NLP Tasks (Continued)

CMSC 473/673 - NATURAL LANGUAGE PROCESSING

Slides modified from Dr. Frank Ferraro & Dr. Jason Eisner

Learning Objectives

Distinguish between different text classification tasks

Formalize NLP Tasks at a high-level:

- What are the input/output for a particular task?
- What might the features be?
- What types of applications could the task be used for?

Similar to HW 1

Review

What's the difference between learning/training and inference/decoding/testing a model?

- Training: adjusting the model's weights to learn how to make good predictions; making the model
- Decoding: using a model's existing weights to make predictions; running the model when it's done

How are the objective function and evaluation function the same?

- Calculation to determine how accurate the predictions are

How are they different?

- Objective function is for training → tells the model how close it's getting to optimal weights
- Evaluation function is for testing

ML/NLP Framework for Learning & Prediction

instances

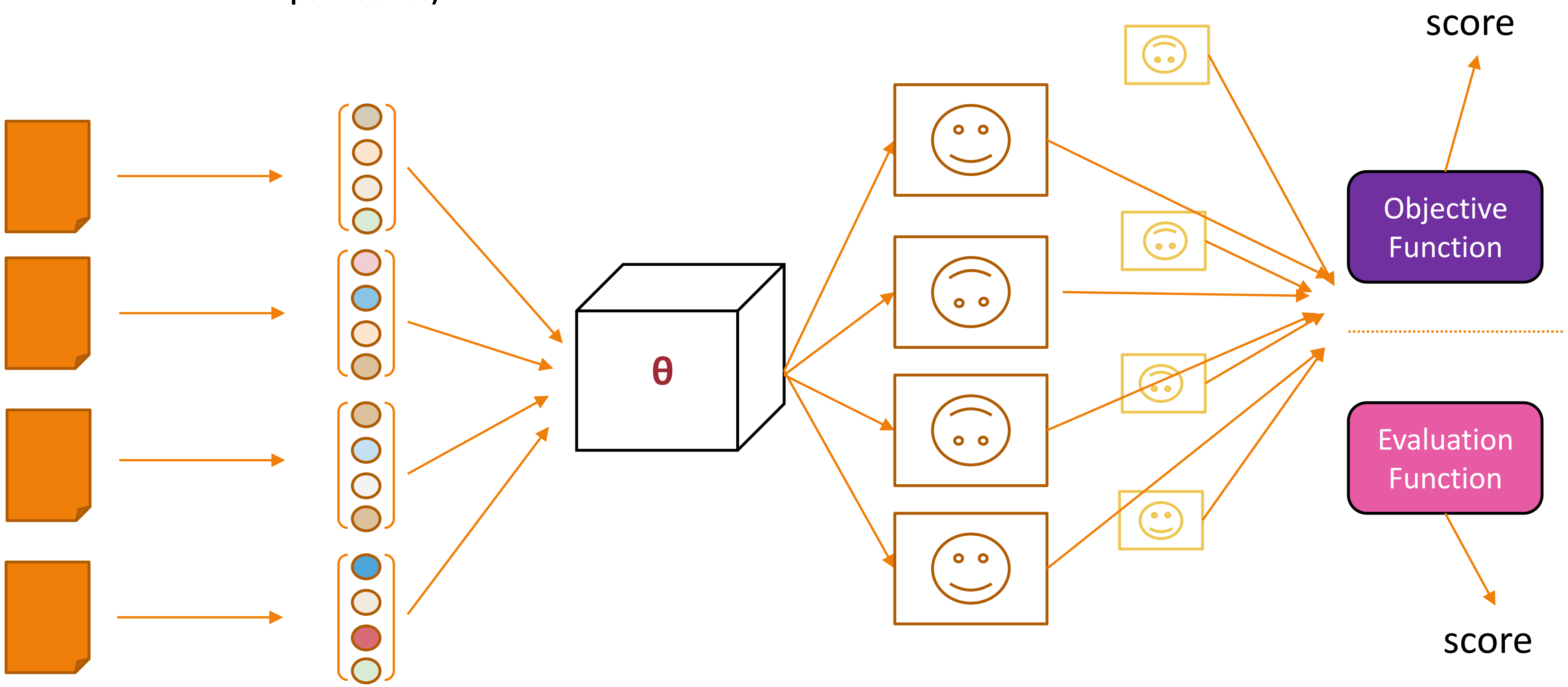
features: **K-dimensional vector representations** (one per instance)

ML model:
- take in featurized input
- output scores/labels
- contains weights θ

output

"Gold" (correct) labels

Objective / Eval Function



What are the three types of features we discussed?

1. Bag-of-words (or bag-of-characters, bag-of-relations)
   - easy to define / extract
   - sometimes still very useful
2. Linguistically-inspired features
   - harder to define
   - helpful for interpretation
   - depending on task: conceptually helpful
   - currently, not freq. used
3. Dense features via embeddings
   - harder to define
   - harder to extract (unless there’s a model to run)
   - currently: freq. used

Classification Types (Terminology)

Name	Number of Tasks (Domains) Labels are Associated with	# Label Types	Example
(Binary) Classification	1	2	Sentiment: Choose one of {positive or negative}
Multi-class Classification	1	> 2	Part-of-speech: Choose one of {Noun, Verb, Det, Prep, …}
Multi-label Classification	1	> 2	Sentiment: Choose multiple of {positive, angry, sad, excited, …}
Multi-task Classification	> 1	Per task: 2 or > 2 (can apply to binary or multi-class)	Task 1: part-of-speech Task 2: named entity tagging … ---------------------- Task 1: document labeling Task 2: sentiment

Text Annotation Tasks ("Classification" Tasks)

1. Classify the entire document ("text categorization")
2. Classify word tokens individually
3. Classify word tokens in a sequence
4. Identify phrases ("chunking")
5. Syntactic annotation (parsing)
6. Semantic annotation
7. Text generation

Slide courtesy Jason Eisner, with mild edits

Text Annotation Tasks ("Classification" Tasks)

1. Classify the entire document ("text categorization")
2. Classify word tokens individually
3. Classify word tokens in a sequence
4. Identify phrases ("chunking")
5. Syntactic annotation (parsing)
6. Semantic annotation

Slide courtesy Jason Eisner, with mild edits

Review: Text Classification

Assigning subject categories, topics, or genres

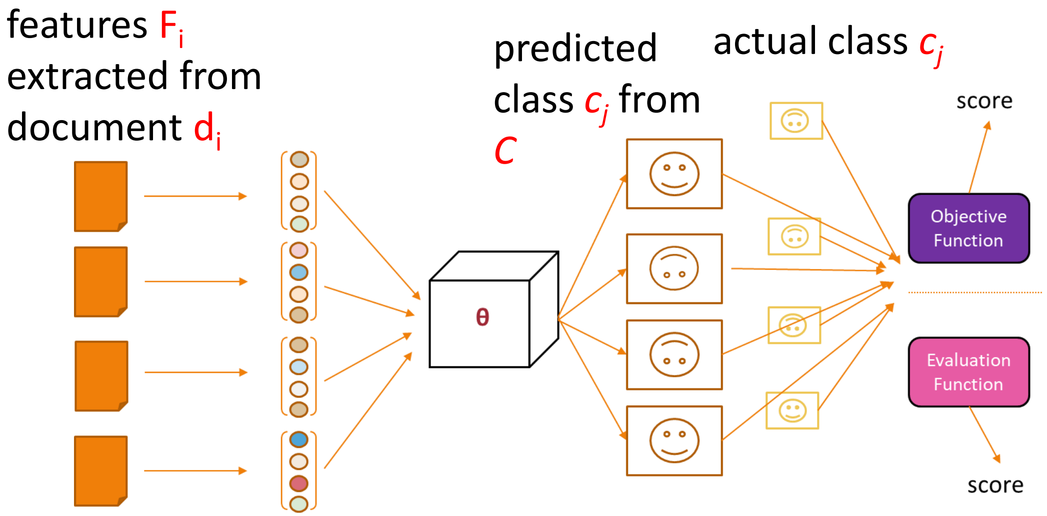
Spam detection

Authorship identification

Language Identification

Sentiment analysis

…



Text Annotation Tasks
("Classification" Tasks)

