Data Science of Text Generation
1. Taking your chances

Ernst C. Wit and Francisco Richter

wite@usi.ch, richterf@usi.ch
Università della Svizzera italiana

Bachelor in Data Science
https://www.usi.ch/en/education/bachelor/data-science

A funny poem

Normally, probability starts with an urn of coloured balls.

We start with a poem:

Do you carrot all for me?
My heart beets for you,
With your turnip nose
And your radish face,
You are a peach.
If we cantaloupe,
Lettuce marry:
Weed make a swell pear.

consisting of 28 different words.

Task 1. Pick one random word from poem and write down.

What are the chances...

The 28 words in poem consist of:

- **9 food items**: carrot, beets, turnip, radish, peach, cantaloupe, lettuce, weed, pear.
- **3 body parts:** heart, nose, face
- **4 verbs:** do, are, marry, make
- **5 pronouns:** you, me, my, your, we
- **7 others:** all, for, with, and, a, if, swell

Task 2. What is the probability that you chose a food item?

What are the chances...

The 28 words in poem consist of:

- **9 food items**: carrot, beets, turnip, radish, peach, cantaloupe, lettuce, weed, pear.
- **3 body parts:** heart, nose, face
- **4 verbs:** do, are, marry, make
- **5 pronouns:** you, me, my, your, we
- **7 others:** all, for, with, and, a, if, swell

Task 2. What is the probability that you chose a food item?

$$P(\text{chose food item}) = \frac{9}{28} = 0.32$$

Probability rules!

Definition of probability according to Marquis de Laplace (1779)

Pierre Simon Laplace

Probability *of event is ratio of*

- *number of cases* ***favorable****, to*
- *number of all cases* ***possible****;*

when ***nothing leads us to expect that any one of these cases should occur more than any other****, which renders them, for us, equally possible.*

Definition of probability according to Marquis de Laplace (1779)

Pierre Simon Laplace

Probability *of event is ratio of*

- *number of cases* ***favorable****, to*
- *number of all cases* ***possible****;*

when ***nothing leads us to expect that any one of these cases should occur more than any other****, which renders them, for us, equally possible.*

In **mathematical terms**:

$$P(E) = \frac{\text{Number of elements in } E}{\text{Total number of elements}}$$

where E is an **event**.

Example: 2 words from poem

Consider words **you and your neighbour** selected from poem.

What is **sample space**?

$$S_{\text{poem2}} =$$

Example: 2 words from poem

Consider words **you and your neighbour** selected from poem.

What is **sample space**?

$$
\begin{aligned}
S_{\text{poem2}} &= \{(\text{carrot, carrot}), \ldots (\text{carrot, pear}), \ldots, (\text{pear, pear})\} \\
&=
\end{aligned}
$$

Example: 2 words from poem

Consider words **you and your neighbour** selected from poem.

What is **sample space**?

$$
\begin{aligned}
S_{\text{poem2}} &= \{(\text{carrot, carrot}), \ldots (\text{carrot, pear}), \ldots, (\text{pear, pear})\} \\
&= \{28 \times 28 \text{ word combinations}\}
\end{aligned}
$$

Let's consider the event

$$E = \{\text{both of you choose food items}\}$$

Note,

$$|E| =$$

Example: 2 words from poem

Consider words **you and your neighbour** selected from poem.

What is **sample space**?

$$\begin{aligned} S_{\text{poem2}} &= \{(\text{carrot}, \text{carrot}), \ldots (\text{carrot}, \text{pear}), \ldots, (\text{pear}, \text{pear})\} \\ &= \{28 \times 28 \text{ word combinations}\} \end{aligned}$$

Let's consider the event

$$E = \{\text{both of you choose food items}\}$$

Note,

$$|E| = 9 \times 9.$$

If you didn't cheat, then

$$P(E) =$$

Example: 2 words from poem

Consider words **you and your neighbour** selected from poem.

