COMP 790.139 (Fall 2017)
Natural Language Processing
(with deep learning and connections to vision/robotics)

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Mohit Bansal

Class Info/Logistics

- COMP 790.139 'Natural Language Processing'
- **3 UNITS**
- Instructor: Mohit Bansal (SN258, http://www.cs.unc.edu/~mbansal/)
- Time: Wed 10.10am-12.40pm
- Room: FB008
- Office Hours: Wed 12.40-1.40pm (by appointment), SN258
- Course Webpage: http://www.cs.unc.edu/~mbansal/teaching/nlp-course-fall17.html
- **Course Email**: nlpcomp790unc@gmail.com

- Your email: check/fwd your connectcarolina xyz@email/live.unc.edu email for my welcome message and send me your preferred email id!

About Me

- Asst. Professor, CS, UNC (joined Fall 2016)
- Research Asst. Professor, TTI-Chicago, 2013-2016
- PhD, UC Berkeley, 2008-2013

- Research Interests:
  - Past: Syntactic parsing, coreference resolution, taxonomy induction, world knowledge and commonsense induction
  - Current: Multimodal and embodied semantics (i.e., language with vision and speech, for robotics); human-like language generation and Q&A/dialogue; interpretable and structured deep learning

- Office SN258

- Webpage: http://www.cs.unc.edu/~mbansal/, Email: mbansal@cs.unc.edu

Your Introductions

- Please say your:
  - Name
  - Department/degree/major
  - Research interests (Why NLP? Past ML/AI/NLP/CV experience? Coding experience?)
  - Fun fact ☺

About the Course (and its Goals)

- Research-oriented graduate course! We will cover lots of interesting chapters+papers, brainstorm, & do fun projects!
- We'll start with some basics of traditional NLP
- Then cover some specific, latest research topics, both based on traditional models and newer neural models (also some paper readings on certain topics, presented by students)
- Will also discuss connections of NLP with vision and robotics, and several deep learning for NLP models
- Brainstorm regularly and code + write up fun/novel projects!
- Some lecture(s) on academic/research quality paper writing

Prerequisites

- Advanced, graduate-level, research-based class
- Some machine learning and coding experience is definitely expected and required! (please talk to me class if you haven't yet discussed this with me over email)
- Homeworks, projects, summaries, and paper presentation will all require solid ML foundations/clarity and coding skills (e.g., linear algebra, diff eqns, logistic regression, supervised/unsup learning setups, classifiers, backprop and MLPs/NNs)
- Moreover, some basic NLP background is highly recommended

Expectations/Grading (tentative)

- Project reports and presentations (40%; midterm+final)
- Homework assignments (30%)
- Paper/Chapter presentation (10%)
- Paper/Chapter written summaries (10%)
- Class participation, discussion and brainstorming (10%)

Homeworks

- Coding based assignments
- On certain topics covered in class, e.g., parsing, Q&A, summarization/translation
- Code should be written from scratch (acknowledge any borrowed pieces)
- Preferably in tensorflow/python/pytorch

Project

- Students will pick (early) their favorite topic among latest cutting-edge research topics covered in class
- And will try a novel idea (implementing+extending or original) -- I am happy to discuss details!
- Midterm and final report + presentation (and possibly some updates)
- Might be in pairs/groups depending on final class size
- Use ACL conference style files and aim for high-quality project write-ups
- Will have some lecture(s) on research-quality paper writing

Paper Presentation

- Lead discussion for 1-2 chapters/papers on a topic some week (may be done in pairs/groups depending on class size)
- Read related chapters/papers and present background to audience
- Present task and ML details of given chapter/papers
- Present demo's of related code, etc.
- Ask interesting questions to initiate brainstorming
- Mention some next steps, future work, extension ideas!

Chapter/Paper Written Summaries

- 0.5-1 page (per paper) write-up for certain week's chapters/papers
- Describe the task
- Summarize the methods/models
- Explain the novelty
- Discuss the next steps or potential improvements

Class Participation and Brainstorming

- Audience students expected to take part in lively discussion in every class and after chapter/paper reading!

- Semi-regularly (i.e., after completing several chapters/papers in 2-3 weeks), we will have a brainstorming and 'idea-generation' session!

- Exact details to be announced soon but students expected to submit and discuss novel idea(s) on the whole general topic, e.g., new related task or dataset, new approach to existing task, combinations of tasks/approaches, etc.

- Don't hesitate to propose fancy ideas ☺, but try to keep them grounded/feasible and think of how to approach them realistically (in terms of datasets, models, speed, memory, etc.)

Reference Books

- **SLP2**: D. Jurafsky & James H. Martin. “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition”. Prentice Hall, Second Edition, 2009.

- **SLP3**: Some draft chapters of the third edition are available online at https://web.stanford.edu/~jurafsky/slp3/

- **FSNLP**: Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999. http://nlp.stanford.edu/fsnlp/

- **ML Background**: Andrew Ng’s Coursera Machine Learning course https://www.coursera.org/learn/machine-learning

- Stanford NLP + Deep Learning Class: http://web.stanford.edu/class/cs224n/

Course Syllabus/Topics (tentative)

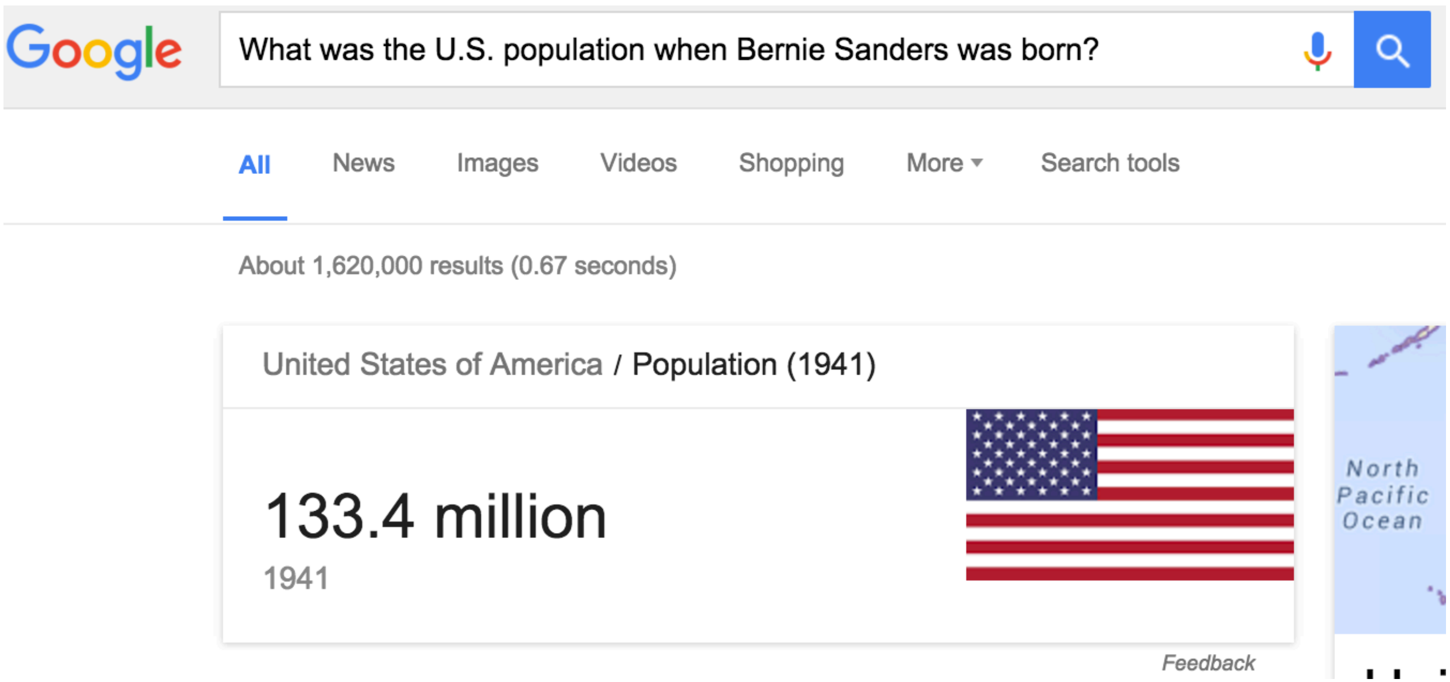
- Language Modeling
- Part-of-speech Tagging
- Syntactic Parsing: Constituent, Dependency, CCG, others
- Coreference Resolution
- Distributional Semantics: PMI, neural, CCA
- Compositional Semantics: Logical-form, Semantic Parsing, Vector-form, neural (RNNs/CNNs)
- Question Answering: Factoid-based, Passage-based
- Sentiment Analysis
- Document Summarization
- Machine Translation
- Dialogue Models
- Language and Vision: Image Captioning, Video Captioning, Visual Question Answering
- Language and Robotics: Instructions for Navigation, Manipulation, Skill Learning; Human-Robot Interaction
- Several interesting machine and deep learning models all along the way, e.g., deep+structured models, interpretable models, adversarial models, reward-based models (reinforcement learning)

What is NLP?

- Question answering

What is NLP?

- Question answering



What is NLP?

- Question answering

Google

which countries border the black sea

All Maps Images News Shopping More Search tools

About 2,710,000 results (0.81 seconds)

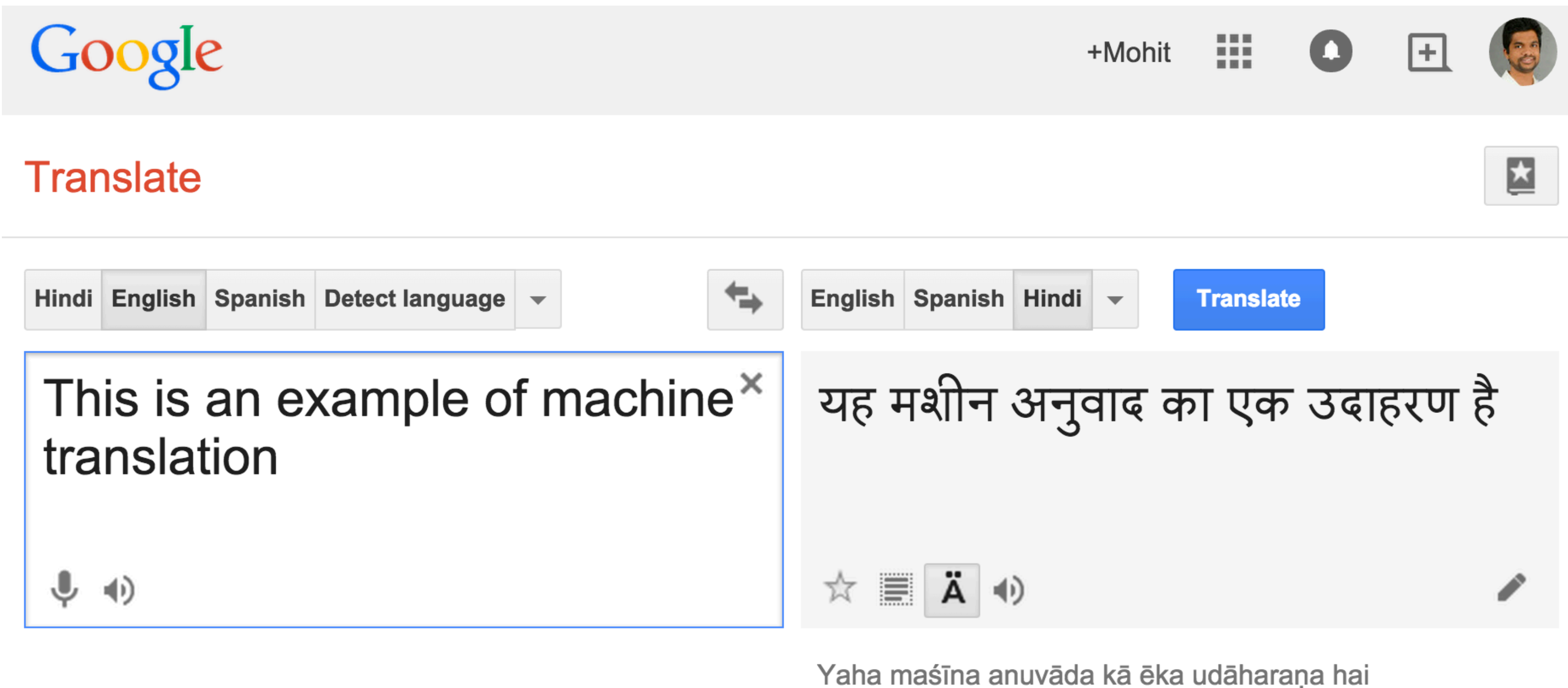
This major inland sea is bordered by six countries — **Romania** and **Bulgaria** to the west; **Ukraine**, **Russia**, and **Georgia** to the north and east; and **Turkey** to the south. Additionally, it is impacted by another 10 nations through the five major rivers that empty into the Black Sea, the largest of which is the Danube River.