1. Classify the entire document ("text categorization")
2. Classify word tokens individually
3. Classify word tokens in a sequence
4. Identify phrases ("chunking")
5. Syntactic annotation (parsing)
6. Semantic annotation
7. Text generation

Slide courtesy Jason Eisner, with mild edits

p(class | token in context)

Word Sense Disambiguation (WSD)

Problem:

The company said the *plant* is still operating ...

⇒ (A) Manufacturing plant or

⇒ (B) Living plant

Training Data: Build a special classifier just for “plant” tokens

Sense	Context
(1) Manufacturing	... union responses to *plant* closures . ...
” ”	... computer disk drive *plant* located in ...
” ”	company manufacturing *plant* is in Orlando ...
(2) Living	... animal rather than *plant* tissues can be ...
” ”	... to strain microscopic *plant* life from the ...
” ”	and Golgi apparatus of *plant* and animal cells

Test Data:

Sense	Context
???	... vinyl chloride monomer *plant* , which is ...
???	... molecules found in *plant* tissue from the ...

slide courtesy of D. Yarowsky (modified)

p(class | token in context)

WSD for Machine Translation
(English → Spanish)

Problem:

... He wrote the last **sentence** two years later ...

⇒ *sentencia* (legal sentence)   or

⇒ *frase* (grammatical sentence)

Training Data: Build a special classifier just for “sentence” tokens

Translation	Context
(1) sentencia	... for a maximum *sentence* for a young offender ...
” ”	... of the minimum *sentence* of seven years in jail ...
” ”	... were under the *sentence* of death at that time ...
(2) frase	... read the second *sentence* because it is just as ...
” ”	... The next *sentence* is a very important ...
” ”	... It is the second *sentence* which I think is at ...

Test Data:

Translation	Context
???	... cannot criticize a *sentence* handed down by ...
???	... listen to this *sentence* uttered by a former ...

slide courtesy of D. Yarowsky (modified)

p(class | token in context)

Accent Restoration in Spanish & French

Problem:

Input:	... deja travaille cote a cote ...
	⇓
Output:	... déjà travaillé côte à côte ...

Examples:

... appeler l'autre **cote** de l'atlantique ...
⇒ *côté* (meaning side) or
⇒ *côte* (meaning coast)

. . . une famille des **pecheurs** . . .
⇒ *pêcheurs* (meaning fishermen) or
⇒ *pécheurs* (meaning sinners)

slide courtesy of D. Yarowsky (modified)

p(class | token in context)

Accent Restoration in Spanish & French

Training Data:

Pattern	Context
(1) côté	... du laisser de *cote* faute de temps ...
" "	... appeler l' autre *cote* de l' atlantique ...
" "	... passe de notre *cote* de la frontiere ...
(2) côte	... vivre sur notre *cote* ouest toujours ...
" "	... creer sur la *cote* du labrador des ...
" "	travaillaient cote a *cote* , ils avaient ...

Test Data:

Pattern	Context
???	... passe de notre *cote* de la frontiere ...
???	... creer sur la *cote* du labrador des ...

slide courtesy of D. Yarowsky (modified)

p(class | token in context)

Spelling Correction

Problem:

... and he fired presidential **aid/aide** Dick Morris after ...

$\Rightarrow$ *aid* or

$\Rightarrow$ *aide*

Training Data:

Spelling	Context
(1) aid	... and cut the foreign *aid/aide* budget in fiscal 1996 ...
" "	... they offered federal *aid/aide* for flood-ravaged states ...
(2) aide	... fired presidential *aid/aide* Dick Morris after ...
" "	... and said the chief *aid/aide* to Sen. Baker, Mr. John ...

Test Data:

Spelling	Context
???	... said the longtime *aid/aide* to the Mayor of St. ...
???	... will squander the *aid/aide* it receives from the ...

slide courtesy of D. Yarowsky (modified)

What features? Example: "word to [the] left [of correction]"

Word to left	Frequency as **Aid**	Frequency as **Aide**
foreign	718	1
federal	297	0
western	146	0
provide	88	0
covert	26	0
oppose	13	0
future	9	0
similar	6	0
presidential	0	63
chief	0	40
longtime	0	26
aids-infected	0	2
sleepy	0	1
disaffected	0	1
indispensable	2	1
practical	2	0
squander	1	0

Spelling correction using an n-gram language model ($n \geq 2$) would use words to left and right to help predict the true word.

Similarly, an HMM would predict a word's class using classes to left and right.

But we'd like to throw in all kinds of other features, too …

slide courtesy of D. Yarowsky (modified)

p(class | token in context)

Text-to-Speech Synthesis

Problem:

... slightly elevated *lead* levels ...

$\Rightarrow$ *lɛd* (as in *lead mine*) or

$\Rightarrow$ *li:d* (as in *lead role*)

Training Data:

Pronunciation	Context		
(1) lɛd	... it monitors the	*lead*	levels in drinking ...
" "	... conference on	*lead*	poisoning in ...
" "	... strontium and	*lead*	isotope zonation ...
(2) li:d	... maintained their	*lead*	Thursday over ...
" "	... to Boston and	*lead*	singer for Purple ...
" "	... Bush a 17-point	*lead*	in Texas , only 3 ...

Test Data:

Pronunciation	Context		
???	... median blood	*lead*	concentration was ..
???	... his double-digit	*lead*	nationwide . The ...

slide courtesy of D. Yarowsky (modified)

An assortment of possible cues ...

	Position	Collocation	lɛd	li:d
N-grams	+1 L	lead *level/N*	219	0
	-1 W	*narrow* lead	0	70
(word,	+1 W	lead *in*	207	898
lemma,	-1W,+1W	*of* lead *in*	162	0
part-of-speech)	-1W,+1W	*the* lead *in*	0	301
	+1P,+2P	lead *, <NOUN>*	234	7
Wide-context	±k W	*zinc* (in ±*k* words)	235	0
collocations	±k W	*copper* (in ±*k* words)	130	0
Verb-object	-V L	*follow/V* + lead	0	527
relationships	-V L	*take/V* + lead	1	665

generates a whole bunch of potential cues – use data to find out which ones work best

Word to left	Frequency as **Aid**	Frequency as **Aide**
foreign	718	1
federal	297	0
western	146	0
provide	88	0

slide courtesy of D. Yarowsky (modified)

An assortment of possible cues ...

