NLP Tasks 2

Instructor: Lara J. Martin (she/they)

TA: Omkar Kulkarni (he)

https://laramartin.net/NLP-class/

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

Learning Objectives

Distinguish between classification and regression, supervised and unsupervised learning

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?

Enumerate different input scopes of tasks when thought of as classification

High-Level View of the Course

Part 1: Intro to NLP terms & ML concepts

Tasks	What you are trying to solve
Models & Evaluation	What you are making and how you know it's doing well
Vector Embeddings	A way of encoding features

High-Level View of the Course

Part 2: Evolution of the language model

- N-Gram Language Models
- Logistic Regression
- Neural Language Models
- Recurrent NNs
- Seq2Seq Models
- Transformers
- Masked LMs
- Large Language Models (Few shot)

High-Level View of the Course

Part 3: Language Model Extensions

- Alignment
- Small “Large” Language Models
- RAG

High-Level View of the Course

Part 4: In-depth dives into certain applications

- Automatic Speech Recognition
- Machine Translation
- Dialog Systems

Review: Helpful ML Terminology

Model: the (computable) way to go from **features** (input) to labels/scores (output)

Weights/parameters (θ): vectors of numbers that control how the model produces labels/scores from inputs. These are learned through **training**.

Objective function: an algorithm/calculation, whose variables are the **weights** of the **model**, that we numerically optimize in order to learn appropriate weights based on the labels/scores. The **model's** weights are adjusted.

Evaluation function: an algorithm/calculation that scores how "correct" the **model's** predictions are. The **model's** weights are not adjusted.

Review: Helpful ML Terminology

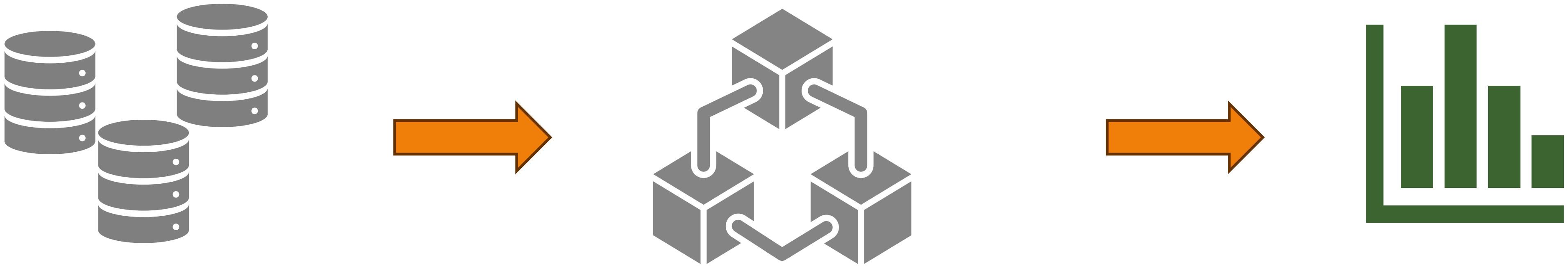
Training / Learning: the process of adjusting the model's weights to learn to make good predictions.

Inference / Prediction / Decoding / Classification:

the process of *using* a model's existing weights to make (hopefully!) good predictions

Features: values that denote an aspect of the data; often a K-dimensional vector

How do we learn models?

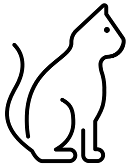


Take past experiences
(lots of data; corpus)

Find patterns
(the ML algorithm)

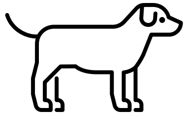
Use on new experiences
(save & test the model)

How do we learn models?



cat

hamster



dog

[position of lines]
[loops and dots]
[any other information relevant to our problem]

P(hamster|[line in this position],…)

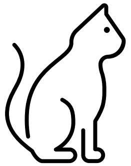
P(dog|[line in this other position],…)

Have data with features extracted (and possibly labels)

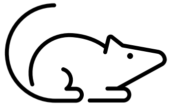
Learn associations between features and labels

Dividing up data for Training

Why would we do this?