Black Sea Geography - College of Earth, Ocean, and Environment
https://www.ceoe.udel.edu/**blacksea**/geography/index.html University of Delaware

About this result • Feedback

What is NLP?

- Machine Translation



What is NLP?

- Sentiment Analysis

Sentiment Analysis with Python NLTK Text Classification

This is a demonstration of **sentiment analysis** using a `NLTK 2.0.4` powered **text classification** process. It can tell you whether it thinks the text you enter below expresses **positive sentiment**, **negative sentiment**, or if it's **neutral**. Using **hierarchical classification**, *neutrality* is determined first, and *sentiment polarity* is determined second, but only if the text is not neutral.

Analyze Sentiment

Language

english

Enter text

It always amazes me how Universal never cares to create anything remotely clever when it comes to their animations, and so once again they come up with a harmless little story that wants to be cute and funny (which it is sometimes) but is only bound to be quickly forgotten.

Enter up to 50000 characters

Analyze

Sentiment Analysis Results

The text is **neg**.

The final sentiment is determined by looking at the classification probabilities below.

Subjectivity

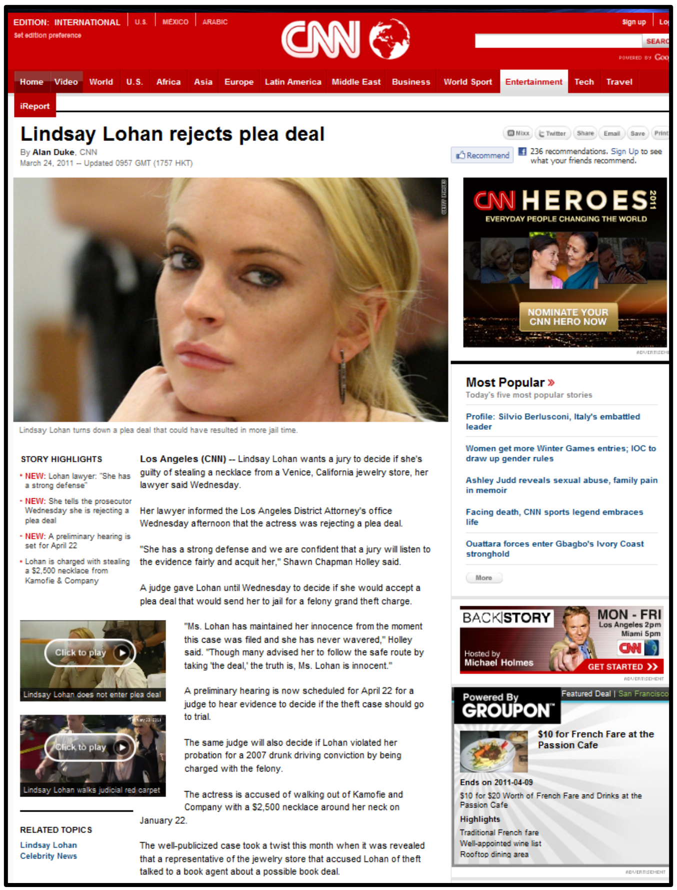
- neutral: 0.3
- **polar: 0.7**

Polarity

- pos: 0.2
- **neg: 0.8**

What is NLP?

- Natural Language Generation: Summarization



EDITION: INTERNATIONAL | U.S. | MÉXICO | ARABIC
Set edition preference

Sign up | Log

Home | Video | World | U.S. | Africa | Asia | Europe | Latin America | Middle East | Business | World Sport | Entertainment | Tech | Travel | iReport

Lindsay Lohan rejects plea deal

By Alan Duke, CNN
March 24, 2011 -- Updated 0957 GMT (1757 HKT)

236 recommendations. Sign Up to see what your friends recommend.

Lindsay Lohan turns down a plea deal that could have resulted in more jail time.

STORY HIGHLIGHTS

- NEW: Lohan lawyer: "She has a strong defense"
- NEW: She tells the prosecutor Wednesday she is rejecting a plea deal
- NEW: A preliminary hearing is set for April 22
- Lohan is charged with stealing a $2,500 necklace from Kamofie & Company

Los Angeles (CNN) -- Lindsay Lohan wants a jury to decide if she's guilty of stealing a necklace from a Venice, California jewelry store, her lawyer said Wednesday.

Her lawyer informed the Los Angeles District Attorney's office Wednesday afternoon that the actress was rejecting a plea deal.

"She has a strong defense and we are confident that a jury will listen to the evidence fairly and acquit her," Shawn Chapman Holley said.

A judge gave Lohan until Wednesday to decide if she would accept a plea deal that would send her to jail for a felony grand theft charge.

"Ms. Lohan has maintained her innocence from the moment this case was filed and she has never wavered," Holley said. "Though many advised her to follow the safe route by taking 'the deal,' the truth is, Ms. Lohan is innocent."

A preliminary hearing is now scheduled for April 22 for a judge to hear evidence to decide if the theft case should go to trial.

The same judge will also decide if Lohan violated her probation for a 2007 drunk driving conviction by being charged with the felony.

The actress is accused of walking out of Kamofie and Company with a $2,500 necklace around her neck on January 22.

The well-publicized case took a twist this month when it was revealed that a representative of the jewelry store that accused Lohan of theft talked to a book agent about a possible book deal.

Click to play
Lindsay Lohan does not enter plea deal

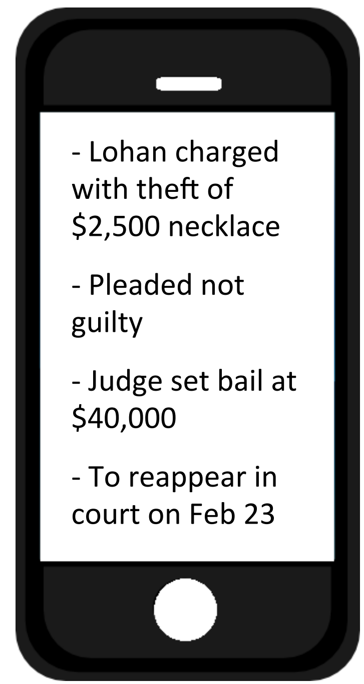
Click to play
Lindsay Lohan walks judicial red carpet

RELATED TOPICS
Lindsay Lohan
Celebrity News

Most Popular »
Today's five most popular stories

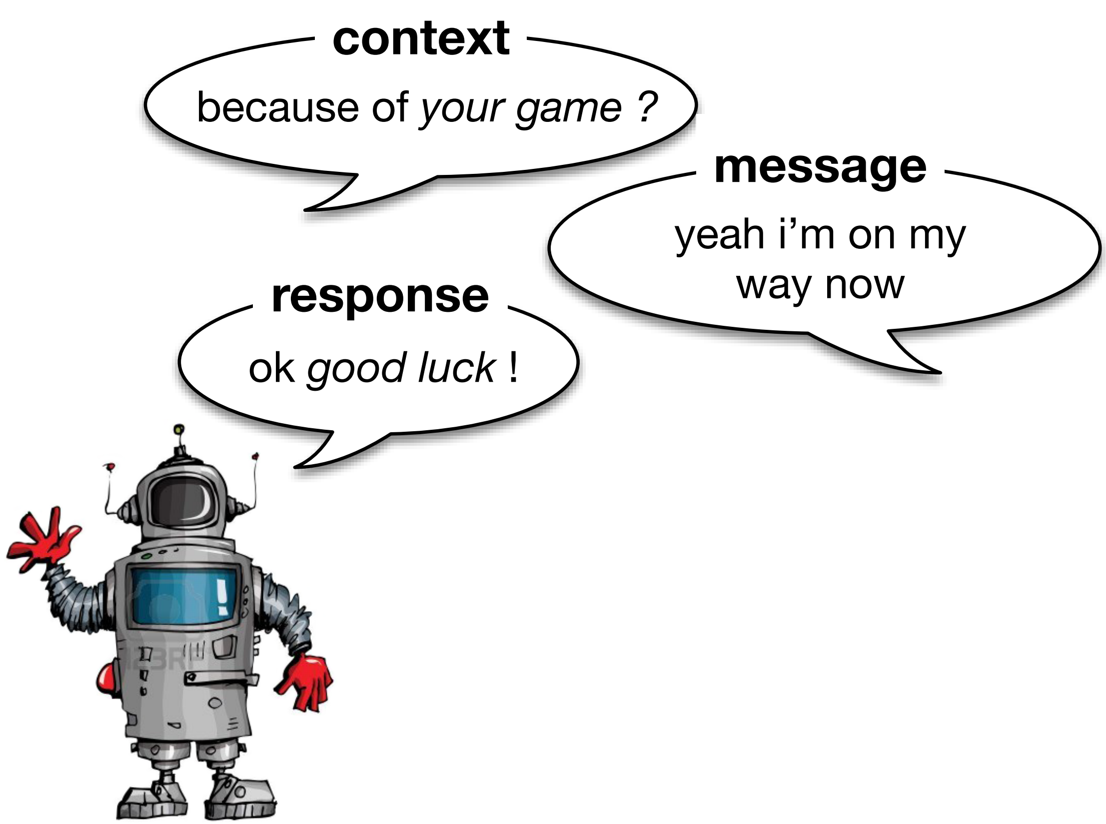
- Profile: Silvio Berlusconi, Italy's embattled leader
- Women get more Winter Games entries; IOC to draw up gender rules
- Ashley Judd reveals sexual abuse, family pain in memoir
- Facing death, CNN sports legend embraces life
- Ouattara forces enter Gbagbo's Ivory Coast stronghold

More



What is NLP?