	Position	Collocation	lɛd	li:d
N-grams	+1 L	lead *level/N*	219	0
	-1 W	*narrow* lead	0	70
(word,	+1 W	lead *in*	$\boxed{207}$	$\boxed{898}$
lemma,	-1W,+1W	*of* lead *in*	162	0
part-of-speech)	-1W,+1W	*the* lead *in*	0	301
	+1P,+2P	lead , <*NOUN*>	234	7
Wide-context	±k W	*zinc* (in ±*k* words)	235	0
collocations	±k W	*copper* (in ±*k* words)	130	0
Verb-object	-V L	*follow/V* + lead	0	527
relationships	-V L	*take/V* + lead	1	665

This feature is relatively weak, but weak features are still useful, especially since very few features will fire in a given context.

merged ranking of all cues of all these types

11.40	*follow/V* + lead	⇒ li:d
11.20	*zinc* (in ±*k* words)	⇒ lɛd
11.10	lead *level/N*	⇒ lɛd
10.66	*of* lead *in*	⇒ lɛd
10.59	*the* lead *in*	⇒ li:d
10.51	lead *role*	⇒ li:d

slide courtesy of D. Yarowsky (modified)

Final decision list for *lead* (abbreviated)

What are the input/output?
What are the features?
What types of applications?

List of all features,
ranked by their weight.

(These weights are for a simple "decision list" model where the single highest-weighted feature that fires gets to make the decision all by itself.

However, a log-linear model, which adds up the weights of all features that fire, would be roughly similar.)

LogL	Evidence	Pronunciation
11.40	*follow/V* + lead	⇒ li:d
11.20	*zinc* (in $\pm k$ words)	⇒ lɛd
11.10	lead *level/N*	⇒ lɛd
10.66	*of* lead *in*	⇒ lɛd
10.59	*the* lead *in*	⇒ li:d
10.51	lead *role*	⇒ li:d
10.35	*copper* (in $\pm k$ words)	⇒ lɛd
10.28	lead *time*	⇒ li:d
10.24	lead *levels*	⇒ lɛd
10.16	lead *poisoning*	⇒ lɛd
8.55	*big* lead	⇒ li:d
8.49	*narrow* lead	⇒ li:d
7.76	*take/V* + lead	⇒ li:d
5.99	lead , *NOUN*	⇒ lɛd
$\boxed{1.15}$	lead *in*	⇒ li:d
	○○○	

slide courtesy of D. Yarowsky (modified)

Token Classification

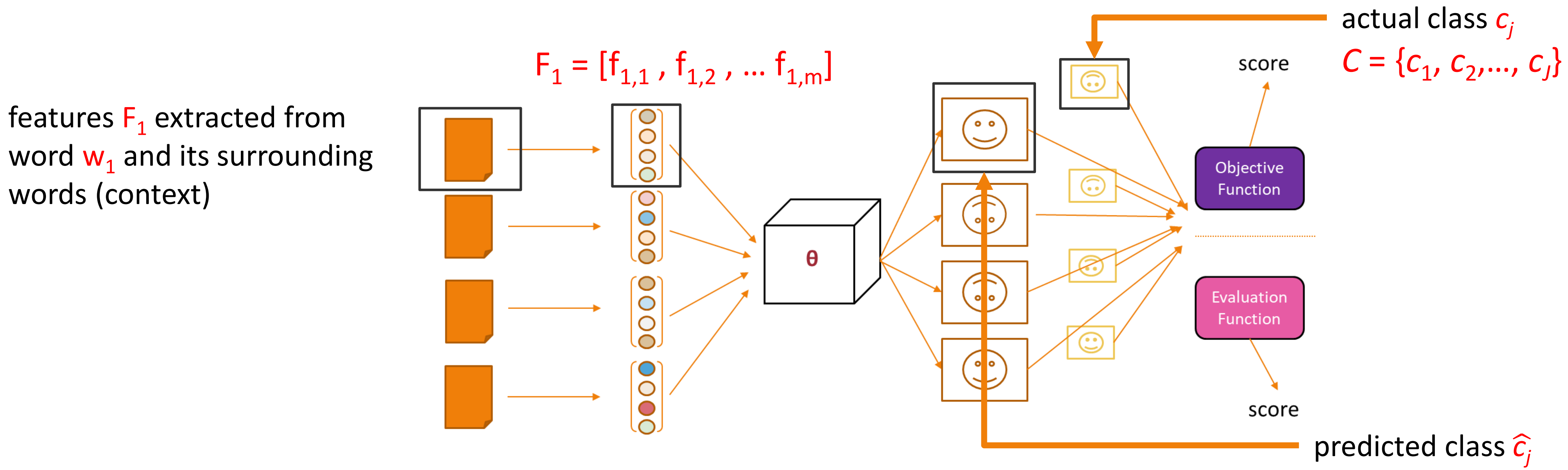
Word pronunciation

Word sense disambiguation (WSD) within or across languages

Accent restoration

…

Other examples?



Text Annotation Tasks (“Classification” Tasks)

1. Classify the entire document (“text categorization”)
2. Classify word tokens individually
3. Classify word tokens in a sequence (i.e., order matters)
4. Identify phrases (“chunking”)
5. Syntactic annotation (parsing)
6. Semantic annotation
7. Text generation

Slide courtesy Jason Eisner, with mild edits

Part of Speech Tagging

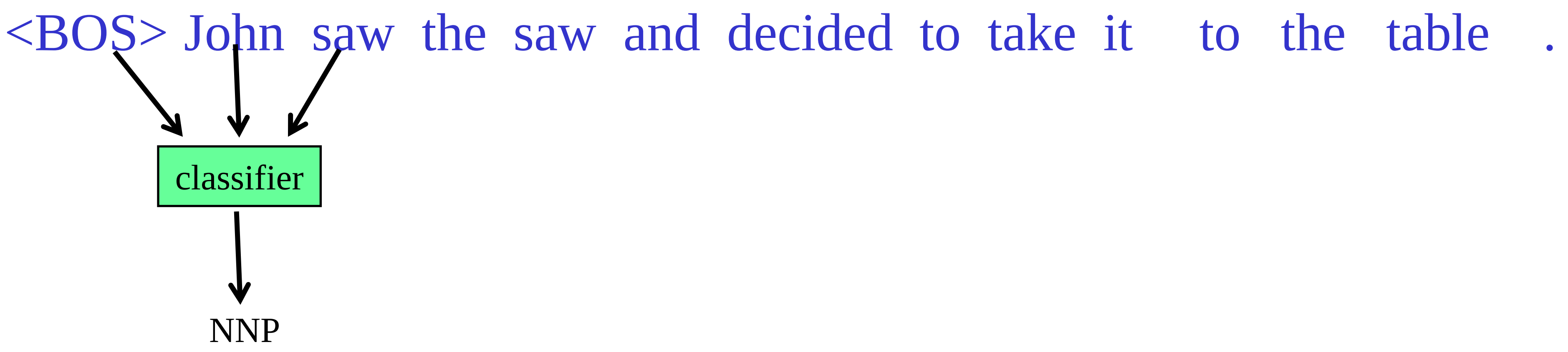
We could treat tagging as a token classification problem

