On the previous screen, we introduced these two equations without much explanation:

$$P(Spam|w_1, w_2, w_3, w_4) \propto P(Spam) \cdot P(w_1|Spam) \cdot P(w_2|Spam) \cdot$$
$$P(Spam^C|w_1, w_2, w_3, w_4) \propto P(Spam^C) \cdot P(w_1|Spam^C) \cdot P(w_2|Spam^C)$$

To explain the mathematics behind these equations, let's start by looking at P(Spam|w_1, w_2, w_3, w_4). Using the conditional probability formula, we can expand P(Spam|w_1, w_2, w_3, w_4) like this (below, make sure you notice the $\cap$ symbol in the numerator):

$$P(Spam|w_1, w_2, w_3, w_4) = \frac{P(Spam \cap (w_1, w_2, w_3, w_4))}{P(w_1, w_2, w_3, w_4)}$$

Recall that we learned in a previous screen that we can ignore the division, which means we can drop P(w_1, w_2, w_3, w_4) to avoid redundant calculations (when we ignore the division, we also replace the equals sign with $\propto$, which means directly proportional):

$$P(Spam|w_1, w_2, w_3, w_4) \propto P(Spam \cap (w_1, w_2, w_3, w_4))$$

Note that (w_1, w_2, w_3, w_4) can be modeled as an intersection of four events:

$$w_1, w_2, w_3, w_4 = w_1 \cap w_2 \cap w_3 \cap w_4$$

For instance, we could think of a message like "thanks for your help" as the intersection of four words inside a single message: "thanks", "for, "your", and "help". In probability jargon, finding the value of $P(w_1 \cap w_2 \cap w_3 \cap w_4)$ means finding the probability that the four words w_1, w_2, w_3, w_4 occur together

With this in mind, our equation above transforms to:

$$P(Spam|w_1, w_2, w_3, w_4) \propto P(Spam \cap \underbrace{(w_1 \cap w_2 \cap w_3 \cap w_4)}_{(w_1, w_2, w_3, w_4)})$$

From set theory, we know that $A \cap (B \cap C) = A \cap B \cap C = C \cap B \cap A$, which means we can transform $P(Spam \cap (w_1 \cap w_2 \cap w_3 \cap w_4))$ in our equation above to make it suitable for further expansion:

$$\begin{aligned} P(Spam \cap (w_1 \cap w_2 \cap w_3 \cap w_4)) &= P(Spam \cap w_1 \cap w_2 \cap w_3 \cap w_4) \\ &= P(w_1 \cap w_2 \cap w_3 \cap w_4 \cap Spam) \end{aligned}$$

Now let's use the multiplication rule to expand $P(w_1 \cap w_2 \cap w_3 \cap w_4 \cap Spam)$:

$$P(w_1 \cap w_2 \cap w_3 \cap w_4 \cap Spam) = P(w_1|w_2 \cap w_3 \cap w_4 \cap Spam) \cdot P(w_2 \cap w_3 \cap w_4 \cap Spam)$$

We can use the multiplication rule again to expand $P(w_2 \cap w_3 \cap w_4 \cap Spam)$:

$$P(w_1 \cap w_2 \cap w_3 \cap w_4 \cap Spam) = P(w_1|w_2 \cap w_3 \cap w_4 \cap Spam) \cdot \underbrace{P(w_2|w_3 \cap w_4 \cap Spam) \cdot P(w_3 \cap w_4}_{P(w_2 \cap w_3 \cap w_4 \cap Spam)} \cap Spam)$$

We can use the multiplication rule successively, until there's nothing more left to expand:

$$P(w_1 \cap w_2 \cap w_3 \cap w_4 \cap Spam) = P(w_1|w_2 \cap w_3 \cap w_4 \cap Spam) \cdot P(w_2 \cap w_3 \cap w_4 \cap Spam)$$