- Natural Language Generation: Conversation/Dialogue



[Sordoni et al., 2015]

What is NLP?

- Natural Language Generation: Image Captioning

www.cs.toronto.edu/~nitish/nips2014demo/results/8832804.html

Results

Tags

- authors
- scones
- luncheon
- breakfast
- seder

Nearest Caption in the Training Dataset

a man cuts a cake while children sit around at the table , looking on .

Generated Captions

- two people at a table with a cake .
- the two people are having a meal that is in a party .
- a man and two children in a blue table with a cake .
- a man sitting at a table with a bunch of cake on it .
- a man and woman sitting at a table with cake at a party .

back

[UToronto]

What is NLP?

- Natural Language Generation: Video Captioning

Ground truth: Two women are shopping in a store.
Two girls are shopping.
Baseline model: A man is doing a monkey in a store.
Multi-task model: A woman is shopping in a store.

Ground truth: Two men are fighting.
A group of boys are fighting.
Baseline model: A group of men are dancing.
Multi-task model: Two men are fighting.

Ground truth: A man is playing a violin.
A man is playing the violin on stage.
Baseline-XE: A man is playing the drums.
CIDEr-RL: A man is playing a guitar.
CIDEnt-RL: A man is playing a violin.

Ground truth: Two men are wrestling.
Two guys are wrestling in a competition.
Baseline-XE: A group of people are playing a game.
CIDEr-RL: A man is playing a wrestling.
CIDEnt-RL: Two men are wrestling.

Ground truth: A person is playing a video game.
Someone is playing video game.
Baseline-XE: A man is riding a motorcycle.
CIDEr-RL: A man is talking about a plane.
CIDEnt-RL: A person is playing a video game.

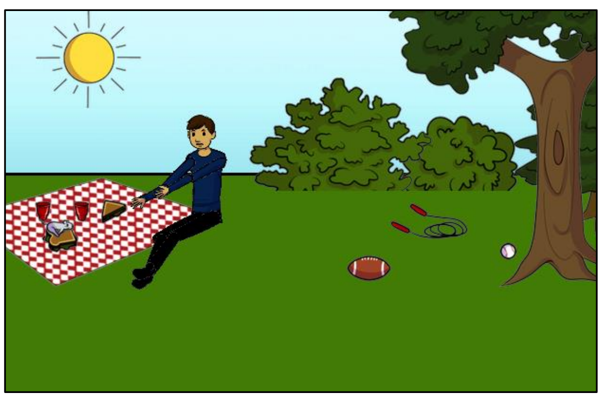
[Pasunuru and Bansal, 2017]

What is NLP?

- Natural Language Generation: Visual Question Answering

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?



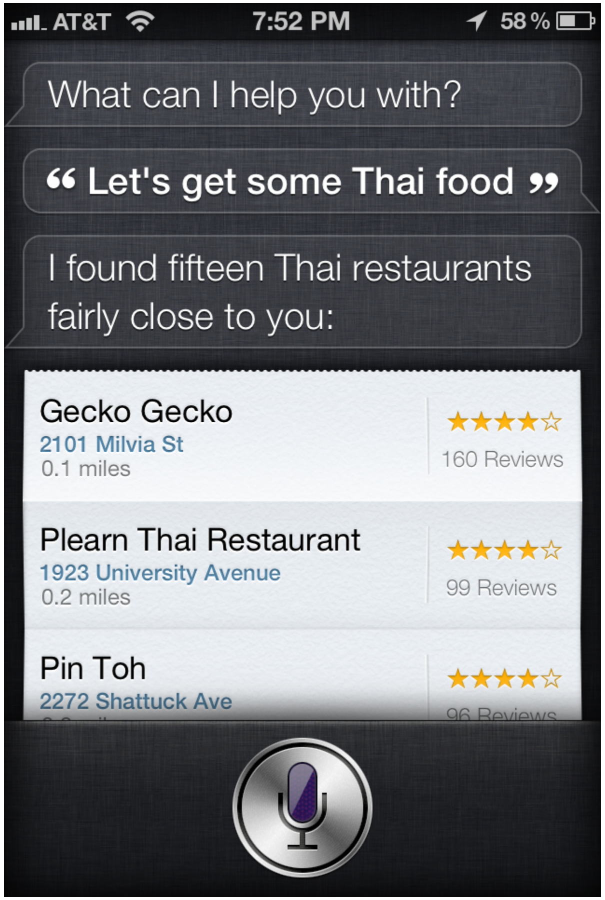
Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?

[Antol et al., 2015]

What is NLP?

- Automatic Speech Recognition



Some Exciting NLP Challenges

Human-like Ambiguous Language

- Non-literal: Idioms, Metaphors

You: *I am under the weather today.*

Siri: *Here is the weather today… 50 F*

Human-like Ambiguous Language

- Non-literal: Idioms, Metaphors

Human-like Ambiguous Language

- Humor, Sarcasm, Politeness/Rudeness

Human-like Ambiguous Language

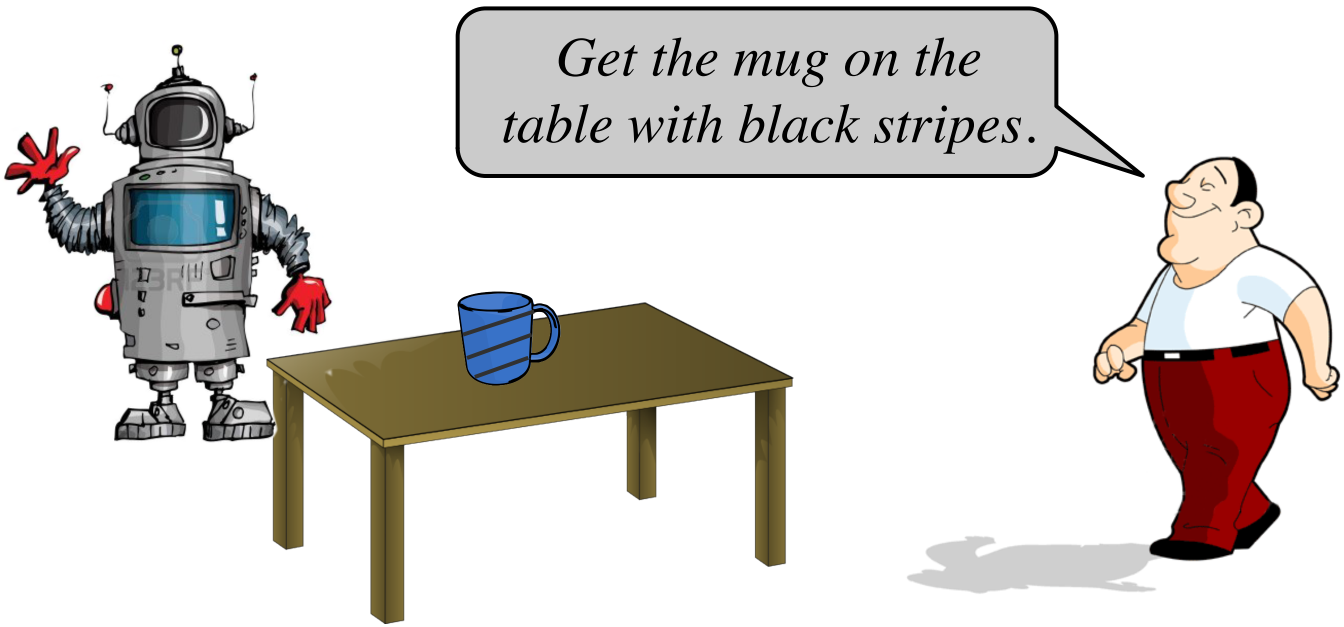
- Prepositional Attachment, Coreference Ambiguities

Human-like Ambiguous Language

- Prepositional Attachment, Coreference Ambiguities

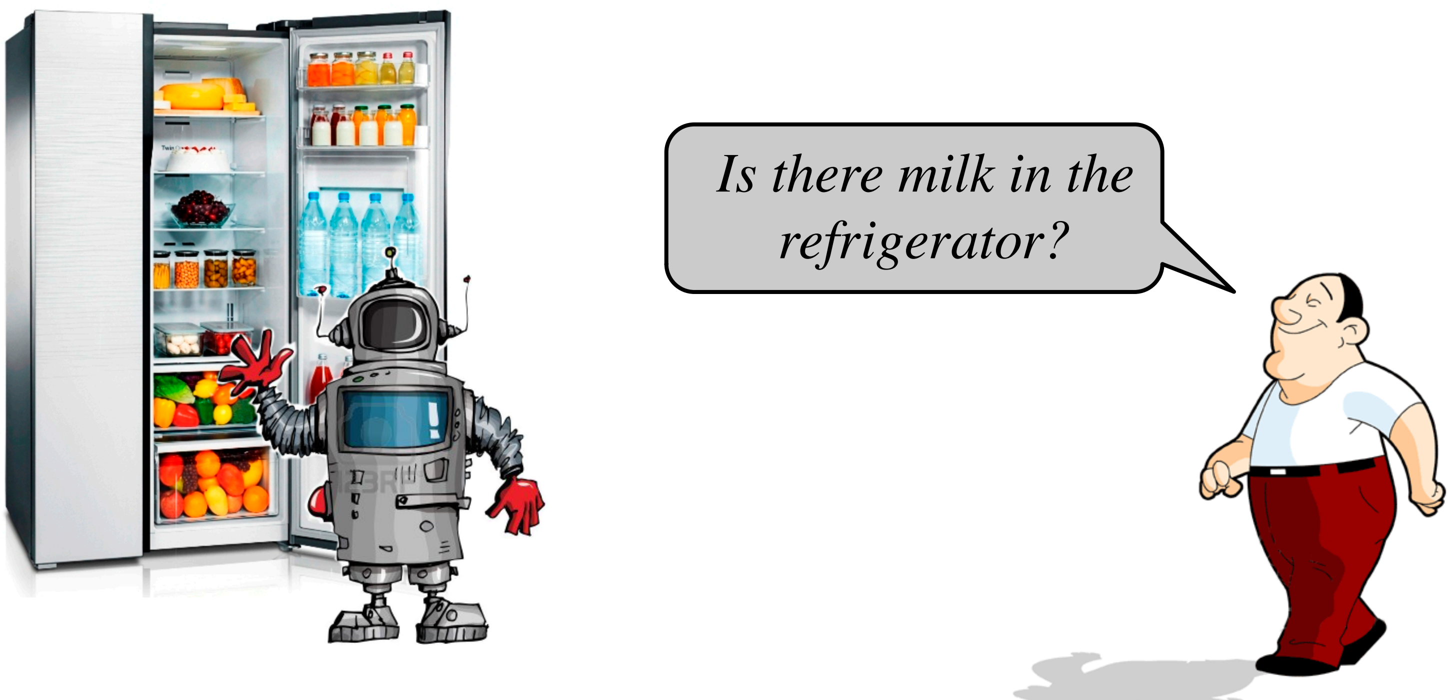
Visually Grounded Language

- Text-Image Alignment: Most of our daily communication language points to several objects in the visual world