- Tag each word independently given features of context
- And features of the word's spelling (suffixes, capitalization)

Slide courtesy Jason Eisner, with mild edits

Sequence Labeling as Classification

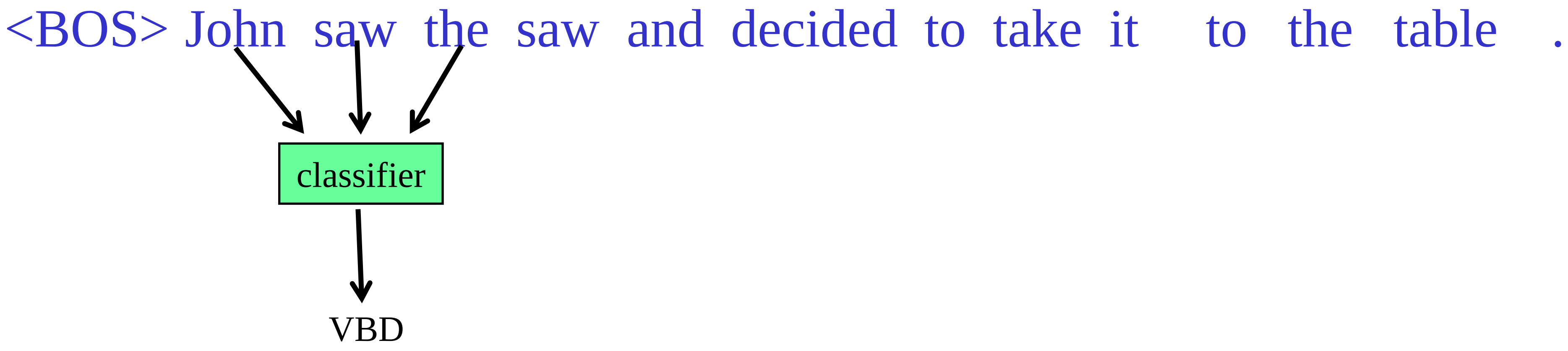
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

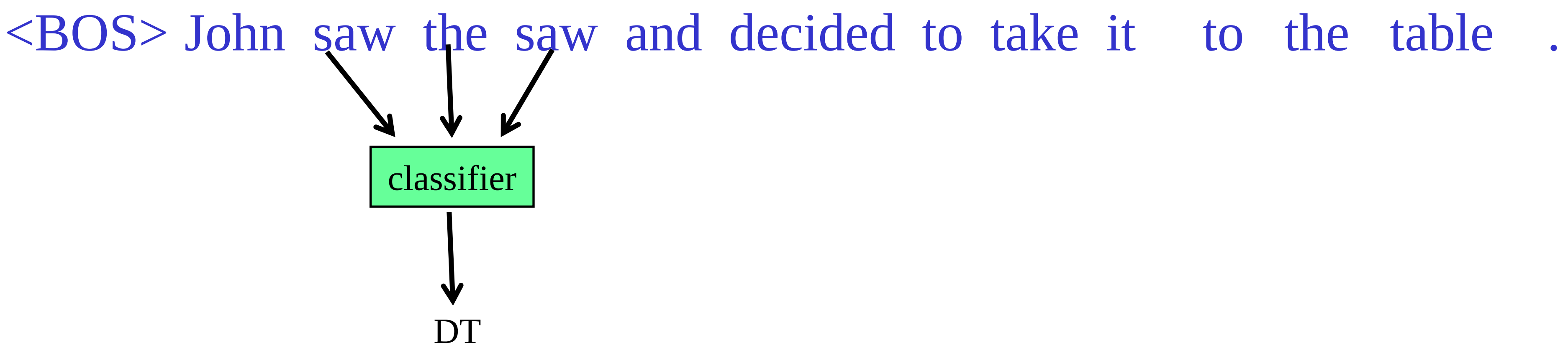
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

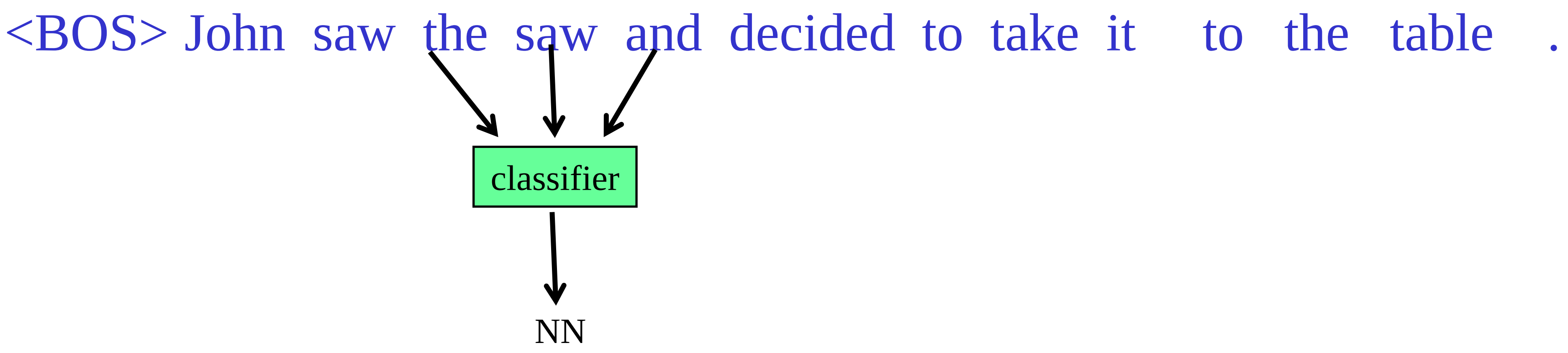
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

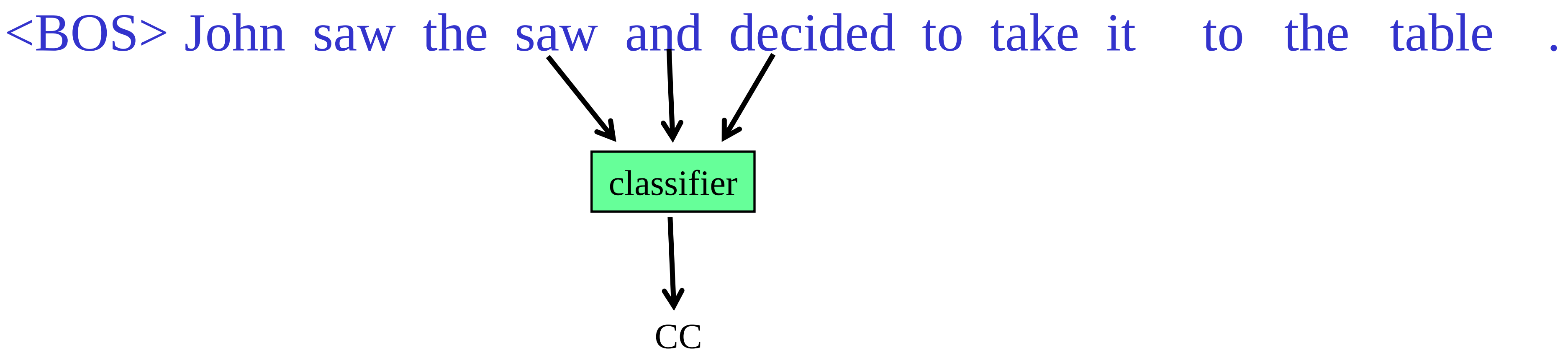
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

