad, I am not learning NLP
I am sad, not happy

| word | Pos | Neg |
|------|-----|-----|
| I | 3 | 3 |
| am | 3 | 3 |
| happy | 2 | 1 |
| because | 1 | 0 |
| learning | 1 | 1 |
| NLP | 1 | 1 |
| sad | 1 | 2 |
| not | 1 | 2 |
| $N_{class}$ | 13 | 12 |

This allows us compute the following table of probabilities:

| word | Pos | Neg |
|------|-----|-----|
| I | 0.24 | 0.25 |
| am | 0.24 | 0.25 |
| happy | 0.15 | 0.08 |
| because | 0.08 | 0 |
| learning | 0.08 | 0.08 |
| NLP | 0.08 | 0.08 |
| sad | 0.08 | 0.17 |
| not | 0.08 | 0.17 |

Once you have the probabilities, you can compute the likelihood score as follows

Tweet: I am happy today; I am learning.

$$\prod_{i=1}^{m} \frac{P(w_i|pos)}{P(w_i|neg)} = \frac{0.14}{0.10} = 1.4 \quad \textbf{> 1}$$

$$\frac{0.20}{0.20} * \frac{0.20}{0.20} * \boxed{\frac{0.14}{0.10}} * \frac{0.20}{0.20} * \frac{0.20}{0.20} * \frac{0.10}{0.10}$$

| word | Pos | Neg |
|------|-----|-----|
| I | 0.20 | 0.20 |
| am | 0.20 | 0.20 |
| happy | 0.14 | 0.10 |
| because | 0.10 | 0.05 |
| learning | 0.10 | 0.10 |
| NLP | 0.10 | 0.10 |
| sad | 0.10 | 0.15 |
| not | 0.10 | 0.15 |

A score greater than 1 indicates that the class is positive, otherwise it is negative.

✓ Complete    Go to next item