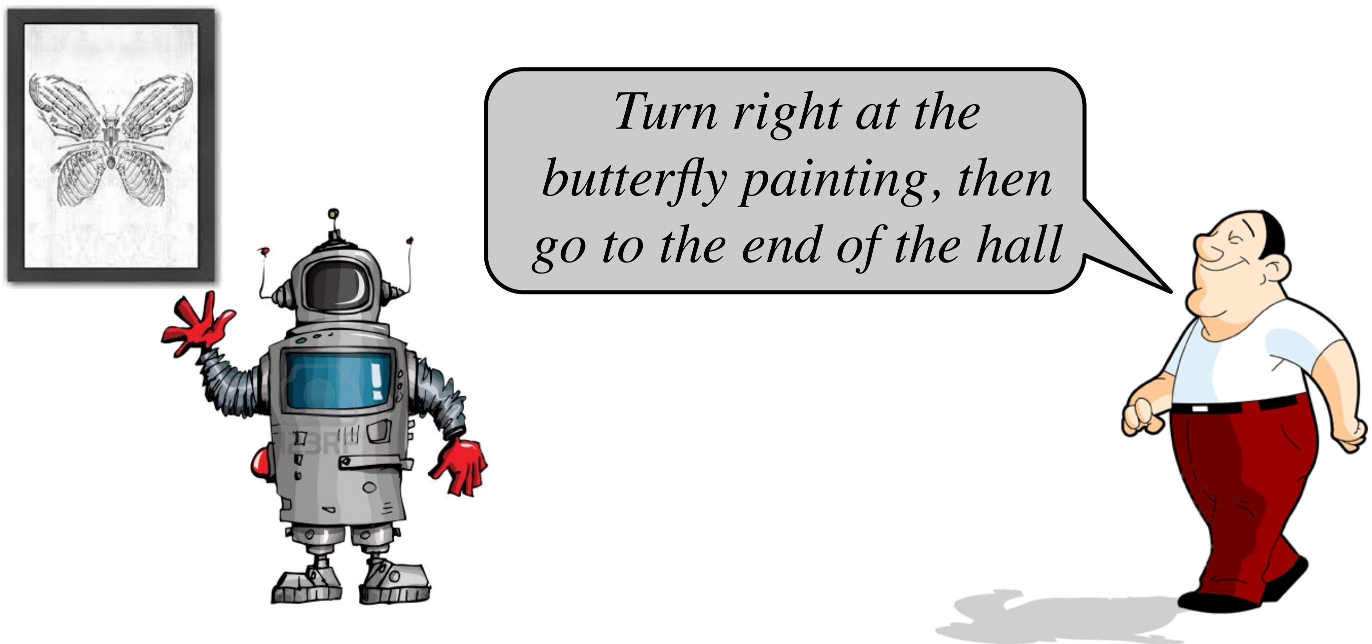
Visually Grounded Language

- Visual Question Answering: Humans asking machines about pictures/videos, e.g., for visually impaired, in remote/ dangerous scenarios, in household service settings



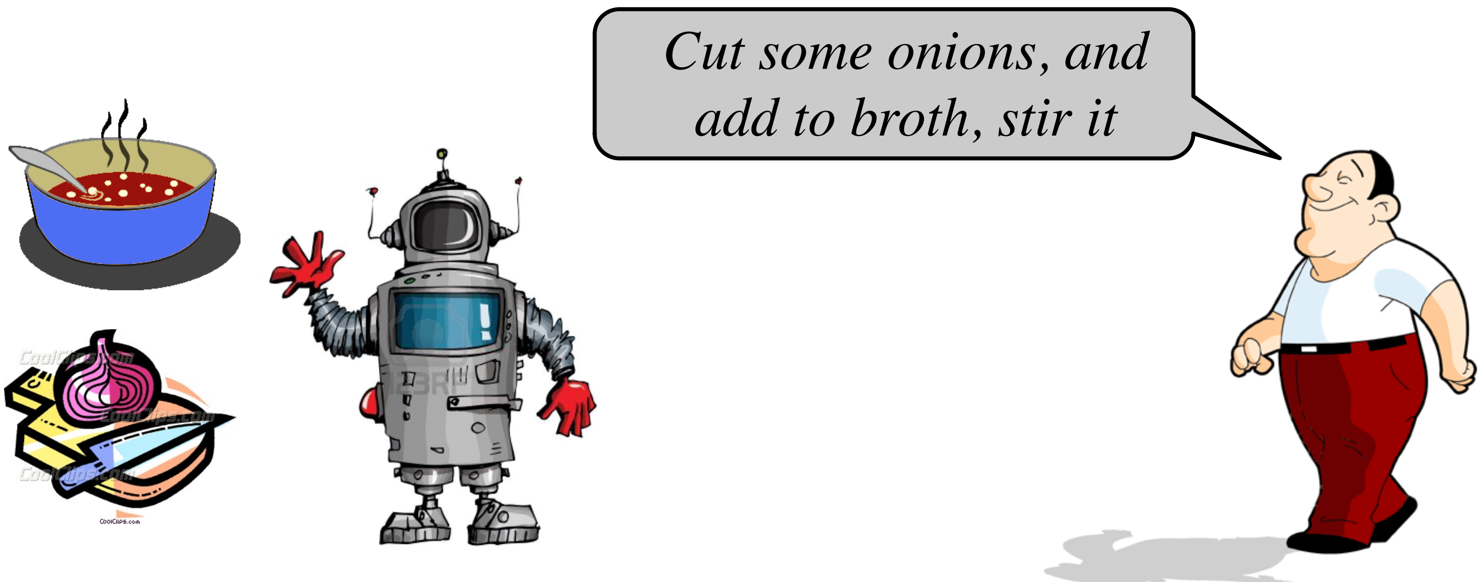
Embodied Language (Robot Instructions)

- Task-based instructions, e.g., navigation, grasping, manipulation, skill learning



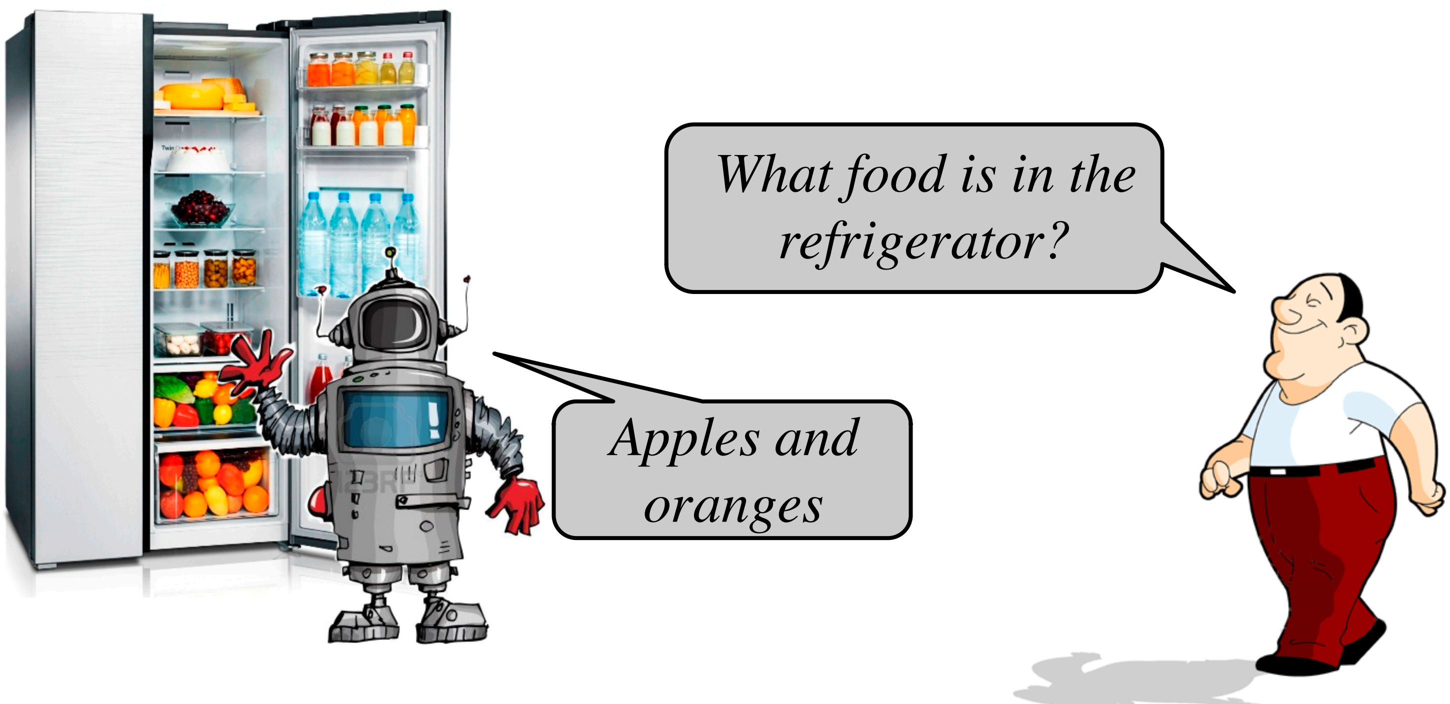
Embodied Language (Robot Instructions)

- Task-based instructions, e.g., navigation, grasping, manipulation, skill learning



