CSE 312 Section 3
The Naive Bayes Classifier

Made by Luxi Wang, Pemi Nguyen, Mitchell Estberg and Shreya Jayaraman
Alex Tsun
Announcements

- Pset2 due yesterday
- Pset3 Due Wednesday, October 20st 11:59 PM PDT
Agenda

- What is Machine Learning?
- Featurizing Emails
- Naive Bayes
Machine Learning in the Real World
From Wikipedia: “Machine learning is the study of computer algorithms that improve automatically through experience.”
You are a machine!

<table>
<thead>
<tr>
<th>Number</th>
<th>Shape</th>
<th>“Label”</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>🟢</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>🟠</td>
<td>15</td>
</tr>
<tr>
<td>-2</td>
<td>🟢</td>
<td>-8</td>
</tr>
<tr>
<td>7</td>
<td>🟠</td>
<td>21</td>
</tr>
<tr>
<td>-4</td>
<td>🟢</td>
<td>???</td>
</tr>
</tbody>
</table>

Given examples with correct “labels”, make predictions!
You are a machine!

<table>
<thead>
<tr>
<th>Number</th>
<th>Shape</th>
<th>“Label”</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>▲</td>
<td>15</td>
</tr>
<tr>
<td>-2</td>
<td>□</td>
<td>-8</td>
</tr>
<tr>
<td>7</td>
<td>▲</td>
<td>21</td>
</tr>
<tr>
<td>-4</td>
<td>□</td>
<td>-16</td>
</tr>
</tbody>
</table>

Given examples with correct “labels”, make predictions!
Regression: Idea

$340,135

$801,353
Classification: Idea

“Green” class
“Red” class
Is this new shape supposed to be "green" or "red"?

"Green" class

"Red" class
Spam Filter

- In real life, you may have seen a lot of spam emails like this.
- Building a good spam filter helps protect users from potential scams, unnecessary advertising, or malware links.
# Evaluating Performance

<table>
<thead>
<tr>
<th>Training Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Email</strong></td>
<td><strong>Email</strong></td>
</tr>
<tr>
<td>Buy Viagra!</td>
<td>You buy viagra!</td>
</tr>
<tr>
<td>You good?</td>
<td>You need viagra sir.</td>
</tr>
<tr>
<td>Viagra help you.</td>
<td>I hope you are healthy.</td>
</tr>
</tbody>
</table>
| Good Viagra help. | ... | Spam | ...
| I need Viagra for my health condition. | ... | Ham | ...

We “**train**” our spam filter on the training set, and **evaluate** performance using a test set (data that is unseen by the spam filter initially). This gives an unbiased estimate of performance.
Spam Filter Task

Training Set

<table>
<thead>
<tr>
<th>Email</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy Viagra!</td>
<td>Spam</td>
</tr>
<tr>
<td>You good?</td>
<td>Ham</td>
</tr>
<tr>
<td>Viagra help you.</td>
<td>Spam</td>
</tr>
<tr>
<td>Good Viagra help.</td>
<td>Spam</td>
</tr>
<tr>
<td>I need Viagra for my health condition.</td>
<td>Ham</td>
</tr>
</tbody>
</table>

Predict whether this email is spam or ham:

You buy Viagra!
**Emails as word collections**

<table>
<thead>
<tr>
<th>Email</th>
<th>Set of Words in the Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUBJECT: Top Secret Business Venture</td>
<td>{top, secret, business, venture, dear, sir, first, I, must, solicit, your, confidence, in, this, transaction, is, by, virtue, of, its, nature, as, being, utterly, confidential, and}</td>
</tr>
<tr>
<td>Dear Sir. First, I must solicit your confidence in this transaction, this is by virtue of its nature as being utterly confidential and top secret…</td>
<td></td>
</tr>
</tbody>
</table>

For simplicity, we will
- Ignore Duplicate Words
- Ignore Punctuation
- Ignore Casing
## Emails as word collections

<table>
<thead>
<tr>
<th>Email</th>
<th>Set of Words in the Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUBJECT: Top Secret Business Venture</td>
<td>{top, secret, business, venture, dear, sir, first, I, must, solicit, your, confidence, in, this, transaction, is, by, virtue, of, its, nature, as, being, utterly, confidencial, and}</td>
</tr>
<tr>
<td>Dear Sir. First, I must solicit your confidence in this transaction, this is by virtue of its nature as being utterly confidential and top secret…</td>
<td></td>
</tr>
<tr>
<td>Hello hello hello there.</td>
<td>{hello, there}</td>
</tr>
</tbody>
</table>
### Emails as word collections

<table>
<thead>
<tr>
<th>Email</th>
<th>Set of Words in the Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUBJECT: Top Secret Business Venture</td>
<td>{top, secret, business, venture, dear, sir, first, I, must, solicit, your, confidence, in, this, transaction, is, by, virtue, of, its, nature, as, being, utterly, confidential, and}</td>
</tr>
<tr>
<td>Dear Sir. First, I must solicit your confidence in this transaction, this is by virtue of its nature as being utterly confidential and top secret…</td>
<td></td>
</tr>
<tr>
<td>Hello hello hello there.</td>
<td>{hello, there}</td>
</tr>
<tr>
<td>You buy Viagra!</td>
<td>{you, buy, viagra}</td>
</tr>
</tbody>
</table>

For simplicity, we will
- Ignore Duplicate Words
- Ignore Punctuation
- Ignore Casing
Our approach

Compute and Compare:

\[ \mathbb{P}(\text{spam} \mid "\text{You buy Viagra!}" ) \quad \mathbb{P}(\text{ham} \mid "\text{You buy Viagra!}" ) \]

Then predict whichever is larger! Can we get away with just computing one of them?
Our approach

Compute and Compare:

$$P(\text{spam} \mid \text{"You buy Viagra!"}) \quad P(\text{ham} \mid \text{"You buy Viagra!"})$$

Then predict whichever is larger! Can we get away with just computing one of them?