What is **sample space**?

$$\begin{aligned} S_{\text{poem2}} &= \{(\text{carrot, carrot}), \ldots (\text{carrot, pear}), \ldots, (\text{pear, pear})\} \\ &= \{28 \times 28 \text{ word combinations}\} \end{aligned}$$

Let's consider the event

$$E = \{\text{both of you choose food items}\}$$

Note,

$$|E| = 9 \times 9.$$

If you didn't cheat, then

$$P(E) = \frac{9 \times 9}{28 \times 28} = .10$$

Independence

There is something special about previous example:

Your word does not affect your neighbour's word

So, events

$$\begin{aligned} A &= \{\text{your word is food item}\} \\ B &= \{\text{your neighbour's word is food item}\} \end{aligned}$$

are so-called **independent events.**

In case of independent events, we can use

$$P(A \cap B) = P(A)P(B)$$

Example. 2 words from poem

$$P(A \cap B) \quad =$$

Independence

There is something special about previous example:

Your word does not affect your neighbour's word

So, events

$$\begin{aligned} A &= \{\text{your word is food item}\} \\ B &= \{\text{your neighbour's word is food item}\} \end{aligned}$$

are so-called **independent events.**

In case of independent events, we can use

$$P(A \cap B) = P(A)P(B)$$

Example. 2 words from poem

$$\begin{aligned} P(A \cap B) &= P(A)P(B) \\ &= \end{aligned}$$

Independence

There is something special about previous example:

Your word does not affect your neighbour's word

So, events

$$\begin{aligned} A &= \{\text{your word is food item}\} \\ B &= \{\text{your neighbour's word is food item}\} \end{aligned}$$

are so-called **independent events.**

In case of independent events, we can use

$$P(A \cap B) = P(A)P(B)$$

Example. 2 words from poem

$$\begin{aligned} P(A \cap B) &= P(A)P(B) \\ &= \frac{9}{28} \times \frac{9}{28} = 0.10 \end{aligned}$$

Conditional Probabilities

Dependence

Independence is great, because

- we can focus on smaller sample space
- which makes calculations easier

However, often events are **not independent.**

Example: Poem. Probability of selecting food item with a “w”?

$$
\begin{aligned}
A &= \{\text{food item}\} \\
B &= \{\text{contains a “w”}\}
\end{aligned}
$$

Dependence

Independence is great, because

- we can focus on smaller sample space
- which makes calculations easier

However, often events are **not independent.**

Example: Poem. Probability of selecting food item with a "w"?

$$
\begin{aligned}
A &= \{\text{food item}\} \\
B &= \{\text{contains a "w"}\}
\end{aligned}
$$

$$0.04 = \frac{1}{28} = P(A \cap B) \neq P(A)P(B) = \frac{9}{28}\frac{4}{28} = 0.05$$

How can we do simple calculations with dependent events?

Definition of conditional probability

Example. Draw 2 cards from deck without replacement.

$$\begin{aligned} E &= \text{1st card is ace} \\ F &= \text{2nd card is ace} \end{aligned}$$

$$P(E \cap F) =$$

Definition of conditional probability

Example. Draw 2 cards from deck without replacement.

$$\begin{aligned} E &= \text{1st card is ace} \\ F &= \text{2nd card is ace} \end{aligned}$$

$$\begin{aligned} P(E \cap F) &= \frac{4}{52}\frac{3}{51} \\ &= \end{aligned}$$

Definition of conditional probability

Example. Draw 2 cards from deck without replacement.

$$\begin{aligned} E &= \text{1st card is ace} \\ F &= \text{2nd card is ace} \end{aligned}$$

$$\begin{aligned} P(E \cap F) &= \frac{4}{52}\frac{3}{51} \\ &= P(E)P(F|E) \end{aligned}$$

where $P(F|E)$ is the probability of F given E.

Definition. Conditional probability of A if B happened:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Using conditional probabilities

Example: Poem. Probability of selecting food item with a "w"?

$$
\begin{aligned}
A &= \{\text{food item}\} \\
B &= \{\text{contains a “w”}\}
\end{aligned}
$$

$$P(A \cap B) \;=$$

Using conditional probabilities

Example: Poem. Probability of selecting food item with a "w"?

$$
\begin{aligned}
A &= \{\text{food item}\} \\
B &= \{\text{contains a "w"}\}
\end{aligned}
$$

$$
\begin{aligned}
P(A \cap B) &= P(A)P(B|A) \\
&=
\end{aligned}
$$

Using conditional probabilities

Example: Poem. Probability of selecting food item with a "w"?

$$\begin{aligned} A &= \{\text{food item}\} \\ B &= \{\text{contains a “w”}\} \end{aligned}$$