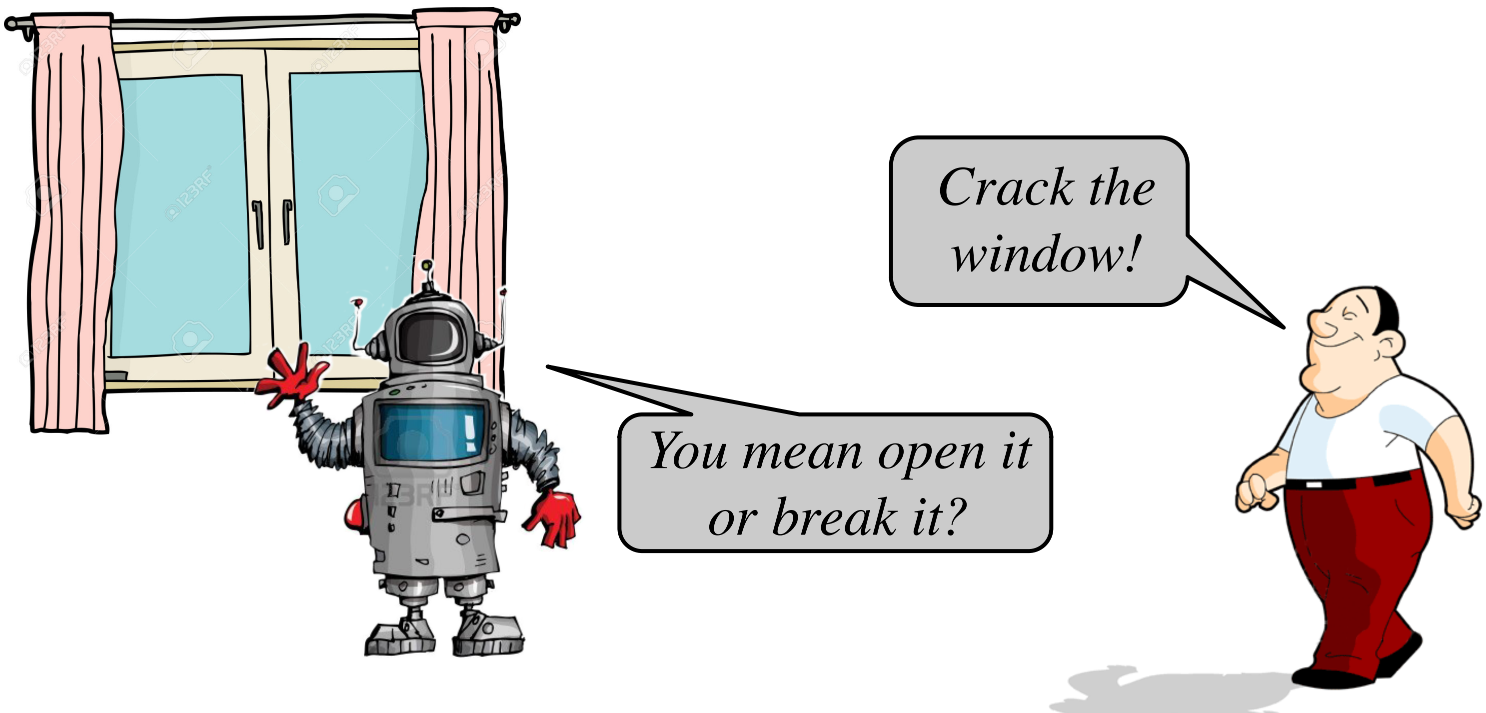
Grounded Language Generation/Dialogue

- Both for answering human questions, and to ask questions back, and for casual chit-chat



Grounded Language Generation/Dialogue

- Both for answering human questions, and to ask questions back, and for casual chit-chat



10-min break?

Language Modeling and Generation

(some slides adapted/borrowed from courses by Dan Klein, Chris Manning, Richard Socher)

Language Modeling

- A language model is a distribution over sequences of words (sentences)

$$P(\mathbf{w}) = P(w_{1} \ldots w_{n})$$

- Purpose is to usually assign high weights to plausible sentences, e.g., in speech recognition or machine translation

- Also used for language generation now (predict next word given previous words), esp. w/ new RNN models

Traditional N-gram LMs

- Use chain rule to generate words left-to-right

$$P(w_1 \ldots w_n) = \prod_i P(w_i | w_1 \ldots w_{i-1})$$

- Can't condition on the entire left context

P(??? | Turn to page 134 and look at the picture of the)

- N-gram models make a Markov assumption

$$P(w_1 \ldots w_n) = \prod_i P(w_i | w_{i-k} \ldots w_{i-1})$$

$$P(\text{please close the door}) =$$

$$P(\text{please}|\text{START}) P(\text{close}|\text{please}) \ldots P(\text{STOP}|door)$$

Traditional N-gram LMs

- How do we know P(w | history)?
  - Use statistics from data (examples using Google N-Grams)
  - E.g. what is P(door | the)?

Training Counts

```
198015222 the first
194623024 the same
168504105 the following
158562063 the world

14112454 the door
-----------------
23135851162 the *
```

$$\hat{P}(\mathrm{door}|\mathrm{the}) = \frac{14112454}{23135851162}$$

$$= 0.0006$$

  - This is the *maximum likelihood* estimate

Sparsity Issue & Parameter Estimation

- New words all the time (*antidisestablishmentarianism, kakorrhaphiophobia,, www.xyzabc156.com)*….worse for new bigrams and trigrams!
  - Maximum likelihood estimates won't get us very far

$$\hat{P}(w|w_{-1}) = \frac{c(w_{-1}, w)}{\sum_{w'} c(w_{-1}, w')}$$

3516 wipe off the excess
1034 wipe off the dust
547 wipe off the sweat
518 wipe off the mouthpiece
…
120 wipe off the grease
0 wipe off the sauce
0 wipe off the mice

28048 wipe off the *

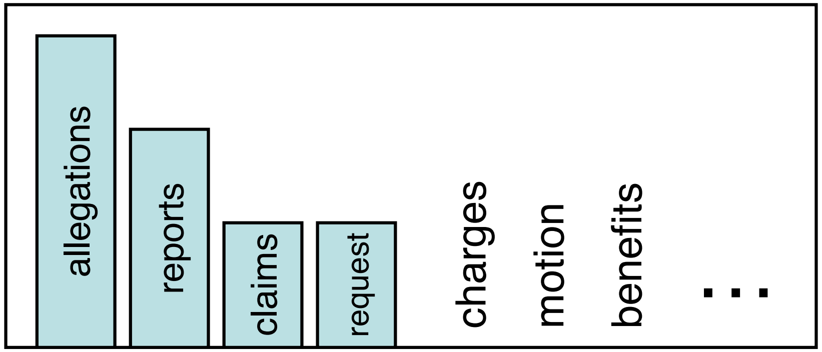
  - Need to *smooth* these estimates
  - General method (procedurally)
    - Take your empirical counts
    - Modify them in various ways to improve estimates
  - General method (mathematically)
    - Often can give estimators a formal statistical interpretation
    - … but not always
    - Approaches that are mathematically obvious aren't always what works

Smoothing Techniques

- We often want to make estimates from sparse statistics:

P(w | denied the)
3 allegations
2 reports
1 claims
1 request

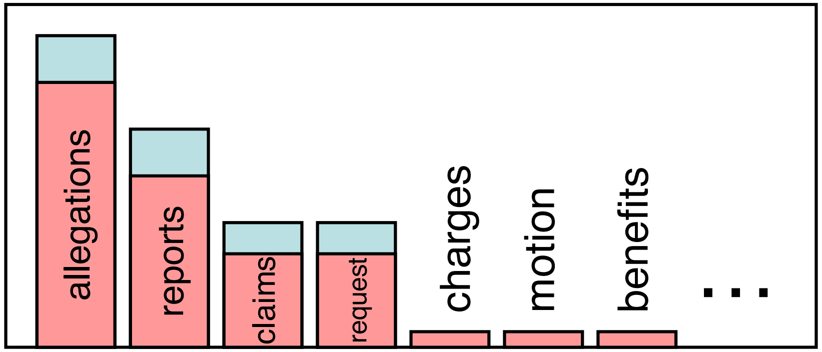
7 total



- Smoothing flattens spiky distributions so they generalize better

P(w | denied the)
2.5 allegations
1.5 reports
0.5 claims
0.5 request
2 other

7 total



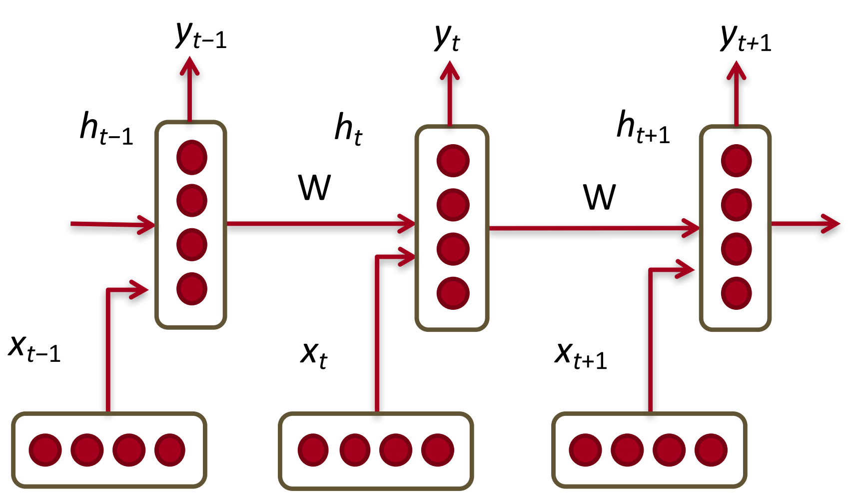
- Very important all over NLP, but easy to do badly!

Smoothing Techniques

- Classic Solution: add-one or add small priors to numer/denom
- Backing off to smaller n-grams
- Held-out Reweighting: Important to optimize/estimate how models generalize! So use held-out data to estimate the map of old count to new count
- Kneser-Ney Discounting: two successful ideas:
  - Idea 1: observed n-grams occur more in training than they will later
  - Idea 2: Type-based fertility (based on how common the word type is)
- Read Chen and Goodman, 1996 for various details and graphs!

RNN Language Models

- Avoid huge number of n-grams; Memory requirement only scales with #words
- Can condition on all previous history (with forget gates)
- Loss function on identity of predicted word at each time step
- But harder/slower to train and reach optimum (and less interpretable)?



Distributional Semantics

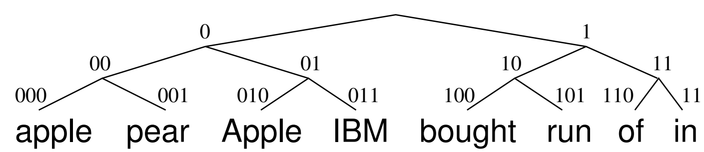
- Words occurring in similar context have similar linguistic behavior (meaning) [Harris, 1954; Firth, 1957]

$$\textit{food} \Rightarrow \begin{bmatrix} 0.6 \\ -0.2 \\ 0.9 \\ 0.3 \\ -0.4 \\ 0.5 \end{bmatrix}$$

- Traditional approach: context-counting vectors
  - Count left and right context in window
  - Reweight with PMI or LLR
  - Reduce dimensionality with SVD or NNMF