Learn

We can use the multiplication rule successively, until there's nothing more left to expand:

$$
\begin{aligned}
P(w_1 \cap w_2 \cap w_3 \cap w_4 \cap Spam) &= P(w_1|w_2 \cap w_3 \cap w_4 \cap Spam) \cdot P(w_2 \cap w_3 \cap w_4 \cap Spam) \\
&= P(w_1|w_2 \cap w_3 \cap w_4 \cap Spam) \cdot P(w_2|w_3 \cap w_4 \cap Spam) \cdot P(w_3 \cap w_4 \cap Spam) \\
&= P(w_1|w_2 \cap w_3 \cap w_4 \cap Spam) \cdot P(w_2|w_3 \cap w_4 \cap Spam) \cdot P(w_3|w_4 \cap Spam) \cdot P(w_4 \cap Spam) \\
&= P(w_1|w_2 \cap w_3 \cap w_4 \cap Spam) \cdot P(w_2|w_3 \cap w_4 \cap Spam) \cdot P(w_3|w_4 \cap Spam) \cdot P(w_4|Spam) \cdot P(Spam)
\end{aligned}
$$

In theory, the last equation you see above is what we'd have to use if we wanted to calculate P(Spam|w_1, w_2, w_3, w_4). However, the equation is pretty long for just four words. Also, imagine how would the equation look for a 50-word message — just think of how many calculations we'd have to perform!!

To make the calculations tractable for messages of all kinds of lengths, we can assume conditional independence between w_1, w_2, w_3, and w_4. This implies that:

$$
\begin{aligned}
P(w_1|w_2 \cap w_3 \cap w_4 \cap Spam) &= P(w_1|Spam) \\
P(w_2|w_3 \cap w_4 \cap Spam) &= P(w_2|Spam) \\
P(w_3|w_4 \cap Spam) &= P(w_3|Spam) \\
P(w_4|Spam) &= P(w_4|Spam)
\end{aligned}
$$

Under the assumption of independence, our lengthy equation above reduces to:

$$P(w_1 \cap w_2 \cap w_3 \cap w_4 \cap Spam) = P(w_1|Spam) \cdot P(w_2|Spam) \cdot P(w_3|Spam) \cdot P(w_4|Spam) \cdot P(Spam$$

$$P(Spam|w_1, w_2, w_3, w_4) \propto P(Spam) \cdot P(w_1|Spam) \cdot P(w_2|Spam) \cdot P(w_3|Spam) \cdot P(w_4|Spam)$$
$$P(Spam^C|w_1, w_2, w_3, w_4) \propto P(Spam^C) \cdot P(w_1|Spam^C) \cdot P(w_2|Spam^C) \cdot P(w_3|Spam^C) \cdot P(w_4|Spam^C)$$

Learn

To make the calculations tractable for messages of all kinds of lengths, we can assume conditional independence between w_1, w_2, w_3, and w_4. This implies that:

$$P(w_1|w_2 \cap w_3 \cap w_4 \cap Spam) = P(w_1|Spam)$$
$$P(w_2|w_3 \cap w_4 \cap Spam) = P(w_2|Spam)$$
$$P(w_3|w_4 \cap Spam) = P(w_3|Spam)$$
$$P(w_4|Spam) = P(w_4|Spam)$$

Under the assumption of independence, our lengthy equation above reduces to:

$$P(w_1 \cap w_2 \cap w_3 \cap w_4 \cap Spam) = P(w_1|Spam) \cdot P(w_2|Spam) \cdot P(w_3|Spam) \cdot P(w_4|Spam) \cdot P(Span$$

The assumption of conditional independence is unrealistic in practice because words are often in a relationship of dependence. For instance, if you see the word "WINNER" in a message, the probability of seeing the word "money" is very likely to increase, so "WINNER" and "money" are most likely dependent. The assumption of conditional independence between words is thus *naive* since it rarely holds in practice, and this is why the algorithm is called **Naive** Bayes (also called **simple Bayes** or **independence Bayes**).

Despite this simplifying assumption, the algorithm works quite well in many real-word situations, and we'll see that ourselves in the guided project.

That being said, on the previous screen we assumed conditional independence when we introduced these two euquation

$$P(Spam|w_1, w_2, w_3, w_4) \propto P(Spam) \cdot P(w_1|Spam) \cdot P(w_2|Spam) \cdot P(w_3|Spam) \cdot P(w_4|Spam)$$
$$P(Spam^C|w_1, w_2, w_3, w_4) \propto P(Spam^C) \cdot P(w_1|Spam^C) \cdot P(w_2|Spam^C) \cdot P(w_3|Spam^C) \cdot P(w_4|Spam^C)$$