

CS 1671/2071

Human Language Technologies

Session 4: Machine learning intro, NLP tasks and applications

Michael Miller Yoder

January 22, 2025



School of Computing and Information

Overview: Machine learning intro, NLP tasks and applications

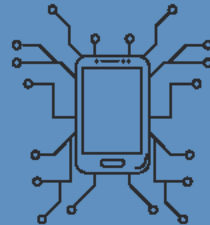
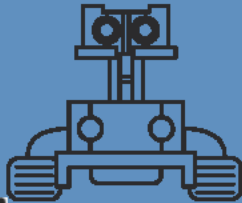
- Intro to machine learning
 - Definitions
 - Models and algorithms
 - Data: training, development, test
- NLP applications
- NLP “core tasks”
- Coding activity: clickbait classification

Course logistics

- I released a new **optional, extra credit** homework assignment, [Homework 0](#) on getting set up for installing Python packages on the CRCD JupyterHub
 - Is due tonight, Wed Jan 22, at 11:59pm
- [Homework 1](#) is **due this Thu Jan 23 at 11:59pm**
- I plan on releasing example projects and a form to submit project ideas you may want to work on Fri Jan 24
 - Project idea submission form will be due next Thu Jan 30

JOIN PITT'S NEWEST TECH CLUB: TECHNOLOGY FOR GOOD

responsible tech is a growing field addressing issues of:



ethical AI, democracy in tech, tech policy

weekly meetings will consist of project workshops, site visits, research opportunities, and more!

join the founding class!!!



scan the qr code to join!



@t4g_at_pitt



Technology for Good
at Pitt

- Intro to (supervised) machine learning

What is machine learning?

- A system that learns a function (maps from an input to an output) from examples/data
- Can predict things and perform tasks **without** explicit instructions
- Learns patterns from data with statistical algorithms

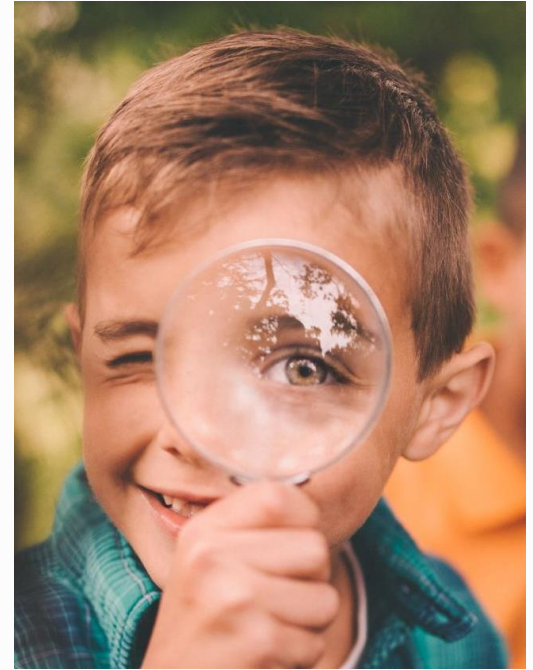
Machine learning models

- Transform an input to an output with a “model”: a simplified mathematical/statistical version of reality
- Models have parameters **learned from patterns in data**
 - Usually encode how variables relate to each other



Machine learning models and algorithms

“All models are wrong. Some are useful.”



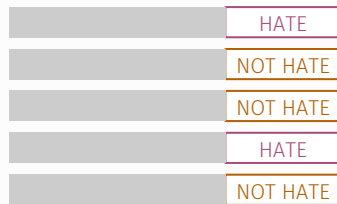
Machine learning algorithms

- Algorithms are systematic ways of doing things
- In machine learning, “algorithms” refers to systematic ways of estimating model parameters from data

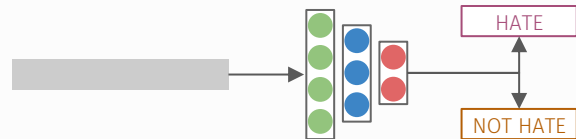
Supervised machine learning process



Data
(input text, X)



Annotate
labels (Y)



Train a model to
predict labels (Y)
from input text (X)

Training and test sets (and phases)