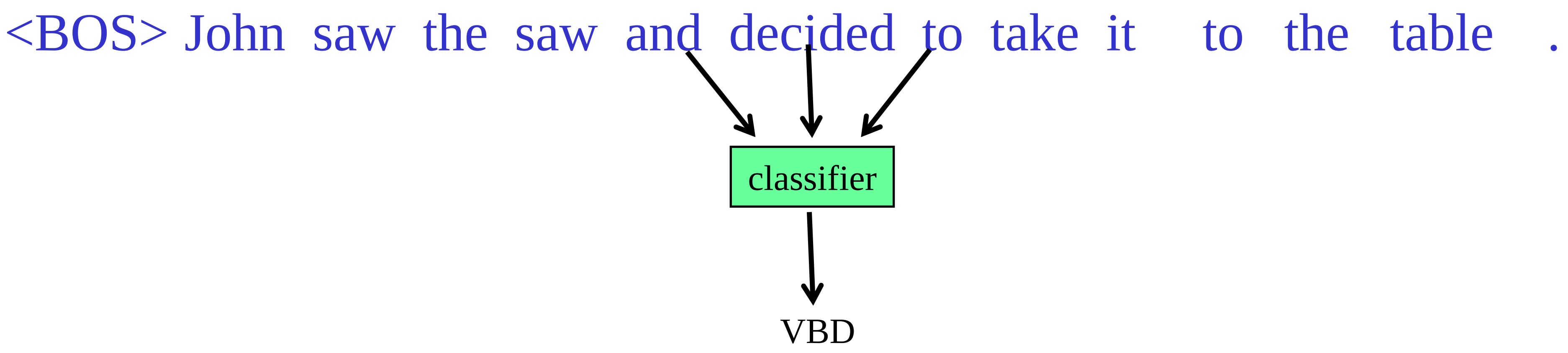
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

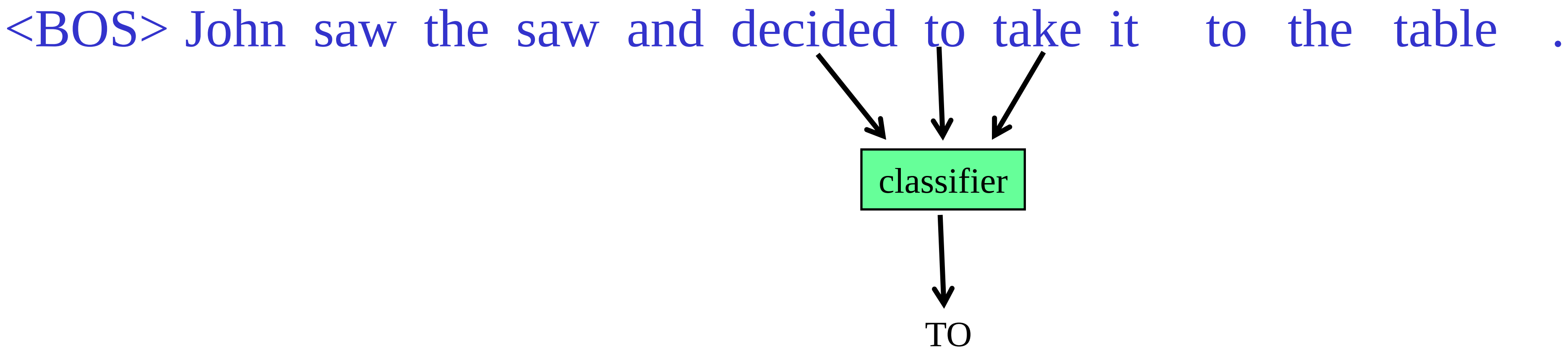
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

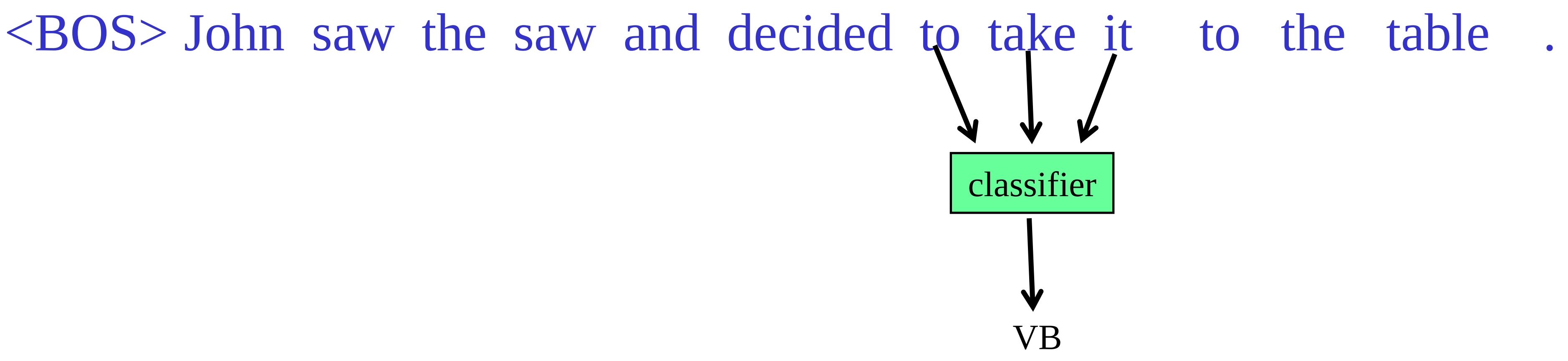
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

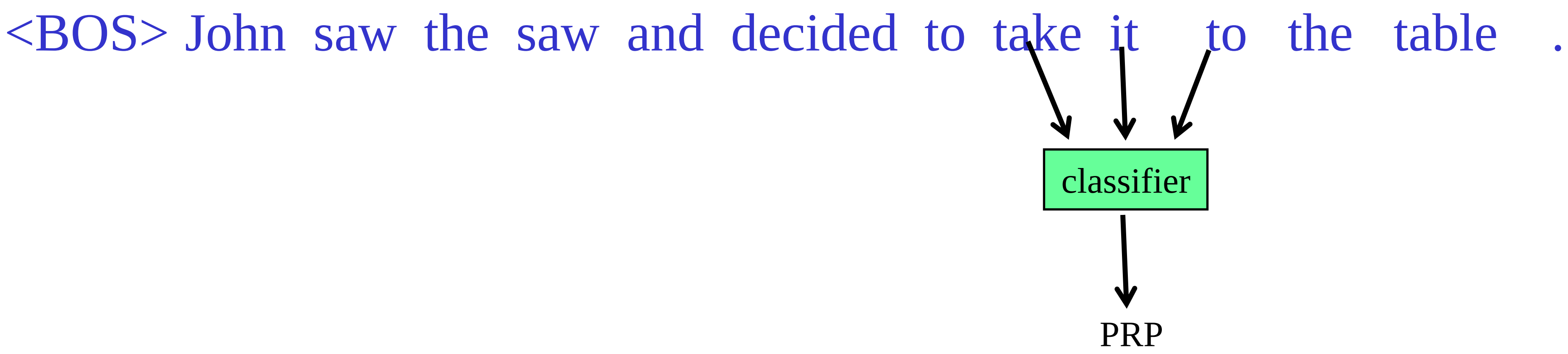
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