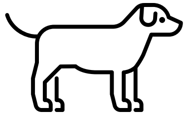


cat



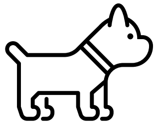
mouse

hamster

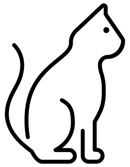


dog

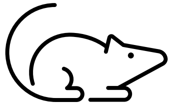
[position of lines]
[loops and dots]
[any other information relevant to our problem]



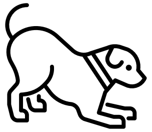
dog



cat



mouse



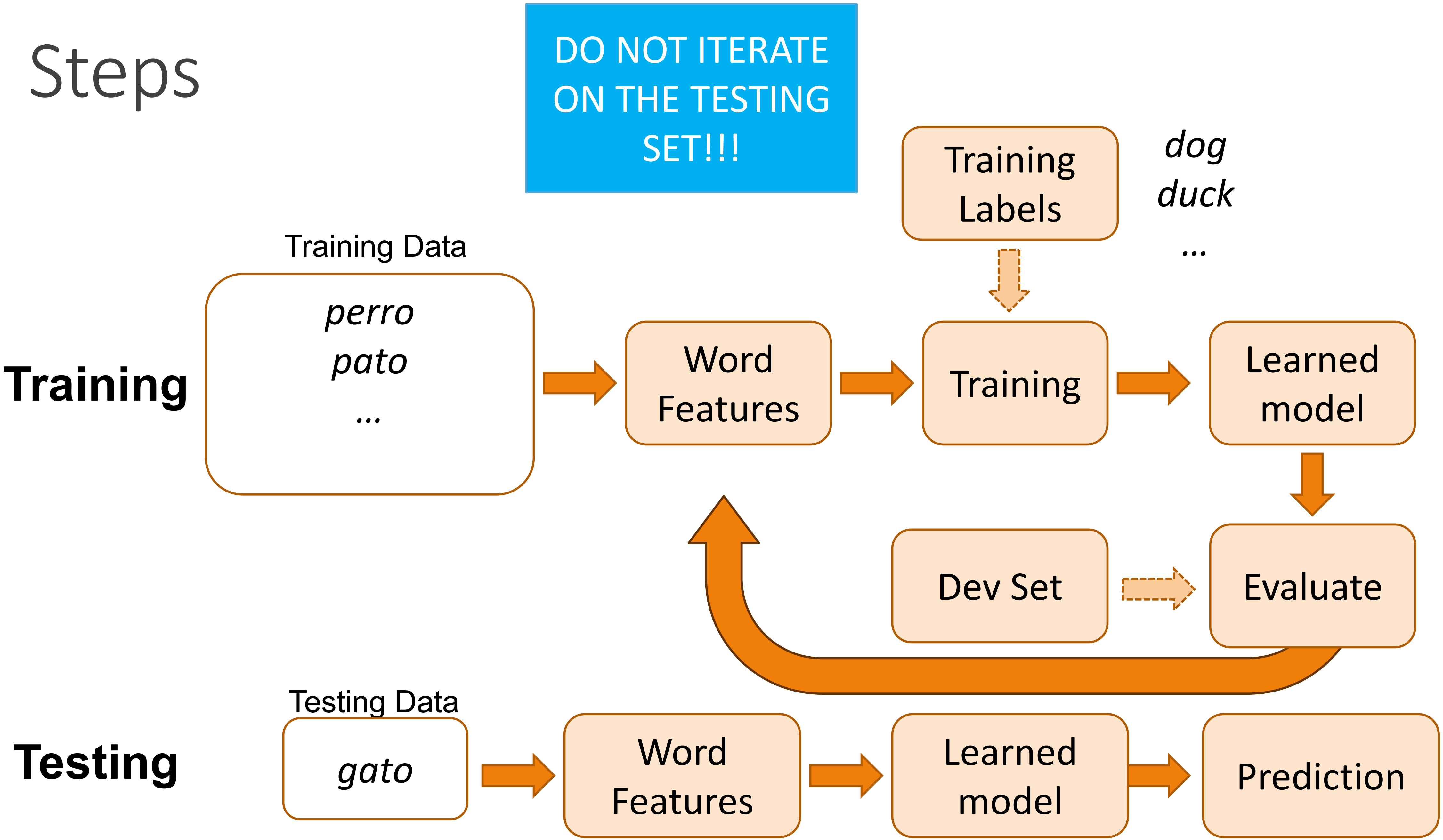
dog

Training
~80%

Validation/Development
~10%

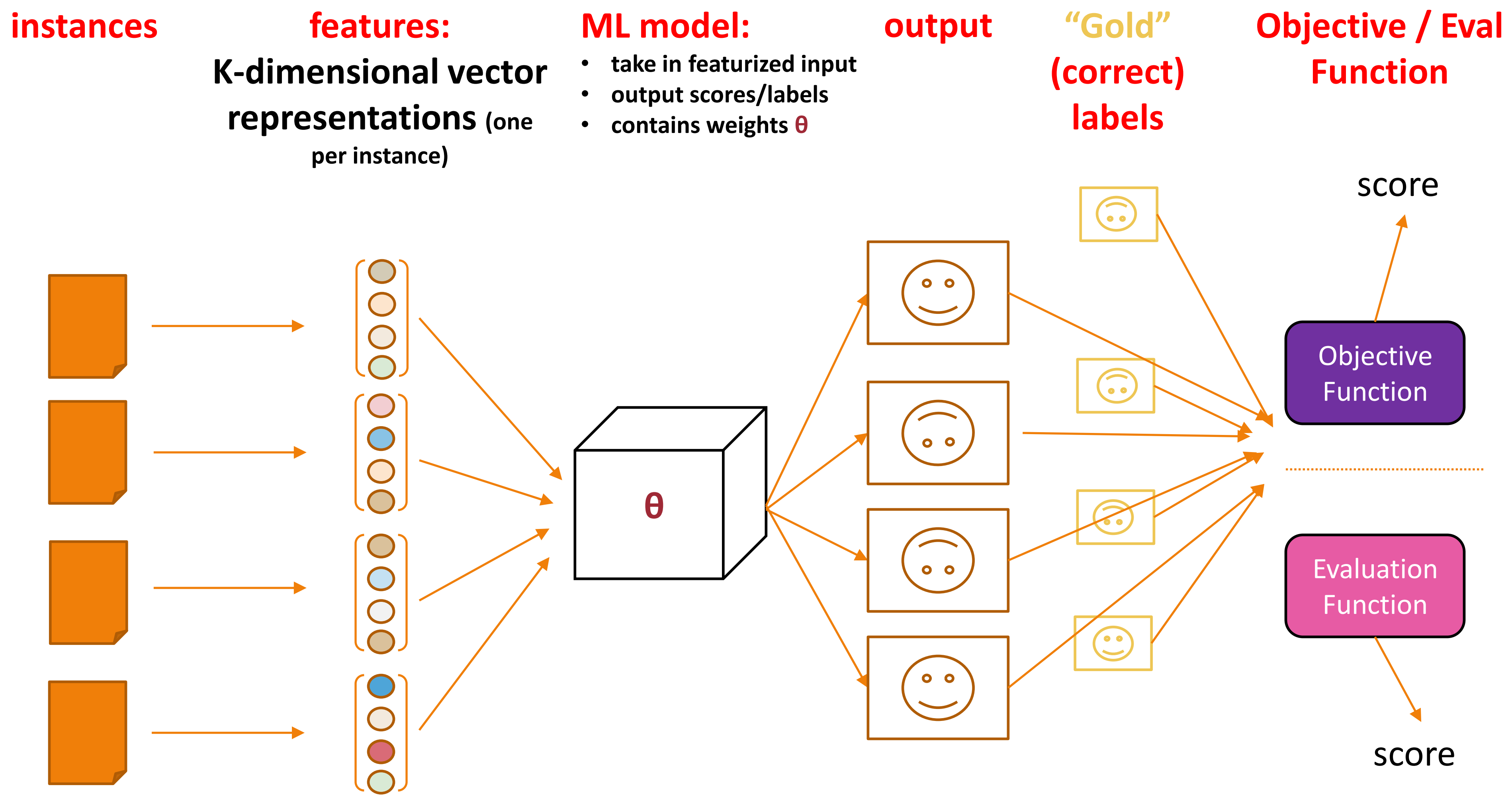
Testing
~10%

Steps



DO NOT ITERATE ON THE TESTING SET!!!
Training Labels
dog
duck
…
Training Data
perro
pato
…
Training
Word Features
Training
Learned model
Dev Set
Evaluate
Testing Data
Testing
gato
Word Features
Learned model
Prediction

Review: ML/NLP Framework for Learning & Prediction



The Machine Learning Framework

$$y = f(\mathbf{x})$$

output

prediction function

feature(s) of input

Training: given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the prediction function f by minimizing the prediction error on the training set

Testing: apply f to a never before seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

Slide credit: Svetlana Lazebnik

Types of models

CLASSIFICATION

Model outputs comes from a finite set of values

Discrete result

Examples:

- What type of animal is this a picture of?
- Predicting the weather (sunny, cloudy, or rainy?)
- Ranking: Is this result *better* than this result?