Equivalently, note that these add to 1, so we can just compute $$P(\text{spam} \mid \text{"You buy Viagra!"})$$

and if it is greater than 0.5, then we predict spam.

Otherwise, we predict ham.

Note: We resolve the tie in favor of ham.
Naive Bayes Classifier - The bayes part

Bayes Theorem:

\[ P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)} \]

Apply it to our example:

\[ P(\text{spam} \mid \text{"You buy Viagra!"}) = \frac{P(\text{"You buy Viagra!"} \mid \text{spam}) P(\text{spam})}{P(\text{"You buy Viagra!"})} \]
Naive Bayes Classifier - What we Calculate

\[
P(\text{spam} \mid "You buy Viagra!") = \frac{P("You buy Viagra!" \mid \text{spam}) P(\text{spam})}{P("You buy Viagra!")}
\]
Naive Bayes Classifier - What we Calculate

\[
P(\text{spam} \mid \text{"You buy Viagra!"}) = \frac{P(\text{"You buy Viagra!"} \mid \text{spam}) P(\text{spam})}{P(\text{"You buy Viagra!"})}
\]

\[
= \frac{P(\{\"you\",\"buy\",\"viagra\}\} \mid \text{spam}) P(\text{spam})}{P(\{\"you\",\"buy\",\"viagra\}\} \mid \text{spam}) P(\text{spam}) + P(\{\"you\",\"buy\",\"viagra\}\} \mid \text{ham}) P(\text{ham})}
\]

[LTP]
Naive Bayes Classifier - What we Calculate

\[
P(\text{spam} \mid \text{"You buy Viagra!"}) = \frac{P(\text{"You buy Viagra!"} \mid \text{spam}) P(\text{spam})}{P(\text{"You buy Viagra!"})}
\]

\[
= \frac{P(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{spam}) P(\text{spam})}{P(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{spam}) P(\text{spam}) + P(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{ham}) P(\text{ham})}
\]

\[
P(\text{spam}) = \frac{\text{total spam emails (in training set)}}{\text{total emails (in training set)}}
\]

\[
P(\text{ham}) = \frac{\text{total ham emails (in training set)}}{\text{total emails (in training set)}}
\]

(our approximation for these probabilities, based on the training set)
Naive Bayes Classifier - The naive part

It is somewhat unlikely that we have the email "You buy Viagra!" in our training data. (In this case we don't!)
Naive Bayes Classifier - The naive part

It is somewhat unlikely that we have the email "You buy Viagra!" in our training data. (In this case we don’t!)

We naively assume that words are conditionally independent from each other, given the label (In reality, they aren’t):
Naive Bayes Classifier - The naive part

It is somewhat unlikely that we have the email "You buy Viagra!" in our training data. (In this case we don’t!)

We naively assume that words are conditionally independent from each other, given the label (In reality, they aren’t):

\[
P\left(\{\text{“you”}, \text{“buy”}, \text{“viagra”}\} \mid \text{spam}\right) \\
\approx P(\text{“you”} \mid \text{spam})P(\text{“buy”} \mid \text{spam})P(\text{“viagra”} \mid \text{spam})
\]
Naive Bayes Classifier - The naive part

It is somewhat unlikely that we have the email ”You buy Viagra!” in our training data. (In this case we don’t!)

We naively assume that words are conditionally independent from each other, given the label (In reality, they aren’t):

\[
P(\{"you", "buy", "viagra"\} \mid \text{spam}) \\
\approx P("you" \mid \text{spam})P("buy" \mid \text{spam})P("viagra" \mid \text{spam})
\]

Then we estimate for example that

\[
P("you" \mid \text{spam}) = \frac{\text{number of spam emails containing "you" (in training set)}}{\text{number of spam emails (in training set)}}
\]
Why is this Naive?

Consider for example the following two emails:

“!!!Lunch free for You!!!!!”

“You free for lunch?”
Why is this Naive?

Consider for example the following two emails:

“!!!Lunch free for You!!!!!”

“You free for lunch?”

One shortfalling of our model is that it will make the same prediction for these since they have the same set of words!
\[ P(\text{spam} \mid \text{"You buy Viagra"}) = \frac{P(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{spam}) \cdot P(\text{spam})}{P(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{ham}) \cdot P(\text{ham})} \]

\[ = \frac{P(\text{"you"} \mid \text{spam}) \cdot P(\text{"buy"} \mid \text{spam}) \cdot P(\text{"viagra"} \mid \text{spam}) \cdot P(\text{spam})}{P(\text{"you"} \mid \text{spam}) \cdot P(\text{"buy"} \mid \text{spam}) \cdot P(\text{"viagra"} \mid \text{spam}) \cdot P(\text{spam}) + P(\text{"you"} \mid \text{ham}) \cdot P(\text{"buy"} \mid \text{ham}) \cdot P(\text{"viagra"} \mid \text{ham}) \cdot P(\text{ham})} \]