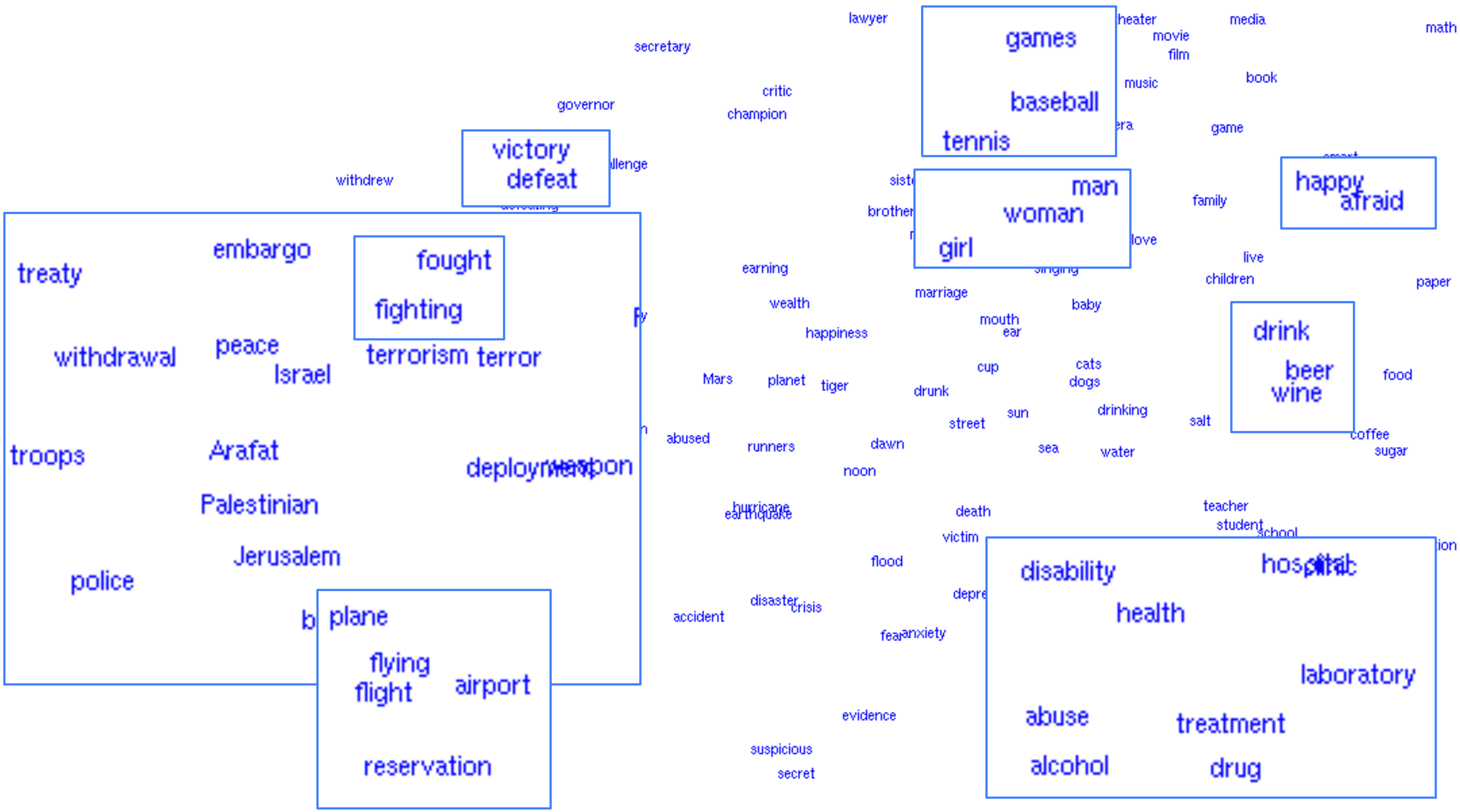
  [Pereira et al., 1993; Lund & Burgess, 1996; Lin, 1998; Lin and Pantel, 2001; Sahlgren, 2006; Pado & Lapata, 2007; Turney and Pantel, 2010; Baroni and Lenci, 2010]

- More word representations: hierarchical clustering based on bigram LM LL
  [Brown et al., 1992]



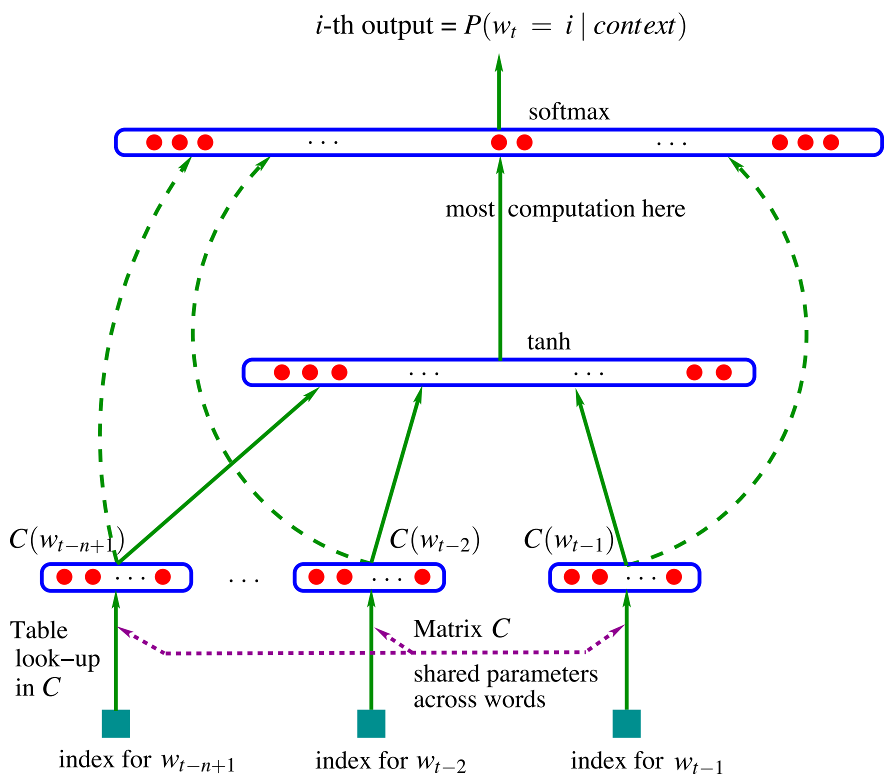
Unsupervised Embeddings

- Vector space representations learned on unlabeled linear context (i.e., left/right words): distributional semantics (Harris, 1954; Firth, 1957)



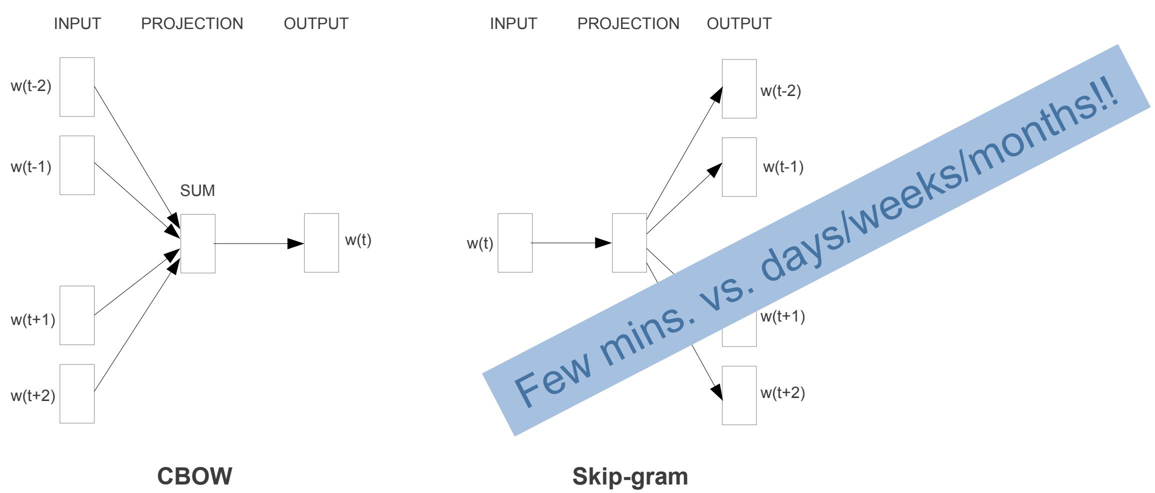
Distributional Semantics -- NNs

- Newer approach: context-predicting vectors (NNs)
  - SENNA [Collobert and Weston, 2008; Collobert et al., 2011]: Multi-layer DNN w/ ranking-loss objective; BoW and sentence-level feature layers, followed by std. NN layers. Similar to [Bengio et al., 2003].



Distributional Semantics -- NNs

- CBOW, SKIP, word2vec [Mikolov et al., 2013]: Simple, super-fast NN w/ no hidden layer. Continuous BoW model predicts word given context, skip-gram model predicts surrounding context words given current word



  - Other: [Mnih and Hinton, 2007; Turian et al., 2010]

- Demos: https://code.google.com/p/word2vec, http://metaoptimize.com/projects/wordreprs/, http://ml.nec-labs.com/senna/

Distributional Semantics

- Other approaches: spectral methods, e.g., CCA
  - Word-context correlation [Dhillon et al., 2011, 2012]
  - Multilingual correlation [Faruqui and Dyer, 2014; Lu et al., 2015]

- Some later ideas: Train task-tailored embeddings to capture specific types of similarity/semantics, e.g.,
  - Dependency context [Bansal et al., 2014, Levy and Goldberg, 2014]
  - Predicate-argument structures [Hashimoto et al., 2014; Madhyastha et al., 2014]
  - Lexicon evidence (PPDB, WordNet, FrameNet) [Xu et al., 2014; Yu and Dredze, 2014; Faruqui et al., 2014; Wieting et al., 2015]
  - Combining advantages of global matrix factorization and local context window methods [GloVe; Pennington et al., 2014]

Compositional Semantics with NNs

- Composing, combining word vectors to representations for longer units: phrases, sentences, paragraphs, …
- Initial approaches: point-wise sum, multiplication [Mitchell and Lapata, 2010; Blacoe and Lapata, 2012]
- Vector-matrix compositionality [Baroni and Zamparelli, 2010; Zanzotto et al., 2010; Grefenstette and Sadrzadeh, 2011; Socher et al., 2011; Yessenalina and Cardie, 2011]
- Linguistic information added via say parses in RvNNs [Socher et al., 2011b, 2012, 2013a, 2013b, 2014; Hermann and Blunsom, 2013]
- Sequential RNNs (with GRU/LSTM gates)

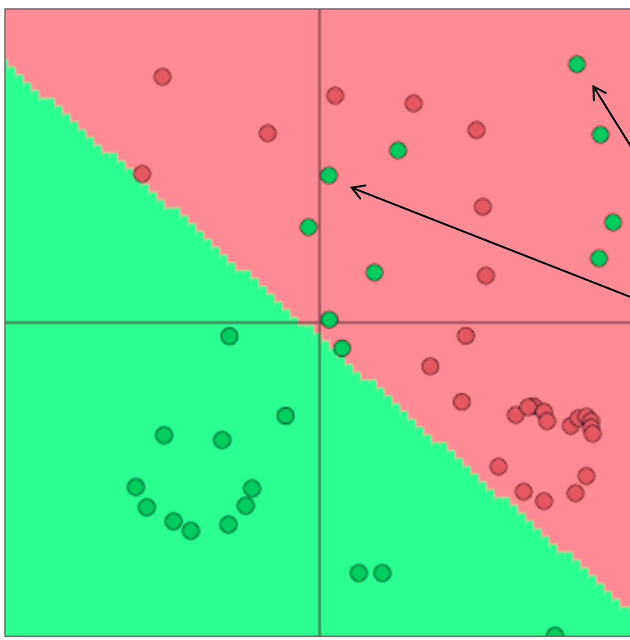
  (Simple vector averaging w/ updating sometimes competitive)