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ &= \frac{9}{28}\frac{1}{9} \\ &= \end{aligned}$$

Using conditional probabilities

Example: Poem. Probability of selecting food item with a “w”?

$$\begin{aligned} A &= \{\text{food item}\} \\ B &= \{\text{contains a “w”}\} \end{aligned}$$

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ &= \frac{9}{28}\frac{1}{9} \\ &= \frac{1}{28} \end{aligned}$$

Multiplication rule

Multiplication rule

For three events A, B, C (not necessarily independent),

$$P(A \cap B \cap C) = P(A) \times P(B \mid A) \times P(C \mid A \cap B)$$

Example. Consider taking a 3 cards from a pack of cards. What is the probability that they are all aces?

$$A_i = \text{Ace in } i\text{th draw}$$

So we want to know,

$$P(A_1 \cap A_2 \cap A_3) \quad =$$

Multiplication rule

Multiplication rule

For three events A, B, C (not necessarily independent),

$$P(A \cap B \cap C) = P(A) \times P(B \mid A) \times P(C \mid A \cap B)$$

Example. Consider taking a 3 cards from a pack of cards. What is the probability that they are all aces?

$$A_i = \text{Ace in } i\text{th draw}$$

So we want to know,

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1) \times P(A_2 \mid A_1) \times P(A_3 \mid A_1 \cap A_2) \\ &= \end{aligned}$$

Multiplication rule

Multiplication rule

For three events A, B, C (not necessarily independent),

$$P(A \cap B \cap C) = P(A) \times P(B \mid A) \times P(C \mid A \cap B)$$

Example. Consider taking a 3 cards from a pack of cards. What is the probability that they are all aces?

$$A_i = \text{Ace in } i\text{th draw}$$

So we want to know,

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1) \times P(A_2 \mid A_1) \times P(A_3 \mid A_1 \cap A_2) \\ &= \frac{4}{52} \times \end{aligned}$$

Multiplication rule

Multiplication rule

For three events A, B, C (not necessarily independent),

$$P(A \cap B \cap C) = P(A) \times P(B \mid A) \times P(C \mid A \cap B)$$

Example. Consider taking a 3 cards from a pack of cards. What is the probability that they are all aces?

$$A_i = \text{Ace in } i\text{th draw}$$

So we want to know,

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1) \times P(A_2 \mid A_1) \times P(A_3 \mid A_1 \cap A_2) \\ &= \frac{4}{52} \times \frac{3}{51} \times \end{aligned}$$

Multiplication rule

Multiplication rule

For three events A, B, C (not necessarily independent),

$$P(A \cap B \cap C) = P(A) \times P(B \mid A) \times P(C \mid A \cap B)$$

Example. Consider taking a 3 cards from a pack of cards. What is the probability that they are all aces?

$$A_i = \text{Ace in } i\text{th draw}$$

So we want to know,

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1) \times P(A_2 \mid A_1) \times P(A_3 \mid A_1 \cap A_2) \\ &= \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \\ &= \end{aligned}$$

Multiplication rule

Multiplication rule

For three events A, B, C (not necessarily independent),

$$P(A \cap B \cap C) = P(A) \times P(B \mid A) \times P(C \mid A \cap B)$$

Example. Consider taking a 3 cards from a pack of cards. What is the probability that they are all aces?

$$A_i = \text{Ace in } i\text{th draw}$$

So we want to know,

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1) \times P(A_2 \mid A_1) \times P(A_3 \mid A_1 \cap A_2) \\ &= \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \\ &= 0.00018 \end{aligned}$$

Bayes' Theorem

Who was Reverend Bayes?

Reverend Thomas Bayes
(c.1702 - 17 April 1761)

- English mathematician and Presbyterian minister
- 1743: Elected Fellow of Royal Society.
- 1761: Thomas Bayes dies
- 1763: *Essay Towards Solving a Problem in the Doctrine of Chances* read before Royal Society of London
- 20th century: famous for solving problem of "inverse probability"

Bayes' Theorem

Assume we get data D from the true state S_0 of reality:

$$\begin{aligned} D &= \{\text{data}\} \\ S &= \{\text{state of the world.}\} \end{aligned}$$

Question. Given data D what is our belief in $S \in \{S_0, S_1, \ldots, S_n\}$?

Note that typically $P(D \mid S_i)$ is easy.