<table>
<thead>
<tr>
<th>Email</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy Viagra!</td>
<td>Spam</td>
</tr>
<tr>
<td>You good?</td>
<td>Ham</td>
</tr>
<tr>
<td>Viagra help you.</td>
<td>Spam</td>
</tr>
<tr>
<td>Good Viagra help.</td>
<td>Spam</td>
</tr>
<tr>
<td>I need Viagra for my health condition.</td>
<td>Ham</td>
</tr>
</tbody>
</table>

\[ P(\text{spam}) = P(\text{ham}) = \]

\[ P(\text{"you"} \mid \text{spam}) = P(\text{"you"} \mid \text{ham}) = \]

\[ P(\text{"buy"} \mid \text{spam}) = P(\text{"buy"} \mid \text{ham}) = \]

\[ P(\text{"viagra"} \mid \text{spam}) = P(\text{"viagra"} \mid \text{ham}) = \]
Example

\[
P(\text{spam} \mid \text{"You buy Viagra"}) = \frac{P(\{\text{"you"},\text{"buy"},\text{"viagra"}\} \mid \text{spam}) \cdot P(\text{spam})}{P(\{\text{"you"},\text{"buy"},\text{"viagra"}\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{\text{"you"},\text{"buy"},\text{"viagra"}\} \mid \text{ham}) \cdot P(\text{ham})}
\]

\[
\begin{align*}
&= \frac{P(\text{"you"} \mid \text{spam}) \cdot P(\text{"buy"} \mid \text{spam}) \cdot P(\text{"viagra"} \mid \text{spam}) \cdot P(\text{spam})}{P(\text{"you"} \mid \text{spam}) \cdot P(\text{"buy"} \mid \text{spam}) \cdot P(\text{"viagra"} \mid \text{spam}) \cdot P(\text{spam}) + P(\text{"you"} \mid \text{ham}) \cdot P(\text{"buy"} \mid \text{ham}) \cdot P(\text{"viagra"} \mid \text{ham}) \cdot P(\text{ham})}
\end{align*}
\]

<table>
<thead>
<tr>
<th>Email</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy Viagra!</td>
<td>Spam</td>
</tr>
<tr>
<td>You good?</td>
<td>Ham</td>
</tr>
<tr>
<td>Viagra help you.</td>
<td>Spam</td>
</tr>
<tr>
<td>Good Viagra help.</td>
<td>Spam</td>
</tr>
<tr>
<td>I need Viagra for my health condition.</td>
<td>Ham</td>
</tr>
</tbody>
</table>

\[
P(\text{spam}) = \frac{3}{5} \quad P(\text{ham}) = \frac{2}{5}
\]

\[
\begin{align*}
P(\text{"you"} \mid \text{spam}) &= \frac{1}{3} \quad P(\text{"you"} \mid \text{ham}) = \frac{1}{2} \\
P(\text{"buy"} \mid \text{spam}) &= \text{ed} \quad P(\text{"buy"} \mid \text{ham}) = \text{ed} \\
P(\text{"viagra"} \mid \text{spam}) &= \text{ed} \quad P(\text{"viagra"} \mid \text{ham}) = \text{ed}
\end{align*}
\]
\[ \mathbb{P}(\text{spam} \mid \text{"You buy Viagra"}) = \frac{\mathbb{P}(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{spam}) \mathbb{P}(\text{spam})}{\mathbb{P}(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{spam}) \mathbb{P}(\text{spam}) + \mathbb{P}(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{ham}) \mathbb{P}(\text{ham})} \]

\begin{array}{|c|c|}
\hline
\text{Email} & \text{Label} \\
\hline
\text{Buy Viagra!} & \text{Spam} \\
\hline
\text{You good?} & \text{Ham} \\
\hline
\text{Viagra help you.} & \text{Spam} \\
\hline
\text{Good Viagra help.} & \text{Spam} \\
\hline
\text{I need Viagra for my health condition.} & \text{Ham} \\
\hline
\end{array}

\[ \mathbb{P}(\text{spam}) = \frac{3}{5} \quad \mathbb{P}(\text{ham}) = \frac{2}{5} \]

\[ \mathbb{P}(\text{"you"} \mid \text{spam}) = \frac{1}{3} \quad \mathbb{P}(\text{"you"} \mid \text{ham}) = \frac{1}{2} \]

\[ \mathbb{P}(\text{"buy"} \mid \text{spam}) = \frac{1}{3} \quad \mathbb{P}(\text{"buy"} \mid \text{ham}) = 0 \]

\[ \mathbb{P}(\text{"viagra"} \mid \text{spam}) = 1 \quad \mathbb{P}(\text{"viagra"} \mid \text{ham}) = \frac{1}{2} \]
\[ P(\text{spam} \mid \text{"You buy Viagra"}) = \frac{P(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{spam}) \cdot P(\text{spam})}{P(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{ham}) \cdot P(\text{ham})} \]

\[
= \frac{P(\text{"you"} \mid \text{spam}) \cdot P(\text{"buy"} \mid \text{spam}) \cdot P(\text{"viagra"} \mid \text{spam}) \cdot P(\text{spam})}{P(\text{"you"} \mid \text{spam}) \cdot P(\text{"buy"} \mid \text{spam}) \cdot P(\text{"viagra"} \mid \text{spam}) \cdot P(\text{spam}) + P(\text{"you"} \mid \text{ham}) \cdot P(\text{"buy"} \mid \text{ham}) \cdot P(\text{"viagra"} \mid \text{ham}) \cdot P(\text{ham})}
\]

