What is Sentiment Analysis?
Bag of Words

What is a Bag of Words?

- One way a computer reads a text is called "Bag of Words"
- The computer makes a list of all words used along with how many times each word occurs in the text

Vocabulary:

- **Bag of words**: a method of analyzing a text that looks at the whole text as a bunch of separate words.

ACTIVITY SLIDE

The following words were all pulled from a poem. Look at the poem like this. What do you think it's about?

Dreams - 5
A - 2
Fast -2
For - 2
Hold - 2
Is - 2
Life - 2
To - 2
Broken-winged - 1
Barren - 1
Bird - 1

Cannot - 1
Die - 1
Fly - 1
Field - 1
Frozen - 1
Go - 1
If - 1
Snow - 1
That - 1
When - 1
With - 1

DISCUSSION

- What can you tell about this poem from the word count?

- What does it seem to be about?

Dreams - Langston Hughes

Hold fast to dreams

For if dreams die

Life is a broken-winged bird

That cannot fly.

Hold fast to dreams

For when dreams go

Life is a barren field

Frozen with snow.

ACTIVITY SLIDE

The following words were all pulled from a review. Can you tell if it is a positive or negative review? How? Label the words "Positive," "Negative," or "Neutral."

Was - 3
Alright - 2
It - 2
And - 1
Generic - 1
Good - 1
Just - 1
No - 1
Nothing - 1
Pretty - 1

Soundtrack - 1
Special - 1
Story - 1
Twists - 1
Unexpected - 1
Visual - 1
Were - 1
With - 1

The Review

It was alright, nothing special, it was just alright, visual and soundtrack were good, story was pretty generic with no unexpected twists (7 stars)

DISCUSSION

- Would you call this review "positive," "negative," or "neutral"?

- Were you correct about the review when you saw it as just words? Why or why not?

- What might have helped you be more accurate?

Vocabulary

N-grams: Sets of n words

- A 3-gram (also called a "trigram") would be a set of three words like "I like pickles"
- A 2-gram (also called a "bigram") would be sets of 2 words such as "I like" and "like pickles."

A bag of words model can use n-grams instead of single words.

DISCUSSION

- What might have helped you understand the bag of words that you saw for the poem and the review better?
  - Would having more words have helped?

Order

- Bag of words doesn't take the order of words into account (except in n-grams).
- Other methods do, for example:
  - Turning words into vectors as described briefly in unit one
  - **Bayesian probability** - Guessing what word is likely the next word based on the word before it

Unit Learning Goals

You will be able to:

- Explain how computers see texts as a "bag of words" that are not necessarily related to each other

- Name ways sentiment analysis could be biased

- Describe impacts sentiment analysis can have on different communities

Exit ticket: Unit 2.07 - Bag of Words

Describe a difference between how a human reads a text and how a computer does.