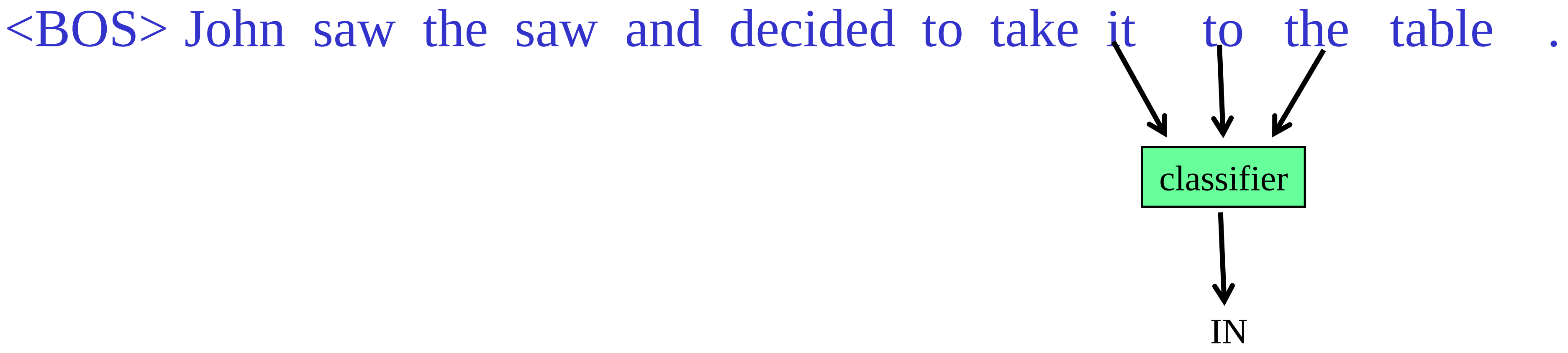
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

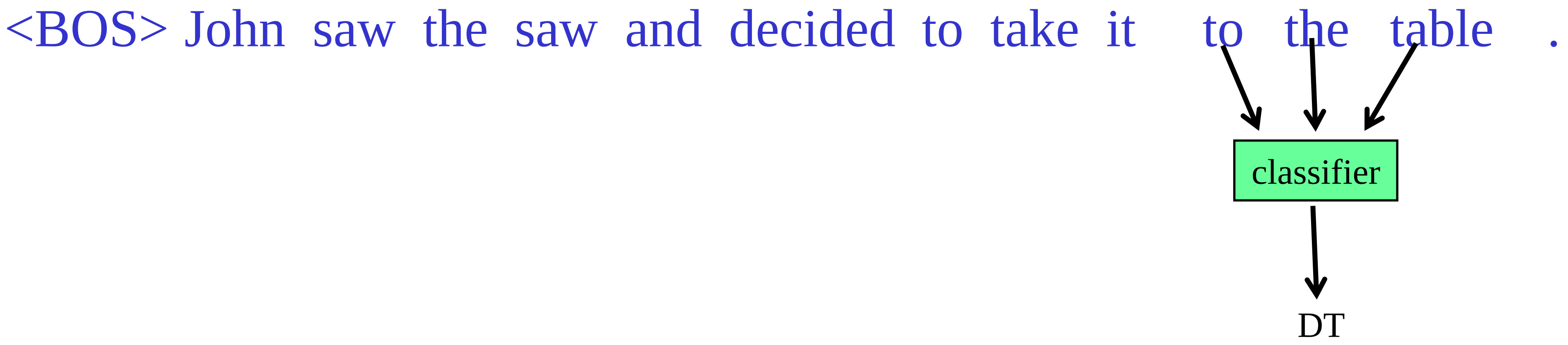
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

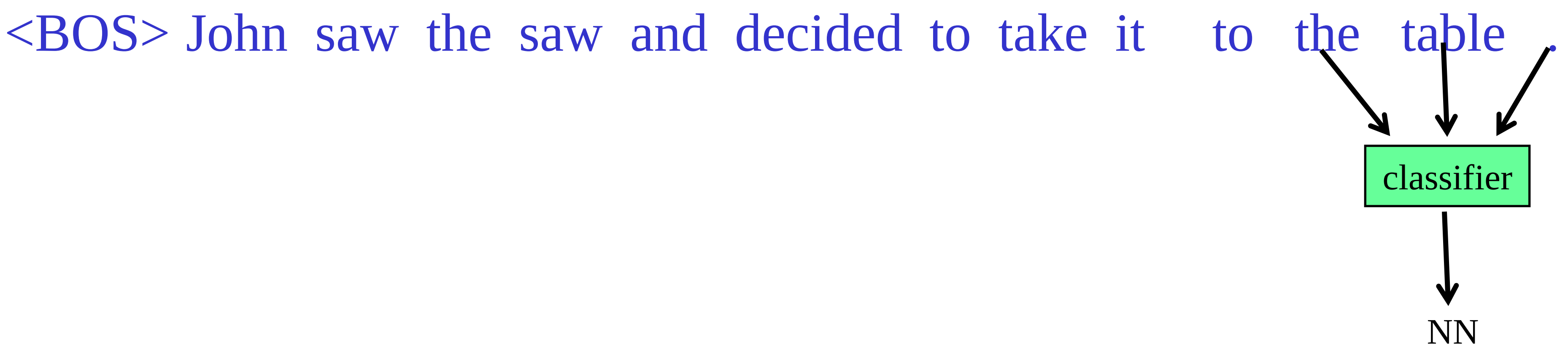
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



What are the input/output?

Slide courtesy Ray Mooney, with mild edits

Part of Speech Tagging

We could treat tagging as a token classification problem

- Tag each word independently given features of context
- And features of the word’s spelling (suffixes, capitalization)