<table>
<thead>
<tr>
<th>Email</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy Viagra!</td>
<td>Spam</td>
</tr>
<tr>
<td>You good?</td>
<td>Ham</td>
</tr>
<tr>
<td>Viagra help you.</td>
<td>Spam</td>
</tr>
<tr>
<td>Good Viagra help.</td>
<td>Spam</td>
</tr>
<tr>
<td>I need Viagra for my health condition.</td>
<td>Ham</td>
</tr>
</tbody>
</table>

\[ P(\text{spam}) = \frac{3}{5} \quad P(\text{ham}) = \frac{2}{5} \]

\[ P(\text{"you"} \mid \text{spam}) = \frac{1}{3} \quad P(\text{"you"} \mid \text{ham}) = \frac{1}{2} \]

\[ P(\text{"buy"} \mid \text{spam}) = \frac{1}{3} \quad P(\text{"buy"} \mid \text{ham}) = 0 \]

\[ P(\text{"viagra"} \mid \text{spam}) = 1 \quad P(\text{"viagra"} \mid \text{ham}) = \frac{1}{2} \]
\[ P(\text{spam} \mid \text{"You buy Viagra"}) \]

\[
\frac{P(\{\text{"you", "buy", "viagra"}\} \mid \text{spam}) \cdot P(\text{spam})}{P(\{\text{"you", "buy", "viagra"}\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{\text{"you", "buy", "viagra"}\} \mid \text{ham}) \cdot P(\text{ham})}
\]

\[
= \frac{P(\text{"you"} \mid \text{spam}) \cdot P(\text{"buy"} \mid \text{spam}) \cdot P(\text{"viagra"} \mid \text{spam}) \cdot P(\text{spam})}{P(\text{"you"} \mid \text{spam}) \cdot P(\text{"buy"} \mid \text{spam}) \cdot P(\text{"viagra"} \mid \text{spam}) \cdot P(\text{spam}) + P(\text{"you"} \mid \text{ham}) \cdot P(\text{"buy"} \mid \text{ham}) \cdot P(\text{"viagra"} \mid \text{ham}) \cdot P(\text{ham})}
\]

<table>
<thead>
<tr>
<th>Email</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy Viagra!</td>
<td>Spam</td>
</tr>
<tr>
<td>You good?</td>
<td>Ham</td>
</tr>
<tr>
<td>Viagra help you.</td>
<td>Spam</td>
</tr>
<tr>
<td>Good Viagra help.</td>
<td>Spam</td>
</tr>
<tr>
<td>I need Viagra for my health condition.</td>
<td>Ham</td>
</tr>
</tbody>
</table>

\[ P(\text{spam}) = \frac{3}{5} \quad P(\text{ham}) = \frac{2}{5} \]

\[ P(\text{"you"} \mid \text{spam}) = \frac{1}{3} \quad P(\text{"you"} \mid \text{ham}) = \frac{1}{2} \]

\[ P(\text{"buy"} \mid \text{spam}) = \frac{1}{3} \quad P(\text{"buy"} \mid \text{ham}) = 0 \]

\[ P(\text{"viagra"} \mid \text{spam}) = 1 \quad P(\text{"viagra"} \mid \text{ham}) = \frac{1}{2} \]
Example

\[
P(\text{spam} \mid \text{"You buy Viagra"}) = \frac{P(\{"you","buy","viagra"\} \mid \text{spam}) \cdot P(\text{spam})}{P(\{"you","buy","viagra"\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{"you","buy","viagra"\} \mid \text{ham}) \cdot P(\text{ham})}
\]

\[
= \frac{P(\text{"you"} \mid \text{spam}) \cdot P(\text{"buy"} \mid \text{spam}) \cdot P(\text{"viagra"} \mid \text{spam}) \cdot P(\text{spam})}{P(\text{"you"} \mid \text{spam}) \cdot P(\text{"buy"} \mid \text{spam}) \cdot P(\text{"viagra"} \mid \text{spam}) \cdot P(\text{spam}) + P(\text{"you"} \mid \text{ham}) \cdot P(\text{"buy"} \mid \text{ham}) \cdot P(\text{"viagra"} \mid \text{ham}) \cdot P(\text{ham})}
\]

\[
= 1
\]

<table>
<thead>
<tr>
<th>Email</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy Viagra!</td>
<td>Spam</td>
</tr>
<tr>
<td>You good?</td>
<td>Ham</td>
</tr>
<tr>
<td>Viagra help you.</td>
<td>Spam</td>
</tr>
<tr>
<td>Good Viagra help.</td>
<td>Spam</td>
</tr>
<tr>
<td>I need Viagra for my health condition.</td>
<td>Ham</td>
</tr>
</tbody>
</table>

\[
P(\text{spam}) = \frac{3}{5} \quad P(\text{ham}) = \frac{2}{5}
\]

\[
P(\text{"you"} \mid \text{spam}) = \frac{1}{3} \quad P(\text{"you"} \mid \text{ham}) = \frac{1}{2}
\]

\[
P(\text{"buy"} \mid \text{spam}) = \frac{1}{3} \quad P(\text{"buy"} \mid \text{ham}) = 0
\]