Bayes' Theorem

$$P(S \mid D) = \frac{P(D \mid S)P(S)}{\sum_{i=0}^{n} P(D \mid S_i)P(S_i)}$$

Probabilities $P(S_i)$ need to be assumed known *a priori*.

God's not playing dice, but flipping coins...

Imagine that at beginning of time, God flips a fair coin:

- If **heads**, then God creates two universes:
  one with black-haired people, other with blond haired people.
- If **tails**, then God creates one black-haired universe.

Now suppose that you are living in black-haired universe.

Then what is probability of God's coin having landed heads?

God's flipping coins...

$$\begin{aligned} E &= \{\text{Living in a black-haired universe.}\} \\ F &= \{\text{Heads}\} \end{aligned}$$

Given data E what is our posterior belief in F?

$$P(F|E) =$$

God's flipping coins...

$$
\begin{aligned}
E &= \{\text{Living in a black-haired universe.}\} \\
F &= \{\text{Heads}\}
\end{aligned}
$$

Given data E what is our posterior belief in F?

$$
\begin{aligned}
P(F|E) &= \frac{P(FE)}{P(E)} \\
&=
\end{aligned}
$$

God's flipping coins...

$$
\begin{aligned}
E &= \{\text{Living in a black-haired universe.}\} \\
F &= \{\text{Heads}\}
\end{aligned}
$$

Given data E what is our posterior belief in F?

$$
\begin{aligned}
P(F|E) &= \frac{P(FE)}{P(E)} \\
&= \frac{P(E|F)P(F)}{}
\end{aligned}
$$

God's flipping coins...

$$\begin{aligned} E &= \{\text{Living in a black-haired universe.}\} \\ F &= \{\text{Heads}\} \end{aligned}$$

Given data E what is our posterior belief in F?

$$\begin{aligned} P(F|E) &= \frac{P(FE)}{P(E)} \\ &= \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)} \\ &= \end{aligned}$$

God's flipping coins...

$$
\begin{aligned}
E &= \{\text{Living in a black-haired universe.}\} \\
F &= \{\text{Heads}\}
\end{aligned}
$$

Given data E what is our posterior belief in F?

$$
\begin{aligned}
P(F|E) &= \frac{P(FE)}{P(E)} \\
&= \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)} \\
&= \frac{0.5 \times 0.5}{0.5 \times 0.5 +}
\end{aligned}
$$

God's flipping coins...

$$\begin{aligned} E &= \{\text{Living in a black-haired universe.}\} \\ F &= \{\text{Heads}\} \end{aligned}$$

Given data E what is our posterior belief in F?

$$\begin{aligned} P(F|E) &= \frac{P(FE)}{P(E)} \\ &= \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)} \\ &= \frac{0.5 \times 0.5}{0.5 \times 0.5 + 1 \times 0.5} \\ &= 1/3 \end{aligned}$$

Sentiment Analysis

Sentiment analysis

We get a piece of text (e.g. tweet) and we want to know:

Does it express a positive or negative sentiment?

- Let's consider following dictionary:

$$C = \{\text{of, great, kind, weird, stuff, mean}\}$$

- Two two sentiments: $S \in \{\text{positive, negative}\}$
- Conditional probabilities $P(word \mid sentiment)$ are:

word	positive	negative
of	0.1	0.1
great	0.3	0.1
kind	0.3	0.1
weird	0.1	0.3
stuff	0.1	0.2
mean	0.1	0.2

Tweet: "weird kind of stuff"

Is above tweet positive or negative?

Define our events:

- W_i = ith word ($i = 1, 2, 3, 4$)
- N = negative sentiment

Tweet: "weird kind of stuff"

Is above tweet positive or negative?

Define our events:

- W_i = ith word ($i = 1, 2, 3, 4$)
- N = negative sentiment

Bayes' Theorem! Let prior probability $P(N) = 0.5$:

$$\begin{aligned} P(N \mid W_1 \cap \ldots \cap W_4) &= \frac{P(W_1 \cap \ldots \cap W_4 \mid N)P(N)}{P(W_1 \cap \ldots \cap W_4 \mid N)P(N) + P(W_1 \cap \ldots \cap W_4 \mid N^c)P(N^c)} \\ &= \end{aligned}$$

Tweet: “weird kind of stuff”

Is above tweet positive or negative?

