

## Example of Naïve Bayes

Tweet	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Label
	Awesome	Great	Good	Hate	Dislike	Boring	
I.	2	0	0	0	0	0	Positive
II.	1	0	0	0	0	0	Positive
III.	0	1	0	0	0	0	Positive
IV.	0	0	1	0	0	0	Positive
V.	0	0	0	1	0	0	Negative
VI.	0	0	0	0	1	0	Negative
VII.	0	0	0	0	0	1	Negative

Total number of labels = 7

Total number of positive = 4

Total number of negative = 3

$$P(\text{positive}) = \frac{4}{7} = 0.57$$

$$P\left(\frac{\text{awesome}}{\text{positive}}\right) = \frac{3}{4} = 0.75$$

$$P\left(\frac{\text{great}}{\text{positive}}\right) = \frac{1}{4} = 0.25$$

$$P\left(\frac{\text{good}}{\text{positive}}\right) = \frac{1}{4} = 0.25$$

$$P\left(\frac{\text{hate}}{\text{positive}}\right) = \frac{0}{4} = 0 + 1 = 1$$

$$P\left(\frac{\text{dislike}}{\text{positive}}\right) = \frac{0}{4} = 0 + 1 = 1$$

$$P\left(\frac{\text{boring}}{\text{positive}}\right) = \frac{0}{4} = 0 + 1 = 1$$

adding 1 is smoothing technique

$$P(\text{negative}) = \frac{3}{7} = 0.42$$

$$P\left(\frac{\text{awesome}}{\text{negative}}\right) = \frac{0}{3} = 0 + 1 = 1$$

$$P\left(\frac{\text{great}}{\text{negative}}\right) = \frac{0}{3} = 0 + 1 = 1$$

$$P\left(\frac{\text{good}}{\text{negative}}\right) = \frac{0}{3} = 0 + 1 = 1$$

$$P\left(\frac{\text{hate}}{\text{negative}}\right) = \frac{1}{3} = 0.33$$

$$P\left(\frac{\text{dislike}}{\text{negative}}\right) = \frac{1}{3} = 0.33$$

$$P\left(\frac{\text{boring}}{\text{negative}}\right) = \frac{1}{3} = 0.33$$

New tweet: The movie was awesome but the graphics were boring.

Formula

$$P\left(\frac{y}{x_1, \dots, x_n}\right) = P(y) \prod_{i=1}^n P\left(\frac{x_i}{y}\right)$$

Now,

$$\begin{aligned}
 P\left(\frac{\text{positive}}{\text{awesome, boring}}\right) &= P(\text{positive}) P\left(\frac{\text{awesome}}{\text{positive}}\right) P\left(\frac{\text{boring}}{\text{positive}}\right) \\
 &= (0.57)(0.75)(1) \\
 &= 0.4275
 \end{aligned}$$

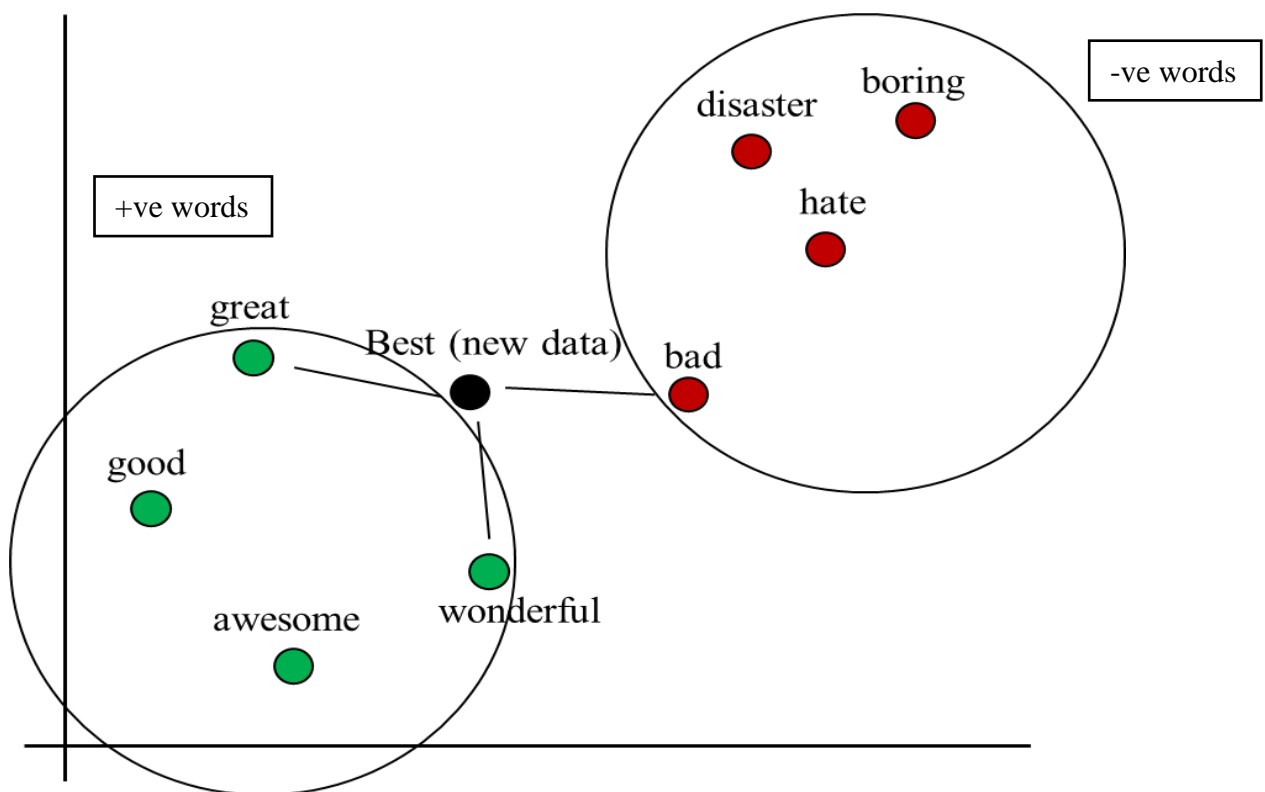
$$\begin{aligned}
 P\left(\frac{negative}{awesome,boring}\right) &= P(negative)P\left(\frac{awesome}{negative}\right)P\left(\frac{boring}{negative}\right) \\
 &= (0.42)(1)(0.33) \\
 &= 0.1386
 \end{aligned}$$

Comparing both  $P\left(\frac{positive}{awesome,boring}\right)$  &  $P\left(\frac{negative}{awesome,boring}\right)$

So,  $P\left(\frac{positive}{awesome,boring}\right)$  has more probability

Therefore, new tweet is positive

## Example of KNN



We choose  $k = 3$

For plotting these points (features) on graph, there is mathematical technique for that.

Now check which has more accuracy out of these 3. The new data point belongs to that category which has more accuracy.