\[
P(\text{"viagra"} \mid \text{spam}) = 1 \quad P(\text{"viagra"} \mid \text{ham}) = \frac{1}{2}
\]
\[
P(\text{spam} \mid \text{"You buy Viagra"}) = \frac{P(\{"you","buy","viagra"\} \mid \text{spam}) \cdot P(\text{spam})}{P(\{"you","buy","viagra"\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{"you","buy","viagra"\} \mid \text{ham}) \cdot P(\text{ham})}
\]

\[
= \frac{P(\text{"you"} \mid \text{spam})P(\text{"buy"} \mid \text{spam})P(\text{"viagra"} \mid \text{spam})P(\text{spam})}{P(\text{"you"} \mid \text{spam})P(\text{"buy"} \mid \text{spam})P(\text{"viagra"} \mid \text{spam})P(\text{spam}) + P(\text{"you"} \mid \text{ham})P(\text{"buy"} \mid \text{ham})P(\text{"viagra"} \mid \text{ham})P(\text{ham})}
\]

\[
= 1 \quad \text{(Marked as spam since no ham email contained "buy")}
\]

<table>
<thead>
<tr>
<th>Email</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy Viagra!</td>
<td>Spam</td>
</tr>
<tr>
<td>You good?</td>
<td>Ham</td>
</tr>
<tr>
<td>Viagra help you.</td>
<td>Spam</td>
</tr>
<tr>
<td>Good Viagra help.</td>
<td>Spam</td>
</tr>
<tr>
<td>I need Viagra for my health condition.</td>
<td>Ham</td>
</tr>
</tbody>
</table>

\[
P(\text{spam}) = \frac{3}{5} \quad P(\text{ham}) = \frac{2}{5}
\]

\[
P(\text{"you"} \mid \text{spam}) = \frac{1}{3} \quad P(\text{"you"} \mid \text{ham}) = \frac{1}{2}
\]

\[
P(\text{"buy"} \mid \text{spam}) = \frac{1}{3} \quad P(\text{"buy"} \mid \text{ham}) = 0
\]

\[
P(\text{"viagra"} \mid \text{spam}) = 1 \quad P(\text{"viagra"} \mid \text{ham}) = \frac{1}{2}
\]
What happens if we got a 0?

\[ P(\text{ham} \mid \text{“You buy Viagra!”}) = 0 \text{ since } P(\text{“buy”} \mid \text{ham}) = 0, \text{ since no ham email in our training data contained the word ‘buy’}. \]

But does that mean we will never encounter a ham email with word ‘buy’?

What about the ham: “I’ll buy sunflowers”
Laplace smoothing

Pretend in spam emails (training set):

- We saw one extra spam email with word $w_i$
- We saw one extra spam email without word $w_i$
Laplace smoothing

Pretend in spam emails (training set):

- We saw one extra spam email with word $w_i$
- We saw one extra spam email without word $w_i$

$$
\mathbb{P}(w_i \mid \text{spam}) = \frac{|\text{total spam emails (training set) containing } w_i| + 1}{|\text{total spam emails (training set)}| + 2}
$$
Laplace smoothing

Pretend in spam emails (training set):

- We saw one extra spam email with word $w_i$
- We saw one extra spam email without word $w_i$

$$\mathbb{P}(w_i \mid \text{spam}) = \frac{|\text{total spam emails (training set) containing } w_i| + 1}{|\text{total spam emails (training set)}| + 2}$$

Same for ham emails:

$$\mathbb{P}(w_i \mid \text{ham}) = \frac{|\text{total ham emails (training set) containing } w_i| + 1}{|\text{total ham emails (training set)}| + 2}$$
Laplace smoothing

Pretend in spam emails (training set):

- We saw one extra spam email with word $w_i$
- We saw one extra spam email without word $w_i$

$$P(w_i \mid \text{spam}) = \frac{|\text{total spam emails (training set) containing } w_i| + 1}{|\text{total spam emails (training set)}| + 2}$$

Same for ham emails:

$$P(w_i \mid \text{ham}) = \frac{|\text{total ham emails (training set) containing } w_i| + 1}{|\text{total ham emails (training set)}| + 2}$$

$$P(\text{“buy” } \mid \text{ ham}) = \frac{0 + 1}{2 + 2} = \frac{1}{4}$$
\[ P(\text{spam} \mid \text{"You buy Viagra"}) = \frac{P(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{spam}) \cdot P(\text{spam})}{P(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{\text{"you"}, \text{"buy"}, \text{"viagra"}\} \mid \text{ham}) \cdot P(\text{ham})} \]

Example

\[
\begin{array}{|c|c|}
\hline
\text{Email} & \text{Label} \\
\hline
\text{Buy Viagra!} & \text{Spam} \\
\text{You good?} & \text{Ham} \\
\text{Viagra help you.} & \text{Spam} \\
\text{Good Viagra help.} & \text{Spam} \\
\text{I need Viagra for my health condition.} & \text{Ham} \\
\hline
\end{array}
\]

\[ P(\text{spam}) = \frac{3}{5} \quad P(\text{ham}) = \frac{2}{5} \]

\[ P(\text{"you"} \mid \text{spam}) = \]

\[ P(\text{"buy"} \mid \text{spam}) = \]

\[ P(\text{"viagra"} \mid \text{spam}) = \]

\[ P(\text{"you"} \mid \text{ham}) = \]

\[ P(\text{"buy"} \mid \text{ham}) = \frac{0 + 1}{2 + 2} = \frac{1}{4} \]