Compositional Semantics with NNs

- Feed-forward NNs with back-propagation

Softmax (= logistic regression) is not very powerful

- Softmax only linear decision boundaries



→ Lame when problem is complex

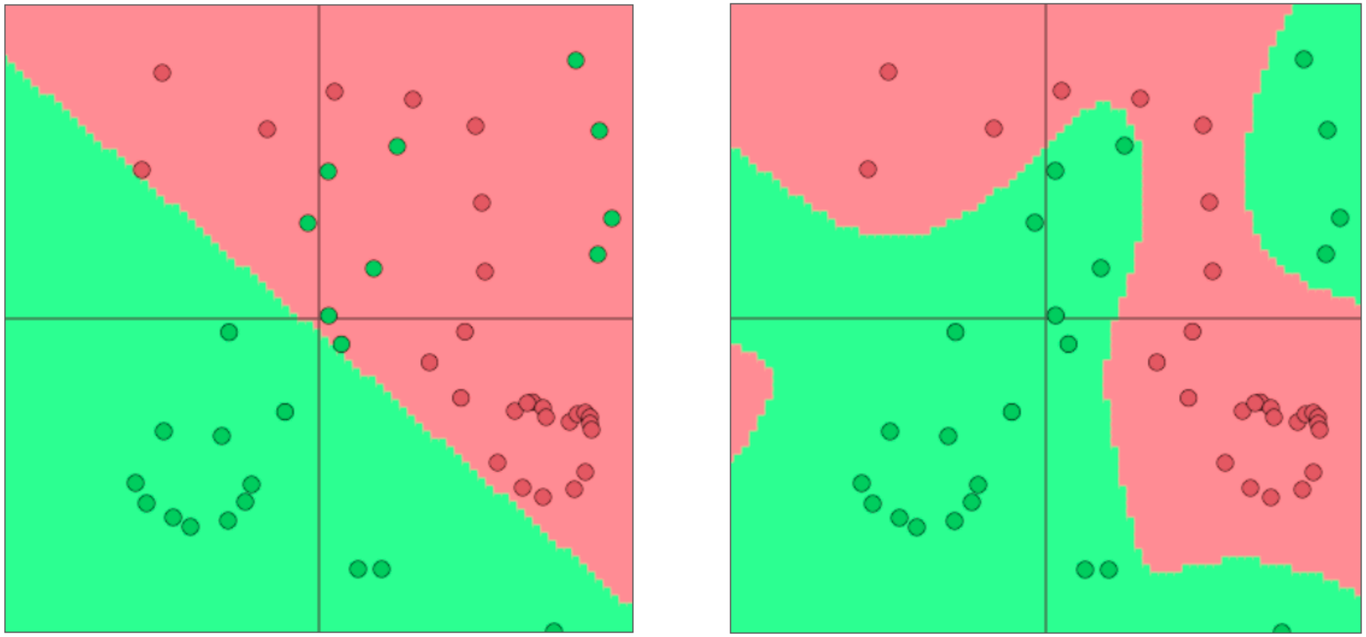
Wouldn't it be cool to get these correct?

Compositional Semantics with NNs

- Feed-forward NNs with back-propagation

Neural Nets for the Win!

- Neural networks can learn much more complex functions and nonlinear decision boundaries!



Compositional Semantics with NNs

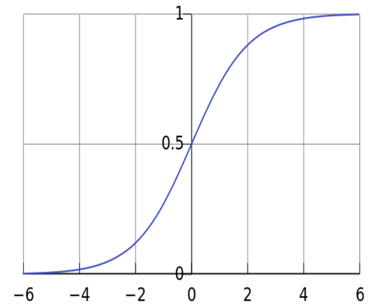
- Feed-forward NNs with back-propagation

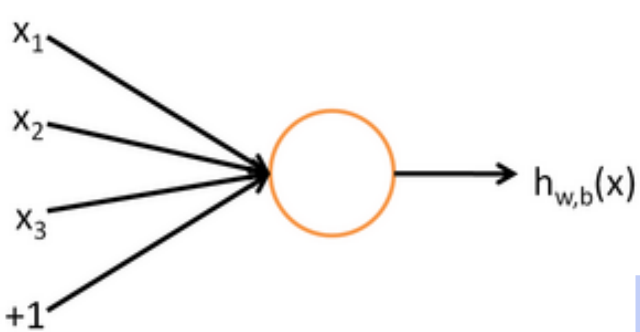
A neuron is essentially a binary logistic regression unit

$$h_{w,b}(x) = f(w^{\mathsf{T}}x + b)$$

b: We can have an "always on" feature, which gives a class prior, or separate it out, as a bias term

$$f(z) = \frac{1}{1 + e^{-z}}$$





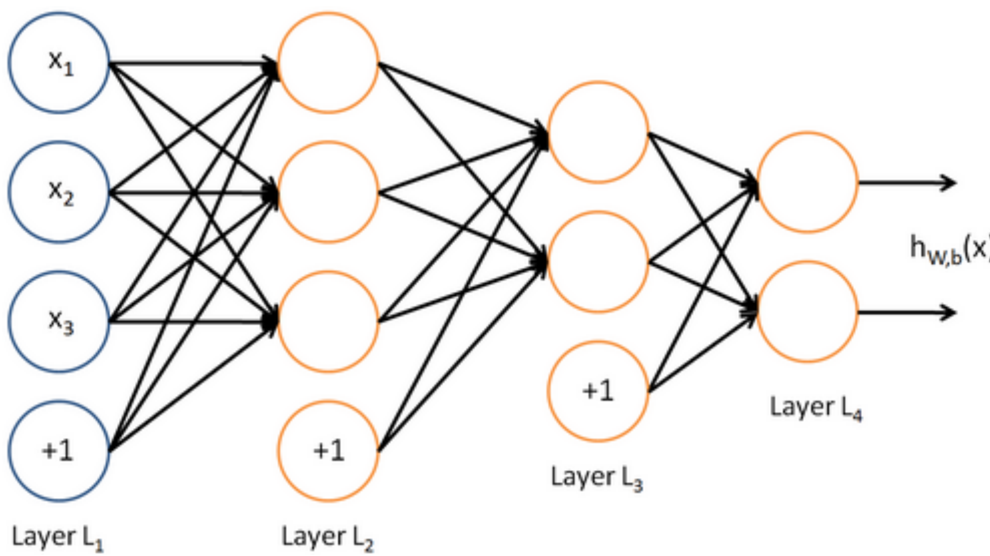
w, *b* are the parameters of this neuron
i.e., this logistic regression model

Compositional Semantics with NNs

- Feed-forward NNs with back-propagation

A neural network
= running several logistic regressions at the same time

Before we know it, we have a multilayer neural network….



Compositional Semantics with NNs

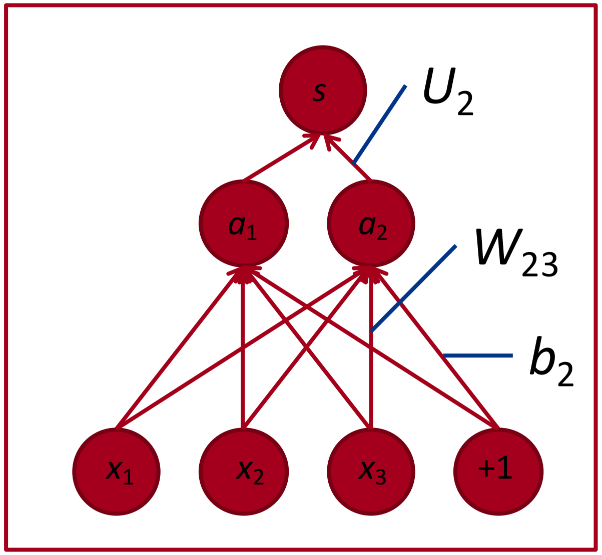
- Feed-forward NNs with back-propagation

Training with Backpropagation

- Let's consider the derivative of a single weight W_{ij}

$$\frac{\partial s}{\partial W} = \frac{\partial}{\partial W} U^T a = \frac{\partial}{\partial W} U^T f(z) = \frac{\partial}{\partial W} U^T f(Wx + b)$$

- This only appears inside a_i
- For example: W_{23} is only used to compute a_2



Compositional Semantics with NNs

- Feed-forward NNs with back-propagation

Training with Backpropagation

$$\frac{\partial s}{\partial W} = \frac{\partial}{\partial W} U^T a = \frac{\partial}{\partial W} U^T f(z) = \frac{\partial}{\partial W} U^T f(Wx+b)$$

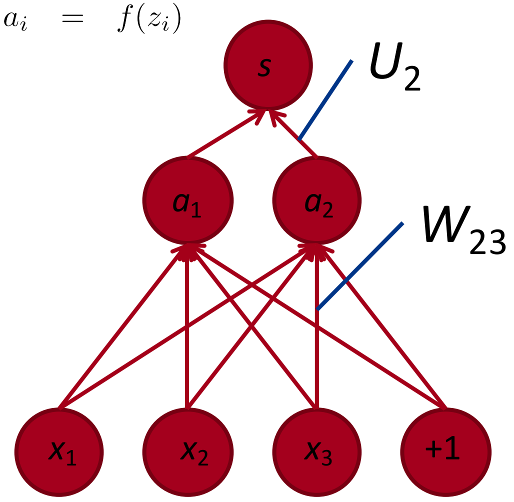
$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial x}$$

Derivative of weight W_{ij}:

$$\begin{aligned} z_i &= W_{i\cdot}x + b_i = \sum_{j=1}^{3} W_{ij}x_j + b_i \\ a_i &= f(z_i) \end{aligned}$$

$$\frac{\partial}{\partial W_{ij}} U^T a \rightarrow \frac{\partial}{\partial W_{ij}} U_i a_i$$

$$\begin{aligned} U_i \frac{\partial}{\partial W_{ij}} a_i &= U_i \frac{\partial a_i}{\partial z_i} \frac{\partial z_i}{\partial W_{ij}} \\ &= U_i \frac{\partial f(z_i)}{\partial z_i} \frac{\partial z_i}{\partial W_{ij}} \\ &= U_i f'(z_i) \frac{\partial z_i}{\partial W_{ij}} \\ &= U_i f'(z_i) \frac{\partial W_{i\cdot}x + b_i}{\partial W_{ij}} \end{aligned}$$



Compositional Semantics with NNs

- Feed-forward NNs with back-propagation

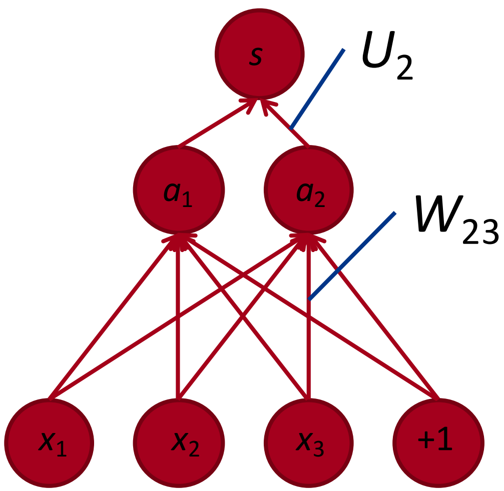
Training with Backpropagation

Derivative of single weight W_{ij} :

$$
\begin{aligned}
z_i &= W_{i\cdot}x + b_i = \sum_{j=1}^{3} W_{ij}x_j + b_i \\
a_i &= f(z_i)
\end{aligned}
$$

$$
\begin{aligned}
U_i \frac{\partial}{\partial W_{ij}} a_i &= U_i f'(z_i) \frac{\partial W_{i\cdot}x + b_i}{\partial W_{ij}} \\
&= U_i f'(z_i) \frac{\partial}{\partial W_{ij}} \sum_k W_{ik}x_k \\
&= \underbrace{U_i f'(z_i)} x_j \\
&= \underbrace{\delta_i}_{\text{Local error signal}} \quad \underbrace{x_j}_{\text{Local input signal}}
\end{aligned}
$$