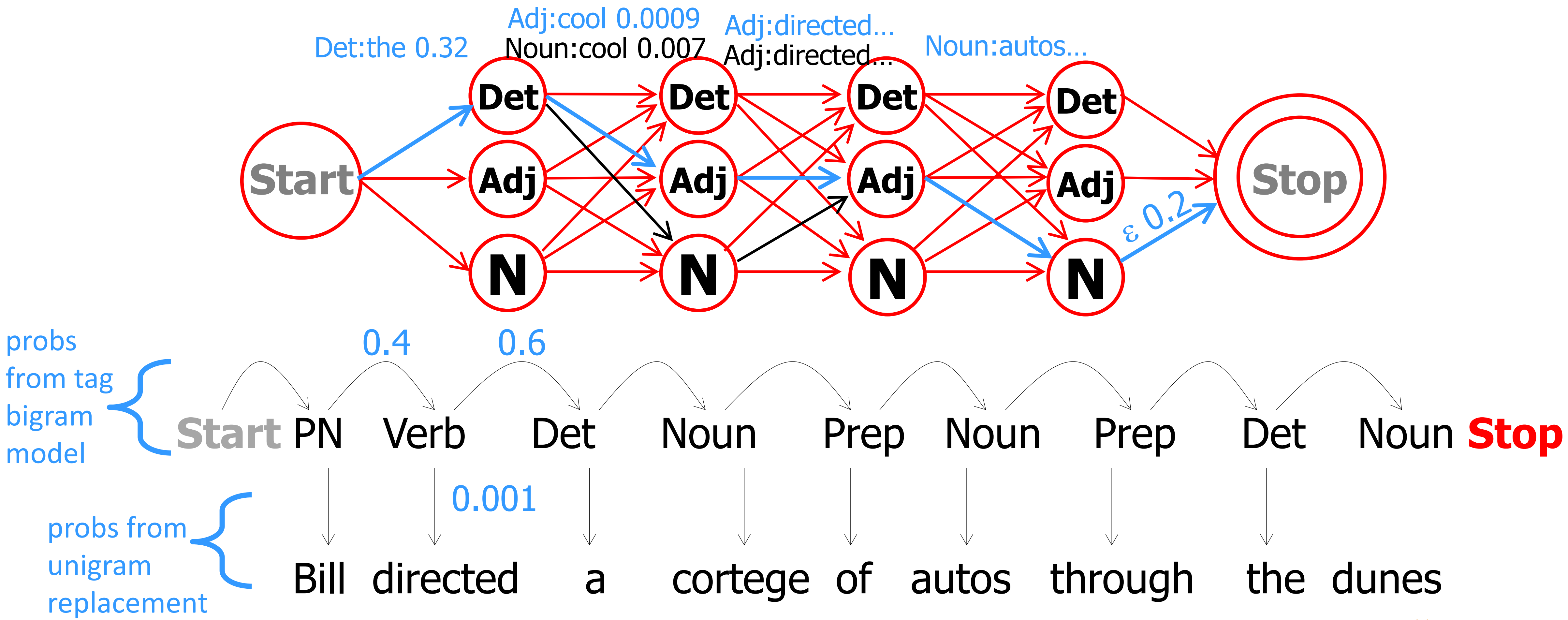
Or we could use an HMM:

- The point of the HMM is basically that the tag of one word might depend on the tags of adjacent words.

Slide courtesy Jason Eisner, with mild edits

Part of Speech Tagging

Or we could use an HMM:



Slide courtesy Jason Eisner, with mild edits

Part of Speech Tagging

We could treat tagging as a token classification problem

- Tag each word independently given features of context
- And features of the word's spelling (suffixes, capitalization)

Or we could use an HMM:

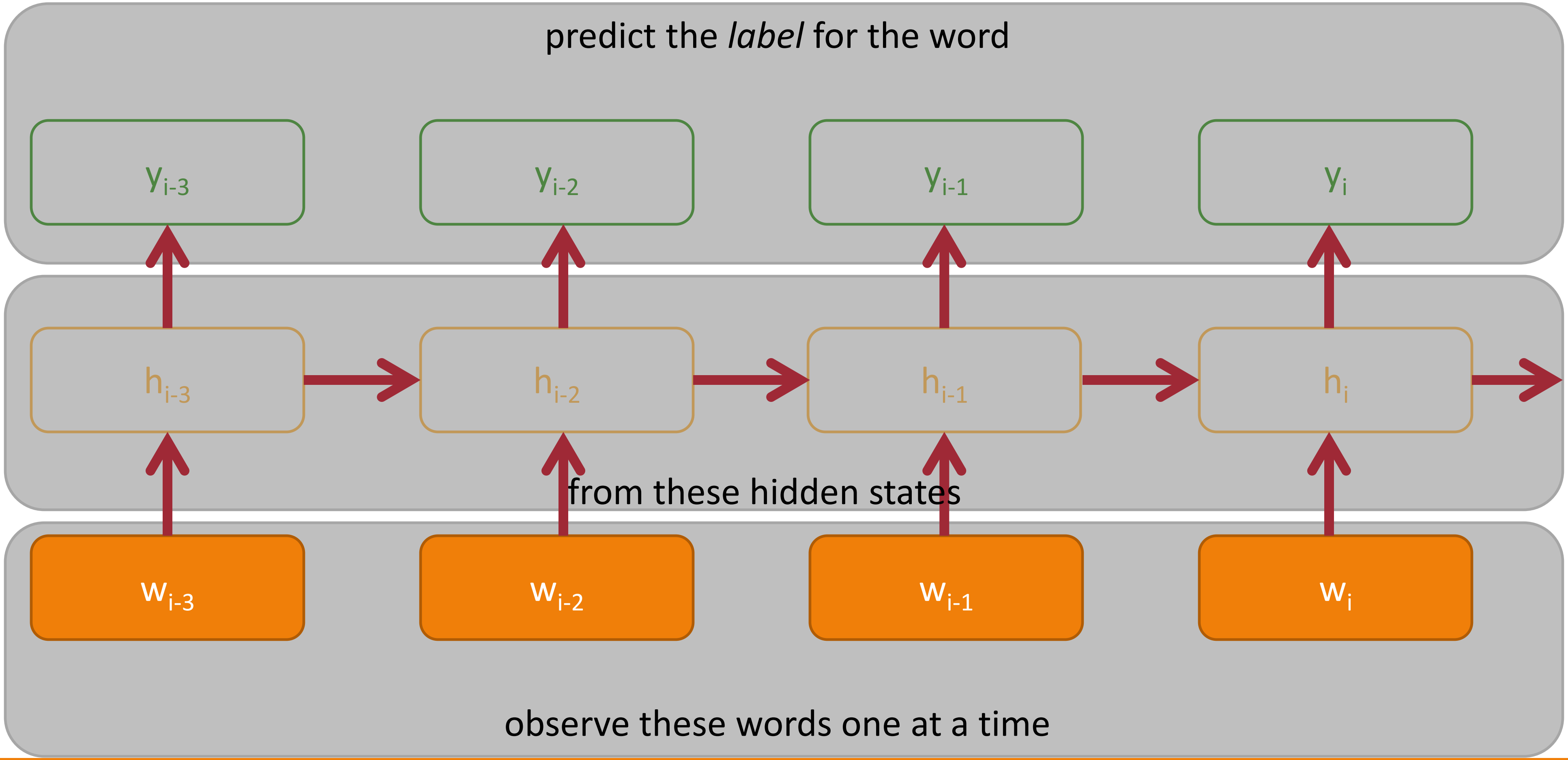
- The point of the HMM is basically that the tag of one word might depend on the tags of adjacent words.

Combine these two ideas??

- We'd like rich features (e.g., in a log-linear model), but we'd also like our feature functions to depend on adjacent tags.
- So, the problem is to predict **all** tags together.