\[ P(\text{"viagra"} \mid \text{ham}) = \]
Example

\[ P(\text{spam} \mid \text{“You buy Viagra”}) = \frac{P(\{\text{“you”, “buy”, “viagra”}\} \mid \text{spam}) \cdot P(\text{spam})}{P(\{\text{“you”, “buy”, “viagra”}\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{\text{“you”, “buy”, “viagra”}\} \mid \text{ham}) \cdot P(\text{ham})} \]

\[ = \frac{P(\text{“you”} \mid \text{spam}) \cdot P(\text{“buy”} \mid \text{spam}) \cdot P(\text{“viagra”} \mid \text{spam}) \cdot P(\text{spam})}{P(\text{“you”} \mid \text{spam}) \cdot P(\text{“buy”} \mid \text{spam}) \cdot P(\text{“viagra”} \mid \text{spam}) \cdot P(\text{spam}) + P(\text{“you”} \mid \text{ham}) \cdot P(\text{“buy”} \mid \text{ham}) \cdot P(\text{“viagra”} \mid \text{ham}) \cdot P(\text{ham})} \]

<table>
<thead>
<tr>
<th>Email</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy Viagra!</td>
<td>Spam</td>
</tr>
<tr>
<td>You good?</td>
<td>Ham</td>
</tr>
<tr>
<td>Viagra help you.</td>
<td>Spam</td>
</tr>
<tr>
<td>Good Viagra help.</td>
<td>Spam</td>
</tr>
<tr>
<td>I need Viagra for my health condition.</td>
<td>Ham</td>
</tr>
</tbody>
</table>

\[ P(\text{spam}) = \frac{3}{5} \]

\[ P(\text{“you”} \mid \text{spam}) = \frac{1 + 1}{3 + 2} = \frac{2}{5} \]

\[ P(\text{“buy”} \mid \text{spam}) = \frac{1 + 1}{2 + 2} = \frac{1}{2} \]

\[ P(\text{“viagra”} \mid \text{spam}) = \frac{1 + 1}{2 + 2} = \frac{1}{4} \]

\[ P(\text{ham}) = \frac{2}{5} \]

\[ P(\text{“you”} \mid \text{ham}) = \frac{0 + 1}{2 + 2} = \frac{1}{4} \]

\[ P(\text{“buy”} \mid \text{ham}) = \frac{1 + 1}{2 + 2} = \frac{1}{4} \]

\[ P(\text{“viagra”} \mid \text{ham}) = \frac{1 + 1}{2 + 2} = \frac{1}{4} \]
\[ P(\text{spam} \mid \text{“You buy Viagra”}) = \frac{P(\{\text{“you”, “buy”, “viagra”}\} \mid \text{spam}) P(\text{spam})}{P(\{\text{“you”, “buy”, “viagra”}\} \mid \text{spam}) P(\text{spam}) + P(\{\text{“you”, “buy”, “viagra”}\} \mid \text{ham}) P(\text{ham})} \]

\[ = \frac{P(\text{“you”} \mid \text{spam}) P(\text{“buy”} \mid \text{spam}) P(\text{“viagra”} \mid \text{spam}) P(\text{spam})}{P(\text{“you”} \mid \text{spam}) P(\text{“buy”} \mid \text{spam}) P(\text{“viagra”} \mid \text{spam}) P(\text{spam}) + P(\text{“you”} \mid \text{ham}) P(\text{“buy”} \mid \text{ham}) P(\text{“viagra”} \mid \text{ham}) P(\text{ham})} \]

<table>
<thead>
<tr>
<th>Email</th>
<th>Label</th>
<th>( P(\text{spam}) )</th>
<th>( P(\text{ham}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy Viagra!</td>
<td>Spam</td>
<td>3/5</td>
<td>2/5</td>
</tr>
<tr>
<td>You good?</td>
<td>Ham</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Viagra help you.</td>
<td>Spam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good Viagra help.</td>
<td>Spam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I need Viagra for my health condition.</td>
<td>Ham</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
P(\text{“you”} \mid \text{spam}) = \frac{1 + 1}{3 + 2} = \frac{2}{5} \quad P(\text{“you”} \mid \text{ham}) = \frac{1 + 1}{2 + 2} = \frac{1}{2}
\]

\[
P(\text{“buy”} \mid \text{spam}) = \frac{1 + 1}{3 + 2} = \frac{2}{5} \quad P(\text{“buy”} \mid \text{ham}) = \frac{0 + 1}{2 + 2} = \frac{1}{4}
\]

\[
P(\text{“viagra”} \mid \text{spam}) = \frac{3 + 1}{3 + 2} = \frac{4}{5} \quad P(\text{“viagra”} \mid \text{ham}) = \frac{1 + 1}{2 + 2} = \frac{1}{2}
\]
\[
P \left( \text{spam} \mid \text{"You buy Viagra"} \right) = \frac{P \left( \{\text{"you", "buy", "viagra"}\} \mid \text{spam} \right) P(\text{spam})}{P \left( \{\text{"you", "buy", "viagra"}\} \mid \text{spam} \right) P(\text{spam}) + P \left( \{\text{"you", "buy", "viagra"}\} \mid \text{ham} \right) P(\text{ham})}
\]

\[
= \frac{P(\text{"you"} \mid \text{spam})P(\text{"buy"} \mid \text{spam})P(\text{"viagra"} \mid \text{spam})P(\text{spam})}{P(\text{"you"} \mid \text{spam})P(\text{"buy"} \mid \text{spam})P(\text{"viagra"} \mid \text{spam})P(\text{spam}) + P(\text{"you"} \mid \text{ham})P(\text{"buy"} \mid \text{ham})P(\text{"viagra"} \mid \text{ham})P(\text{ham})}
\]

\[
= \frac{\frac{2}{5} \cdot \frac{2}{5} \cdot \frac{4}{5} \cdot \frac{3}{5}}{\frac{2}{5} \cdot \frac{2}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{2}{5}} \approx 0.7544
\]