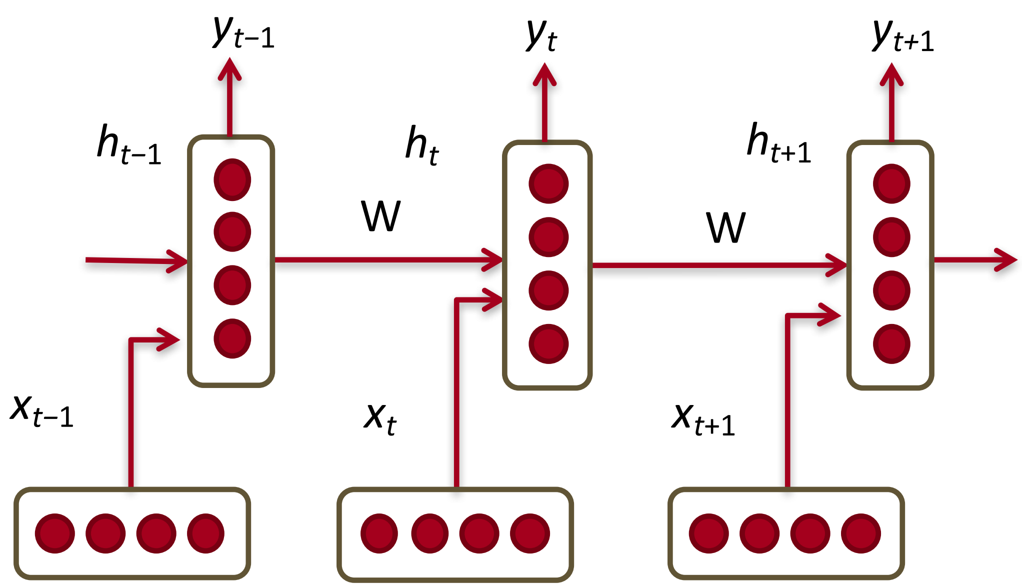
where $f'(z) = f(z)(1 - f(z))$ for logistic f



Recurrent NNs

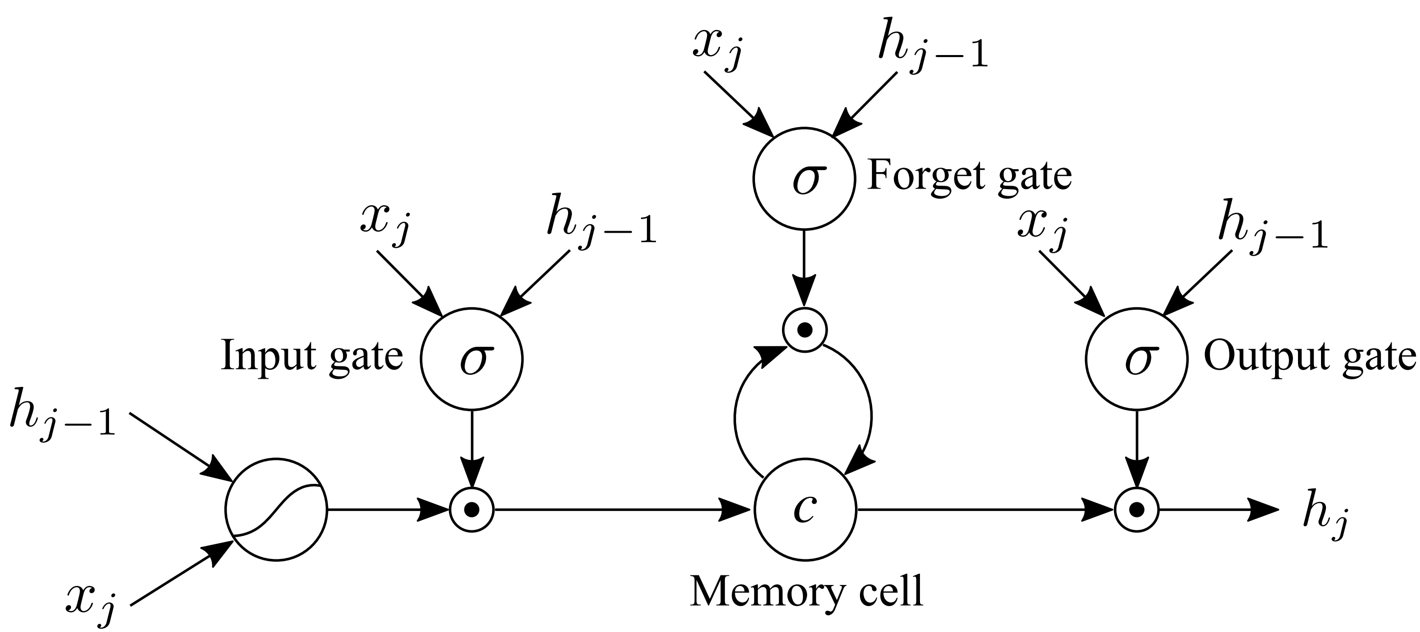
- Recurrent NNs (RNNs) are non-tree, sequential versions of recursive RvNNs
- Weights tied together for each time step

- Loss function on identity of predicted word at each time step



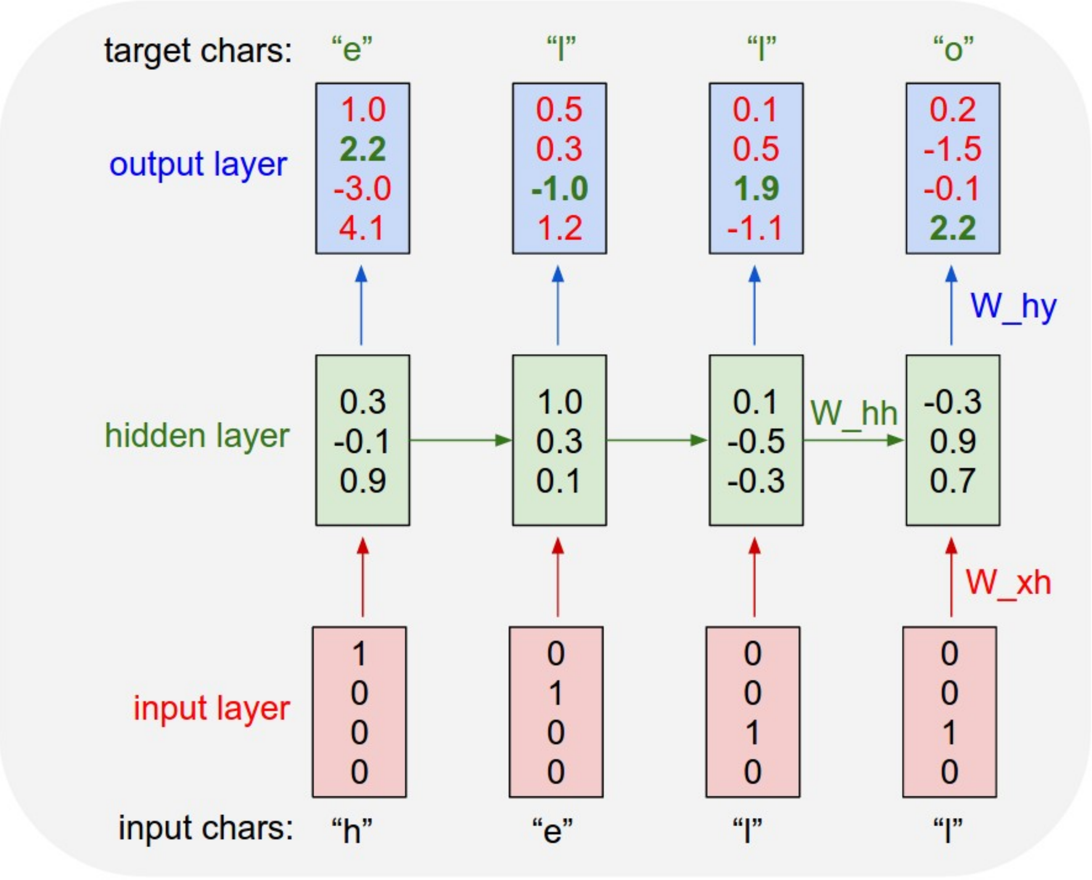
LSTM RNNs

- LSTM (Long short term memory) RNNs have gates for forgetting, allowing learning of longer-term connections by avoiding vanishing/ exploding gradients



Character RNNs

- Can directly process each character as a unit!
- Helps learn prefixes, stems, suffixes (form vs. function, rare/unseen words, etc.)



http://karpathy.github.io/2015/05/21/rnn-effectiveness/

RNN Generations

- Automatically generate Shakespeare from RNNs!

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

RNN Generations

- Automatically generate Wikipedia-style text from RNNs!

```
Naturalism and decision for the majority of Arab countries' capitalide was grounded
by the Irish language by [[John Clair]], [[An Imperial Japanese Revolt]], associated
with Guangzham's sovereignty. His generals were the powerful ruler of the Portugal
in the [[Protestant Immineners]], which could be said to be directly in Cantonese
Communication, which followed a ceremony and set inspired prison, training. The
emperor travelled back to [[Antioch, Perth, October 25|21]] to note, the Kingdom
of Costa Rica, unsuccessful fashioned the [[Thrales]], [[Cynth's Dajoard]], known
in western [[Scotland]], near Italy to the conquest of India with the conflict.
Copyright was the succession of independence in the slop of Syrian influence that
was a famous German movement based on a more popular servicious, non-doctrinal
and sexual power post. Many governments recognize the military housing of the
[[Civil Liberalization and Infantry Resolution 265 National Party in Hungary]],
that is sympathetic to be to the [[Punjab Resolution]]
(PJS)[http://www.humah.yahoo.com/guardian.
cfm/7754800786d17551963s89.htm Official economics Adjoint for the Nazism, Montgomery
was swear to advance to the resources for those Socialism's rule,
was starting to signing a major tripad of aid exile.]]
```

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

RNN Generations

- Automatically generate source code from RNNs!

```
{ { cite journal | id=Cerling Nonforest Department|format=Newlymeslated|none } }
''www.e-complete''.

'''See also''': [[List of ethical consent processing]]

== See also ==
*[[Iender dome of the ED]]
*[[Anti-autism]]

===[[Religion|Religion]]===
*[[French Writings]]
*[[Maria]]
*[[Revelation]]
*[[Mount Agamul]]

== External links==
* [http://www.biblegateway.nih.gov/entrepre/ Website of the World Festival. The labour (

==External links==
* [http://www.romanology.com/ Constitution of the Netherlands and Hispanic Competition :
```

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Various Applications of such RNNs

- Language Modeling and Language Generation
- Classification: Sentiment Analysis
- Conditioned Generation: End-to-end MT, Summarization
- Others: Parsing, Captioning, Q&A, Dialogue (some will be covered in future weeks)