Define our events:

- W_i = ith word ($i = 1, 2, 3, 4$)
- N = negative sentiment

Bayes' Theorem! Let prior probability $P(N) = 0.5$:

$$
\begin{aligned}
P(N \mid W_1 \cap \ldots \cap W_4) &= \frac{P(W_1 \cap \ldots \cap W_4 \mid N)P(N)}{P(W_1 \cap \ldots \cap W_4 \mid N)P(N) + P(W_1 \cap \ldots \cap W_4 \mid N^c)P(N^c)} \\
&= 0.3 \times
\end{aligned}
$$

Tweet: "weird kind of stuff"

Is above tweet positive or negative?

Define our events:

- W_i = ith word ($i = 1, 2, 3, 4$)
- N = negative sentiment

Bayes' Theorem! Let prior probability $P(N) = 0.5$:

$$
\begin{aligned}
P(N \mid W_1 \cap \ldots \cap W_4) &= \frac{P(W_1 \cap \ldots \cap W_4 \mid N)P(N)}{P(W_1 \cap \ldots \cap W_4 \mid N)P(N) + P(W_1 \cap \ldots \cap W_4 \mid N^c)P(N^c)} \\
&= 0.3 \times 0.1 \times 0.1 \times 0.2 \times
\end{aligned}
$$

Tweet: "weird kind of stuff"

Is above tweet positive or negative?

Define our events:

- W_i = ith word ($i = 1, 2, 3, 4$)
- N = negative sentiment

Bayes' Theorem! Let prior probability $P(N) = 0.5$:

$$\begin{aligned} P(N \mid W_1 \cap \ldots \cap W_4) &= \frac{P(W_1 \cap \ldots \cap W_4 \mid N)P(N)}{P(W_1 \cap \ldots \cap W_4 \mid N)P(N) + P(W_1 \cap \ldots \cap W_4 \mid N^c)P(N^c)} \\ &= \frac{0.3 \times 0.1 \times 0.1 \times 0.2 \times 0.5}{0.3 \times 0.1 \times 0.1 \times 0.2 \times 0.5 + 0.1 \times 0.3 \times 0.1 \times 0.1 \times 0.5} \\ &= \end{aligned}$$

Tweet: "weird kind of stuff"

Is above tweet positive or negative?

Define our events:

- W_i = ith word ($i = 1, 2, 3, 4$)
- N = negative sentiment

Bayes' Theorem! Let prior probability $P(N) = 0.5$:

$$
\begin{aligned}
P(N \mid W_1 \cap \ldots \cap W_4) &= \frac{P(W_1 \cap \ldots \cap W_4 \mid N)P(N)}{P(W_1 \cap \ldots \cap W_4 \mid N)P(N) + P(W_1 \cap \ldots \cap W_4 \mid N^c)P(N^c)} \\
&= \frac{0.3 \times 0.1 \times 0.1 \times 0.2 \times 0.5}{0.3 \times 0.1 \times 0.1 \times 0.2 \times 0.5 + 0.1 \times 0.3 \times 0.1 \times 0.1 \times 0.5} \\
&= 0.67
\end{aligned}
$$

Tweet: "weird kind of stuff"

Is above tweet positive or negative?

Define our events:

- W_i = ith word ($i = 1, 2, 3, 4$)
- N = negative sentiment

Bayes' Theorem! Let prior probability $P(N) = 0.5$:

$$\begin{aligned} P(N \mid W_1 \cap \ldots \cap W_4) &= \frac{P(W_1 \cap \ldots \cap W_4 \mid N)P(N)}{P(W_1 \cap \ldots \cap W_4 \mid N)P(N) + P(W_1 \cap \ldots \cap W_4 \mid N^c)P(N^c)} \\ &= \frac{0.3 \times 0.1 \times 0.1 \times 0.2 \times 0.5}{0.3 \times 0.1 \times 0.1 \times 0.2 \times 0.5 + 0.1 \times 0.3 \times 0.1 \times 0.1 \times 0.5} \\ &= 0.67 \end{aligned}$$

We have *secretly* made use of conditional independence!
We relax this assumption in the afternoon.

Conclusions

In this class we have learned:

1. Laplace's definition of probability
2. Independence simplifies calculations.
3. Conditional probabilities are also easy.
4. Bayes' Theorem to learn about reality from data.