REGRESSION

Model outputs are continuous values

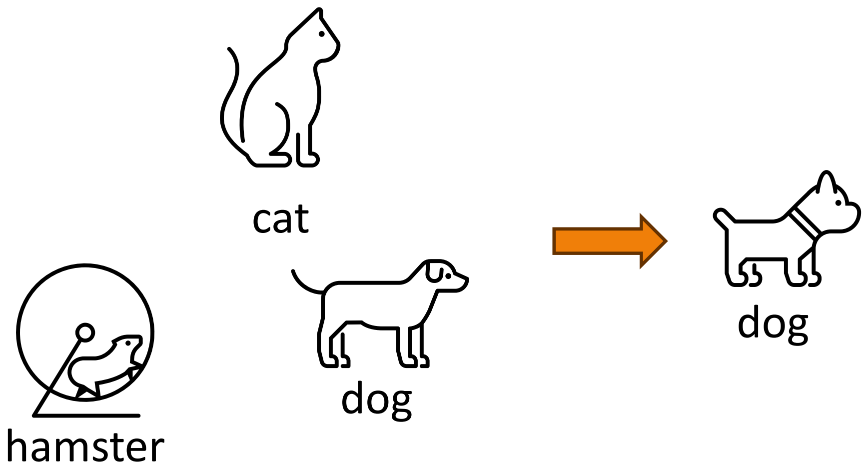
Continuous result

Examples:

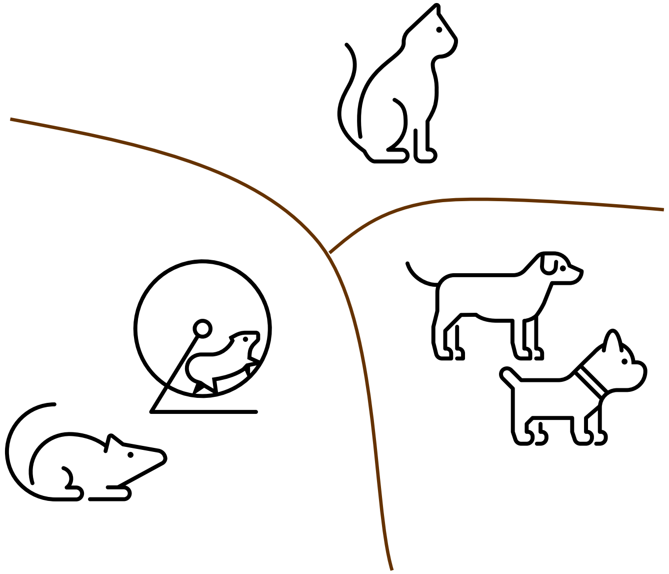
- How far will I move if I drive my motors at this speed for 1 second?
- Predicting the weather (temperature)
- Ranking: *how good* is this result?

Types of Learning

SUPERVISED LEARNING



UNSUPERVISED LEARNING



Types of Learning

SUPERVISED LEARNING

Data has feedback (labels)

Data consists of input-output pairs

Learn mapping from input to output

Examples:

- Dataset classification
- How likely is it that this person will get into a car accident?

UNSUPERVISED LEARNING

No explicit feedback in data

Learn patterns directly from data

Examples:

- Clustering
- Do these people fall under multiple groups?

What are some other examples of these?