<table>
<thead>
<tr>
<th>Email</th>
<th>Label</th>
<th>( P(\text{spam}) )</th>
<th>( P(\text{ham}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy Viagra!</td>
<td>Spam</td>
<td>( \frac{3}{5} )</td>
<td>( \frac{2}{5} )</td>
</tr>
<tr>
<td>You good?</td>
<td>Ham</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Viagra help you.</td>
<td>Spam</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Good Viagra help.</td>
<td>Spam</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>I need Viagra for my health condition.</td>
<td>Ham</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
\[ \mathbb{P}(\text{spam } | \text{ "You buy Viagra"}) = \frac{\mathbb{P}(\{\text{"you", "buy", "viagra"}\} | \text{ spam}) \mathbb{P}(\text{spam})}{\mathbb{P}(\{\text{"you", "buy", "viagra"}\} | \text{ spam}) \mathbb{P}(\text{spam}) + \mathbb{P}(\{\text{"you", "buy", "viagra"}\} | \text{ ham}) \mathbb{P}(\text{ham})} \]

\[ = \frac{\mathbb{P}(\text{"you"} | \text{ spam}) \mathbb{P}(\text{"buy"} | \text{ spam}) \mathbb{P}(\text{"viagra"} | \text{ spam}) \mathbb{P}(\text{spam})}{\mathbb{P}(\text{"you"} | \text{ spam}) \mathbb{P}(\text{"buy"} | \text{ spam}) \mathbb{P}(\text{"viagra"} | \text{ spam}) \mathbb{P}(\text{spam}) + \mathbb{P}(\text{"you"} | \text{ ham}) \mathbb{P}(\text{"buy"} | \text{ ham}) \mathbb{P}(\text{"viagra"} | \text{ ham}) \mathbb{P}(\text{ham})} \]

\[ = \frac{\frac{2}{5} \cdot \frac{2}{5} \cdot \frac{4}{5} \cdot \frac{3}{5}}{\frac{2}{5} \cdot \frac{2}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{5}} \approx 0.7544 \]

<table>
<thead>
<tr>
<th>Email</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy Viagra!</td>
<td>Spam</td>
</tr>
<tr>
<td>You good?</td>
<td>Ham</td>
</tr>
<tr>
<td>Viagra help you.</td>
<td>Spam</td>
</tr>
<tr>
<td>Good Viagra help.</td>
<td>Spam</td>
</tr>
<tr>
<td>I need Viagra for my health condition.</td>
<td>Ham</td>
</tr>
</tbody>
</table>

\[ \mathbb{P}(\text{spam}) = \frac{3}{5} \]

\[ \mathbb{P}(\text{ham}) = \frac{2}{5} \]

\[ \mathbb{P}(\text{"you"} | \text{ spam}) = \frac{1 + 1}{3 + 2} = \frac{2}{5} \]

\[ \mathbb{P}(\text{"you"} | \text{ ham}) = \frac{1 + 1}{2 + 2} = \frac{1}{2} \]

\[ \mathbb{P}(\text{"buy"} | \text{ spam}) = \frac{1 + 1}{3 + 2} = \frac{2}{5} \]

\[ \mathbb{P}(\text{"buy"} | \text{ ham}) = \frac{0 + 1}{2 + 2} = \frac{1}{4} \]

\[ \mathbb{P}(\text{"viagra"} | \text{ spam}) = \frac{3 + 1}{3 + 2} = \frac{4}{5} \]

\[ \mathbb{P}(\text{"viagra"} | \text{ ham}) = \frac{1 + 1}{2 + 2} = \frac{1}{2} \]
Underflow Prevention

- Multiplication of many probabilities, each of which will be between 0 and 1, can result in floating-point underflow. The product will be too small and will result in arithmetic underflow.
Underflow Prevention

- Multiplication of many probabilities, each of which will be between 0 and 1, can result in floating-point underflow. The product will be too small and will result in arithmetic underflow.

- Reminder: Log property:

\[ \log(xy) = \log(x) + \log(y) \]
Underflow Prevention

- Multiplication of many probabilities, each of which will be between 0 and 1, can result in floating-point underflow. The product will be too small and will result in arithmetic underflow.