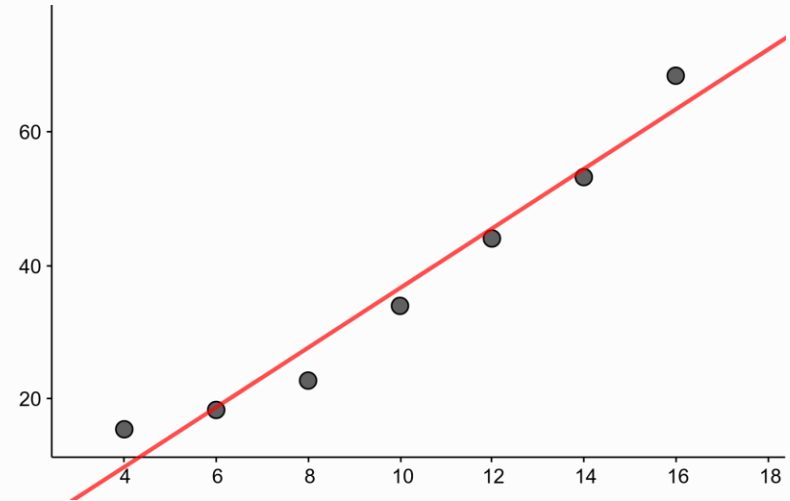
Training set

Test Set

- Train parameters of the model on training set (training phase)
 - Sees examples of input and (assumed correct) output that it will mimic
- Test time
 - Freeze parameters of the model
 - Predict input from an unseen set
 - Evaluate on correct answers and see how well the model performs
- **Don't look at the test set too much when developing/choosing models**

What can you do with machine learning models?

- Prediction: predict an output from an unseen input
 - That fits the pattern learned by looking at input it has seen before
- Interpretation
 - Examine the learned model weights to characterize the relationship between variables



$$y = 4x - 10$$

NLP applications

Core tasks and applications of NLP

machine translation

chatbots

information retrieval

APPLICATIONS

summarization

question answering

NLP applications: email classification

The screenshot displays an email inbox interface with a left sidebar and a main content area. The sidebar includes a 'COMPOSE' button, 'Inbox (7)', 'Starred', 'Drafts', and 'Sent Mail' sections. Below these are contact icons and a search bar for people. The main area shows a list of emails categorized into 'Primary', 'Social', 'Promotions', and 'Updates'. Each category has a 'new' badge indicating unread messages. The 'Social' category is currently selected.

Category	Sender	Subject
Primary	Google+	You were tagged in 3 photos on Google+ - Google+ You were tagged in three pl
Primary	YouTube	LauraBlack just uploaded a video. - Jess, have you seen the video LauraBlack u
Primary	Emily Million (Google+)	[Knitting Club] Are we knitting tonight? - [Knitting Club] Are we knitting tonight?
Primary	Sean Smith (Google+)	Photos of the new pup - Sean Smith shared an album with you. View album be tho
Primary	Google+	Kate Baynham shared a post with you - Follow and share with Kate by adding her
Primary	Google+	Danielle Hoodhood added you on Google+ - Follow and share with Danielle by
Primary	YouTube	Just for You From YouTube: Daily Update - Jun 19, 2013 - Check out the latest
Primary	Google+	You were tagged in 3 photos on Google+ - Google+ You were tagged in three phot
Primary	Hilary Jacobs (Google+)	Check out photos of my new apt - Hilary Jacobs shared an album with you. View
Primary	Google+	Kate Baynham added you on Google+ - Follow and share with Kate by adding her