Slide courtesy Jason Eisner, with mild edits

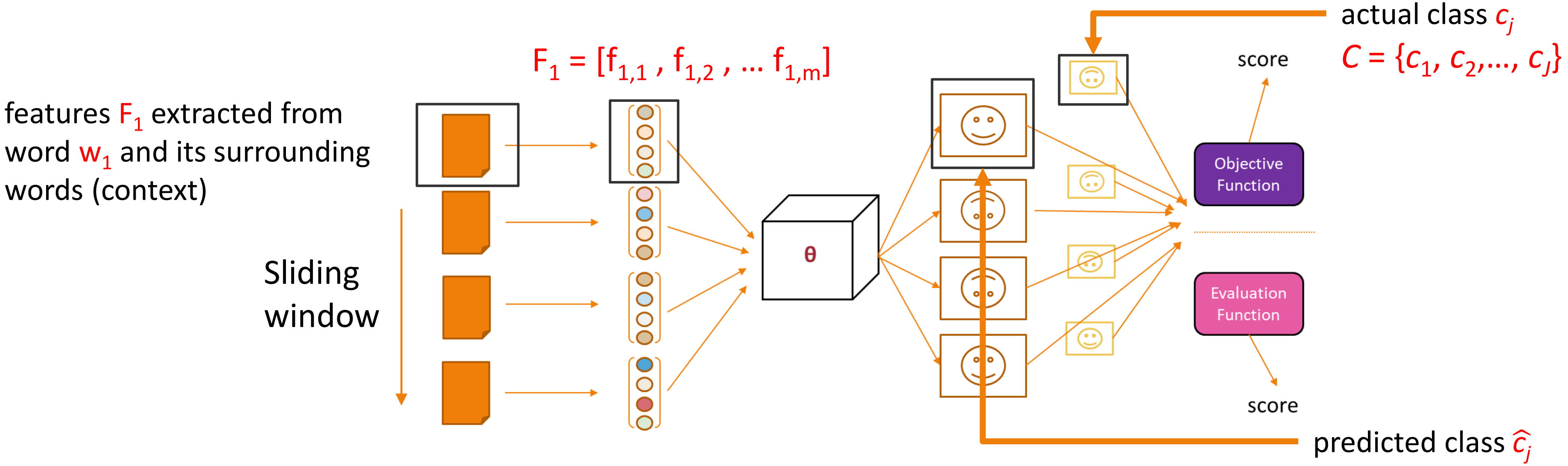
Can We Use Neural, Recurrent Methods for PoS Tagging?



Token Classification in a Sequence

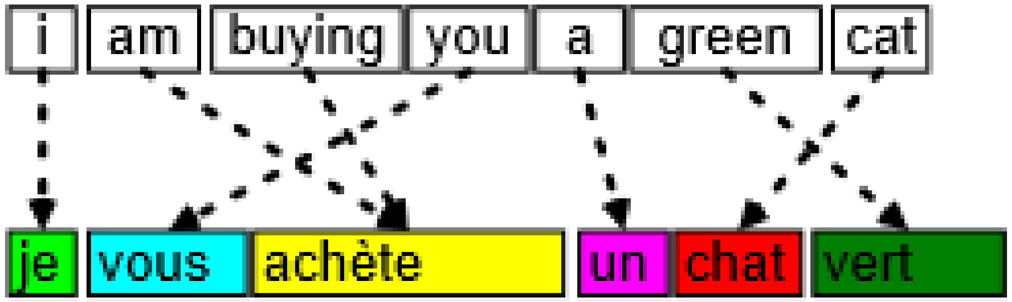
Part of speech tagging

Word alignment



Machine Translation: Word Alignment

What are the input/output?



What kinds of features might we want to consider here?

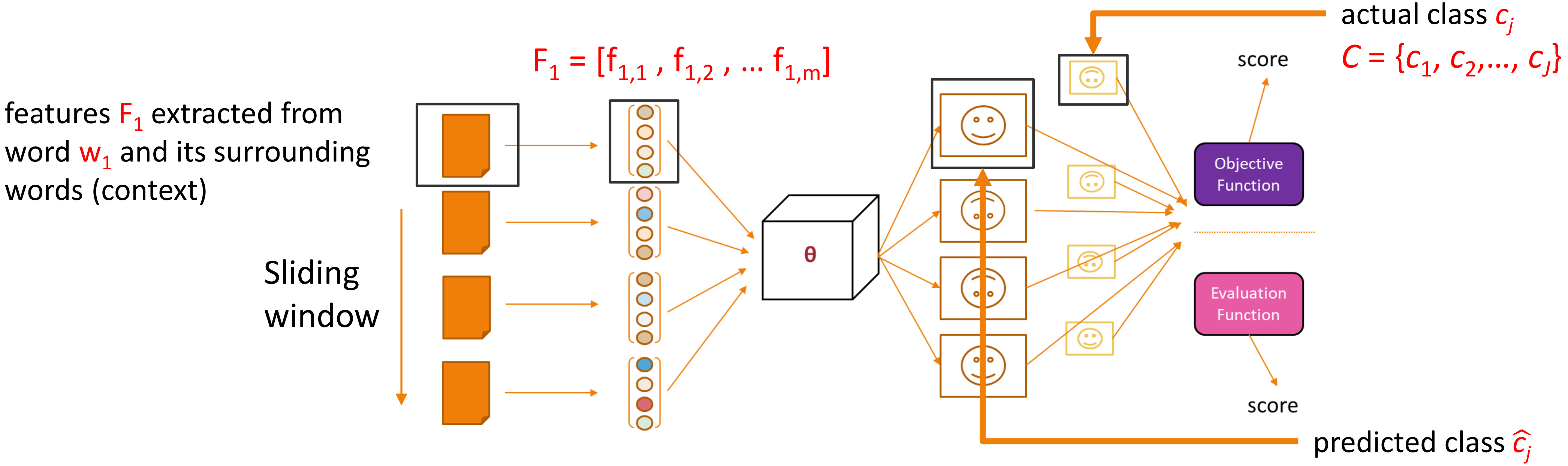
https://towardsdatascience.com/machine-translation-a-short-overview-91343ff39c9f

Token Classification in a Sequence

Part of speech tagging

Word alignment

Other examples?



Text Annotation Tasks (“Classification” Tasks)

1. Classify the entire document (“text categorization”)
2. Classify word tokens individually
3. Classify word tokens in a sequence
4. Identify phrases (“chunking”)
5. Syntactic annotation (parsing)
6. Semantic annotation
7. Text generation

Slide courtesy Jason Eisner, with mild edits

Example: Finding Named Entities

Named entity recognition (NER)

Identify proper names in texts, and classification into a set of predefined categories of interest

NE Types

TYPE	DESCRIPTION
PERSON	People, including fictional
NORP	Nationalities or religious or political groups
FACILITY	Buildings, airports, highways, bridges, etc
ORG	Companies, agencies, institutions, etc
GPE	Countries, cities, states
LOC	Non-GPE locations, mountain ranges, bodies of water
PRODUCT	Objects, vehicles, foods, etc (Not services)
EVENT	Named hurricanes, battles, wars, sports events, etc
WORK_OF_ART	Titles of books, songs, etc
LAW	Named documents made into laws
LANGUAGE	Any named language
DATE	Absolute or relative dates or periods.
TIME	Times smaller than a day
PERCENT	Percentage, including "%".
MONEY	Monetary values, including unit
QUANTITY	Measurements, as of weight or distance
ORDINAL	"first", "second", etc
CARDINAL	Numerals that do not fall under another type

https://medium.com/@rajat.jain1/natural-language-extraction-using-spacy-on-a-set-of-novels-88b159d68686

Named Entity Recognition

CHICAGO (AP) — Citing high fuel prices, United Airlines said Friday it has increased fares by $6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit AMR, immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL, said the increase took effect Thursday night and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York.

Slide courtesy Jim Martin