- Reminder: Log property:

\[ \log(xy) = \log(x) + \log(y) \]

- Summing logs of probabilities is better than multiplying probabilities

\[
\log \left( \prod_{i=1}^{n} p_i \right) = \log(p_1p_2 \ldots p_n) = \log(p_1) + \log(p_2)+\ldots+\log(p_n) \\
= \sum_{i=1}^{n} \log(p_i)
\]
Applying underflow prevention

\[
\begin{align*}
P(\text{spam} \mid \{w_1, w_2, \ldots, w_n\}) &\approx \frac{P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) \cdot P(\text{spam})}{P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) \cdot P(\text{ham})} \\
P(\text{ham} \mid \{w_1, w_2, \ldots, w_n\}) &\approx \frac{P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) \cdot P(\text{ham})}{P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) \cdot P(\text{ham})}
\end{align*}
\]

We will output \text{spam} iff:
\[
P(\text{spam} \mid \{w_1, w_2, \ldots, w_n\}) > P(\text{ham} \mid \{w_1, w_2, \ldots, w_n\})
\]
Applying underflow prevention

\[ P(\text{spam} \mid \{w_1, w_2, \ldots, w_n\}) \approx \frac{P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) \cdot P(\text{spam})}{P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) \cdot P(\text{ham})} \]

\[ P(\text{ham} \mid \{w_1, w_2, \ldots, w_n\}) \approx \frac{P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) \cdot P(\text{ham})}{P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) \cdot P(\text{ham})} \]

We will output \text{spam} iff:

\[ P(\text{spam} \mid \{w_1, w_2, \ldots, w_n\}) > P(\text{ham} \mid \{w_1, w_2, \ldots, w_n\}) \]

\[ \iff \ P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) \cdot P(\text{spam}) > P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) \cdot P(\text{ham}) \]

Denominators are equal and cancel when comparing.
Applying underflow prevention

\[
P(\text{spam} \mid \{w_1, w_2, \ldots, w_n\}) \approx \frac{P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) \cdot P(\text{spam})}{P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) \cdot P(\text{ham})}
\]

\[
P(\text{ham} \mid \{w_1, w_2, \ldots, w_n\}) \approx \frac{P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) \cdot P(\text{ham})}{P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) \cdot P(\text{spam}) + P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) \cdot P(\text{ham})}
\]

We will output \text{spam} iff:

\[P(\text{spam} \mid \{w_1, w_2, \ldots, w_n\}) > P(\text{ham} \mid \{w_1, w_2, \ldots, w_n\})\]

\[\iff P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) \cdot P(\text{spam}) > P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) \cdot P(\text{ham})\]

\[\iff P(w_1 \mid \text{spam}) \cdot P(w_2 \mid \text{spam}) \cdots P(w_n \mid \text{spam}) \cdot P(\text{spam}) > P(w_1 \mid \text{ham}) \cdot P(w_2 \mid \text{ham}) \cdots P(w_n \mid \text{ham}) \cdot P(\text{ham})\]
Applying underflow prevention

\[
P(\text{spam} \mid \{w_1, w_2, \ldots, w_n\}) \approx \frac{P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) P(\text{spam})}{P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) P(\text{spam}) + P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) P(\text{ham})}
\]

\[
P(\text{ham} \mid \{w_1, w_2, \ldots, w_n\}) \approx \frac{P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) P(\text{ham})}{P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) P(\text{ham}) + P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) P(\text{spam})}
\]

We will output spam iff:

\[
P(\text{spam} \mid \{w_1, w_2, \ldots, w_n\}) > P(\text{ham} \mid \{w_1, w_2, \ldots, w_n\})
\]

\[\iff P(\{w_1, w_2, \ldots, w_n\} \mid \text{spam}) P(\text{spam}) > P(\{w_1, w_2, \ldots, w_n\} \mid \text{ham}) P(\text{ham})
\]

\[\iff P(w_1 \mid \text{spam}) P(w_2 \mid \text{spam}) \cdots P(w_n \mid \text{spam}) P(\text{spam}) > P(w_1 \mid \text{ham}) P(w_2 \mid \text{ham}) \cdots P(w_n \mid \text{ham}) P(\text{ham})
\]

Taking the log of two sides:

\[\iff \log(P(\text{spam})) + \sum_{i=1}^{n} \log(P(w_i \mid \text{spam})) > \log(P(\text{ham})) + \sum_{i=1}^{n} \log(P(w_i \mid \text{ham}))\]
Summary: Naive Bayes Algorithm steps

1. **TRAINING**

   1.1. Compute the proportion of emails in the **training set** that is spam or ham:
   
   $$P(\text{spam}) = \frac{\text{total spam emails (in training set)}}{\text{total emails (in training set)}} \quad P(\text{ham}) = \frac{\text{total ham emails (in training set)}}{\text{total emails (in training set)}}$$

   1.2. Iterate over the **training set**, for each unique word $x$, count:
   
   - How many **spam emails** in the training set contain $x$
   - How many **ham emails** in the training set contain $x$
2. **TESTING**

Iterate over the **test set**, for each unlabelled email $D$:

- Create a set $S$ of $n$ unique words appearing in $D$:\$\{w_1, w_2, \ldots, w_n\}\$
- For each word $w_i$ in set $S$, calculate:

\[
P(x \mid \text{spam}) = \frac{|\text{total spam emails (training set) containing } w_i| + 1}{|\text{total spam emails (training set)}| + 2}
\]

\[
P(w_i \mid \text{ham}) = \frac{|\text{total ham emails (training set) containing } w_i| + 1}{|\text{total ham emails (training set)}| + 2}
\]

  - Note: If word $w_i$ doesn’t appear in the training set, we still calculate the above probabilities, with:

\[
|\text{total spam emails (training set) containing } w_i| = |\text{total ham emails (training set) containing } w_i| = 0
\]

- if $\log(P(\text{spam})) + \sum_{i=1}^{n} \log(P(w_i \mid \text{spam})) > \log(P(\text{ham})) + \sum_{i=1}^{n} \log(P(w_i \mid \text{ham}))$

  Predict email $D$ as **spam**

  Otherwise, predict email $D$ as **ham**