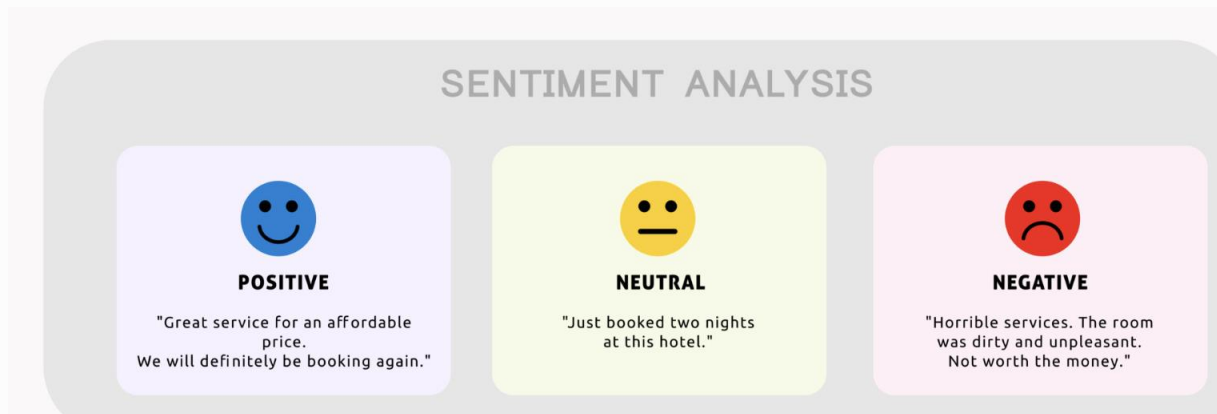
NLP applications: email classification

The screenshot displays an email inbox interface. On the left, there is a sidebar with a 'COMPOSE' button, 'Inbox (7)', 'Starred', 'Drafts', and 'Sent Mail' sections. Below these is a contact list with names like Jenny Kang, Peter H, Jonathan Pelleg, Brett C, Max Stein, Jen Hart, and Eric Lowery. The main inbox area is divided into four tabs: 'Primary', 'Social' (3 new), 'Promotions' (2 new), and 'Updates' (2 new). The 'Social' tab is selected, showing a list of social media notifications from Google+, YouTube, and Emily Million. Each email entry includes a checkbox, a star icon, the sender, and a snippet of the email body.

Category	Sender	Subject / Snippet
Primary	Google+	You were tagged in 3 photos on Google+ - Google+ You were tagged in three pl
Primary	YouTube	LauraBlack just uploaded a video. - Jess, have you seen the video LauraBlack u
Primary	Emily Million (Google+)	[Knitting Club] Are we knitting tonight? - [Knitting Club] Are we knitting tonight?
Primary	Sean Smith (Google+)	Photos of the new pup - Sean Smith shared an album with you. View album be tho
Primary	Google+	Kate Baynham shared a post with you - Follow and share with Kate by adding her
Primary	Google+	Danielle Hoodhood added you on Google+ - Follow and share with Danielle by
Primary	YouTube	Just for You From YouTube: Daily Update - Jun 19, 2013 - Check out the latest
Primary	Google+	You were tagged in 3 photos on Google+ - Google+ You were tagged in three phot
Primary	Hilary Jacobs (Google+)	Check out photos of my new apt - Hilary Jacobs shared an album with you. View
Primary	Google+	Kate Baynham added you on Google+ - Follow and share with Kate by adding her

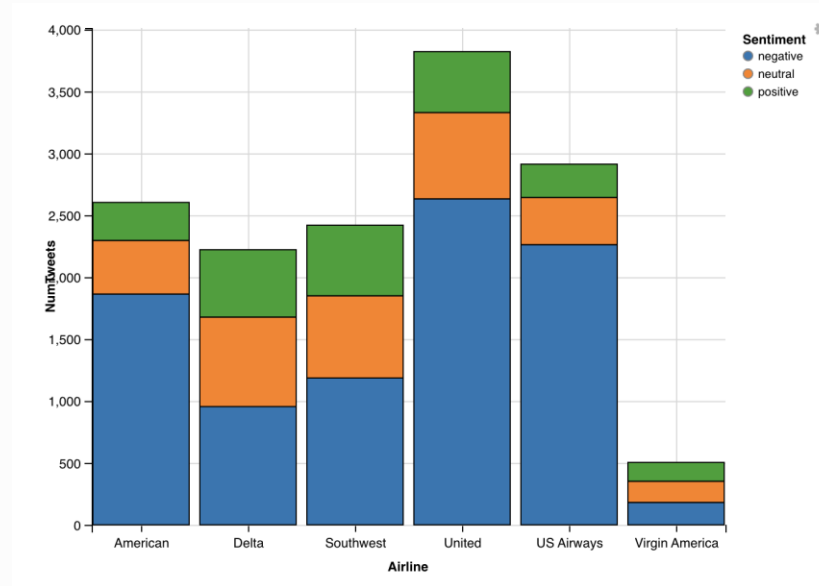
- Spam / Not spam
- Priority Level
- Category (primary / social / promotions / updates)

NLP applications: sentiment analysis



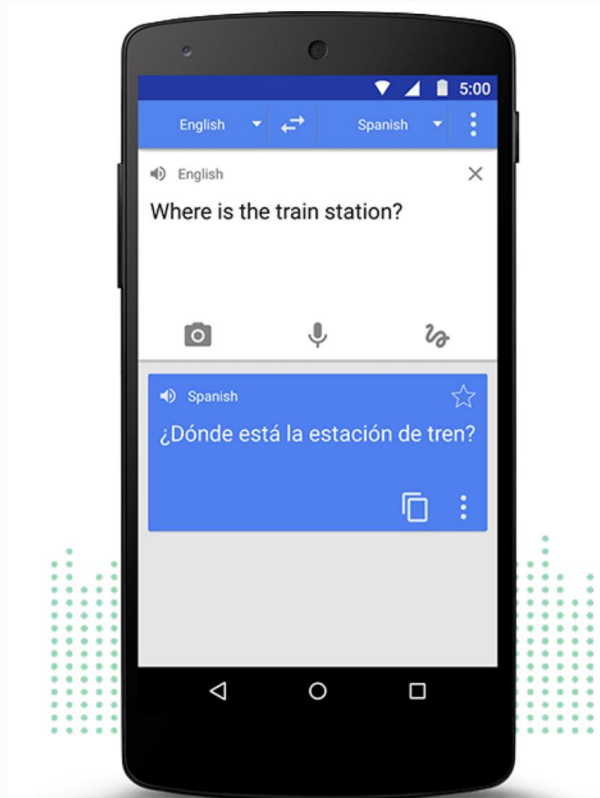
- Hotel review sentiment

NLP applications: sentiment analysis



- US Airline review sentiment

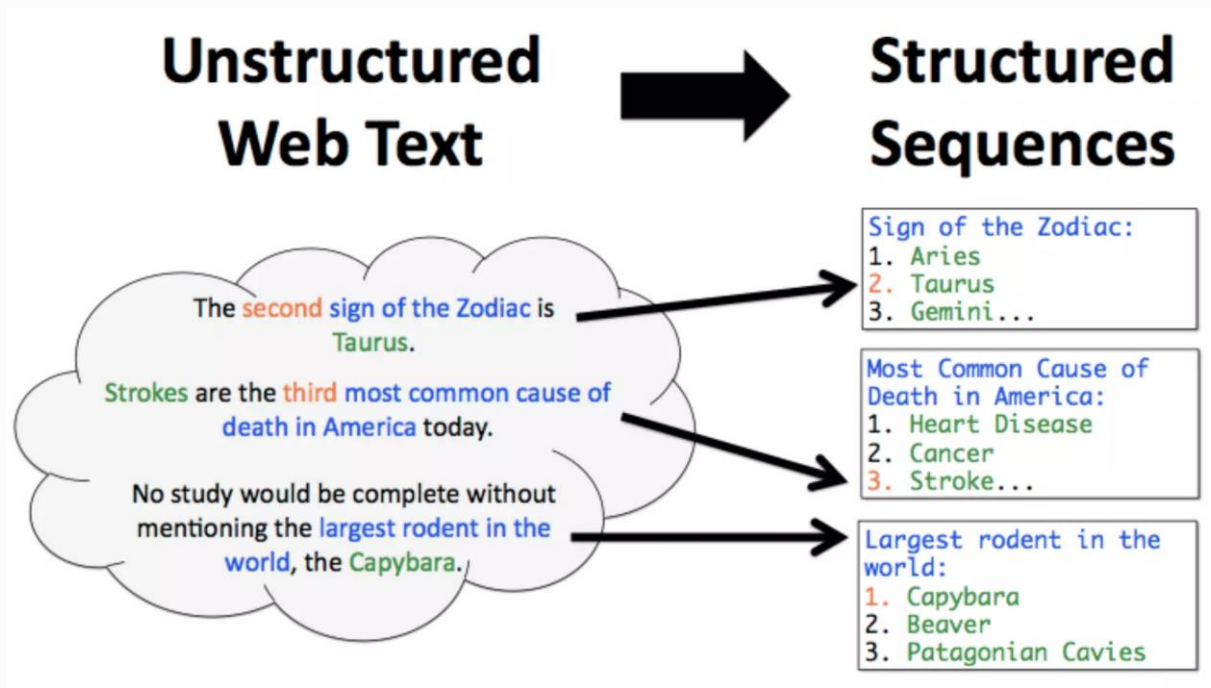
NLP applications: machine translation



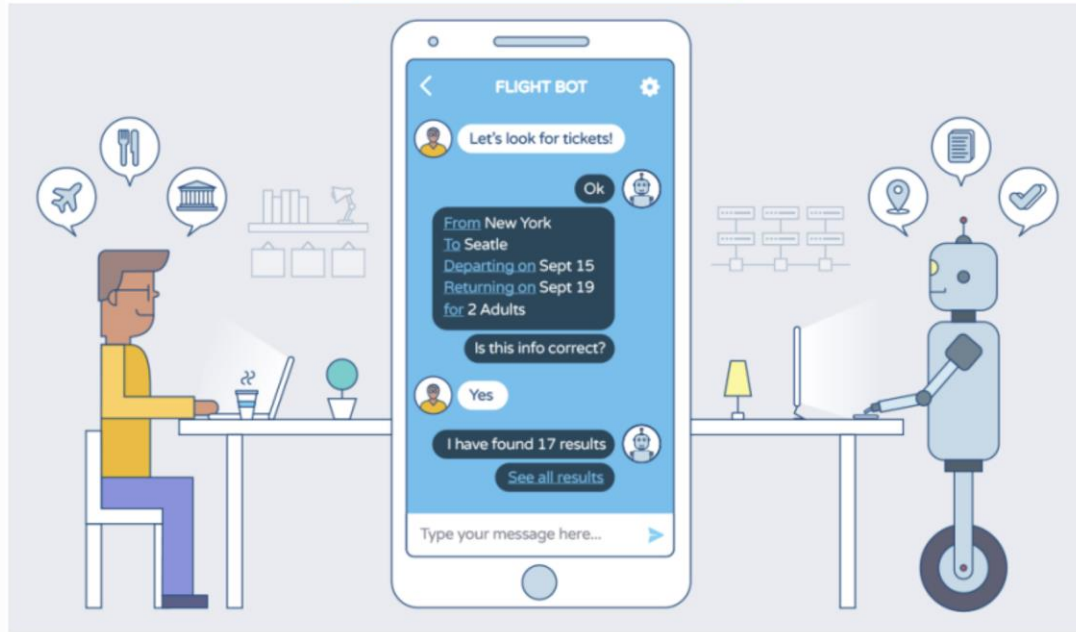
NLP applications: summarization



NLP applications: information extraction



NLP applications: dialogue systems/chatbots



NLP applications: question answering



 amazon alexa

“Alexa, who was President when Barack Obama was nine?”

“Alexa, how’s my commute?”

“Alexa, what’s the weather?”

“Alexa, did the 49ers win?”



- NLP core tasks

Core tasks and applications of NLP

CORE TASKS

text classification

language modeling

sequence labeling



APPLICATIONS

machine translation

chatbots

information retrieval

summarization

question answering

Text classification

- Input: a span of text
- Output: a label from a set of discrete options
- *Example:* sentiment analysis
 - *Text* -> {positive, neutral, negative}

Language modeling

- Input: a span of text, or no text at all
- Output: the next word
- *Example:* text generation for chatbots (ChatGPT)
 - *context text -> next word*

Sequence labeling

- Input: a span of text
- Output: a sequence of labels, one for each word (token)
- *Example*: part-of-speech tagging
 - *The book was brilliant* -> *DET NOUN VERB ADJ*

- Coding activity: clickbait classification

Clickbait classification on JupyterHub

- [Click on this nbgitpuller link](#)
- Open `session4_clickbait_classification.ipynb`