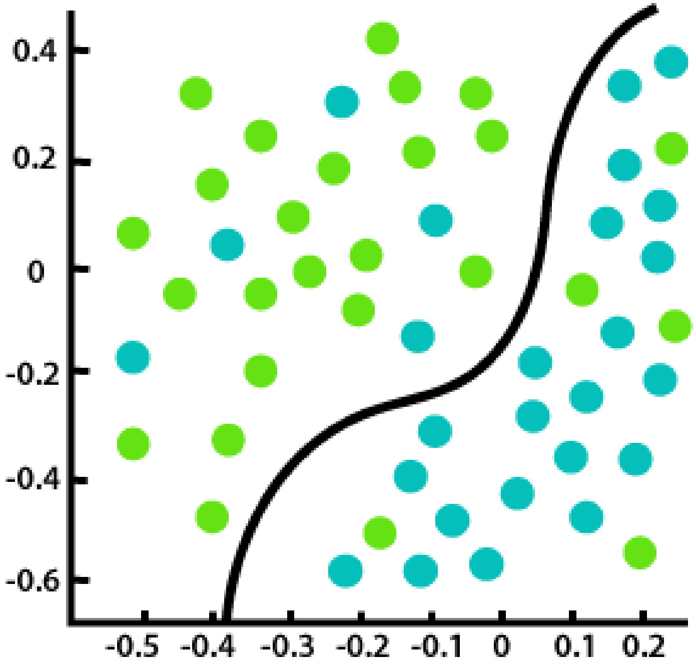
SUPERVISED LEARNING

- Machine translation
- Object segmentation (vision)
- Document classification

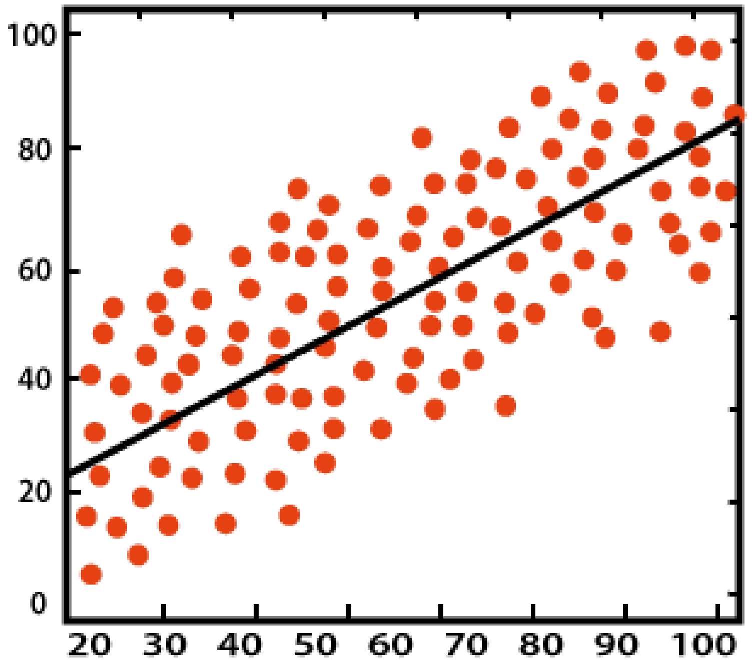
UNSUPERVISED LEARNING

- Clustering
- Language modeling

Types of models



Classification



Regression

https://medium.com/unpackai/classification-regression-in-machine-learning-7cf3b13b0b09

What are some other examples of these?

CLASSIFICATION

Tone tagging

Sentiment classification

Named entity recognition

REGRESSION

Quantity/scale of how much it sounds like a specific author

Numerical sentiment value

Political “score” from document

Likelihoods

Predicted Goodreads score

Types of Algorithms

	Supervised Learning	Unsupervised Learning
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

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

Different ways of categorizing tasks

By purpose:

- Capabilities: test key abilities (linguistic, social, cultural, etc.) of language understanding
  e.g., part-of-speech tagging, parsing, commonsense reasoning
- Application: a use case with potential products in mind
  e.g., machine translation, question answering
- NLP + X: new dimensions of capabilities and applications
  e.g., multilingual, multimodal

By model:

- Classification: output is a categorical variable
- Structured prediction: output is a chain, tree, or graph
- Generation: output is free-form text

Slide courtesy He He 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: Questions to consider...

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

Input

Electronic alerts have been used to assist the authorities in moments of chaos and potential danger: after the Boston bombing in 2013, when the Boston suspects were still at large, and last month in Los Angeles, during an active shooter scare at the airport.

Output

TECH

NOT TECH

Review: Document Classification Features

Electronic alerts have been used to assist the authorities in moments of chaos and potential danger: after the Boston bombing in 2013, when the Boston suspects were still at large, and last month in Los Angeles, during an active shooter scare at the airport.

TECH

NOT TECH

feature extraction

feature $f_i(x)$	value
alerts	1
assist	1
bombing	1
Boston	2
…	
sniffle	0
…	

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

Review: Document Classification Applications

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

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

Review: Token Classification

Word pronunciation

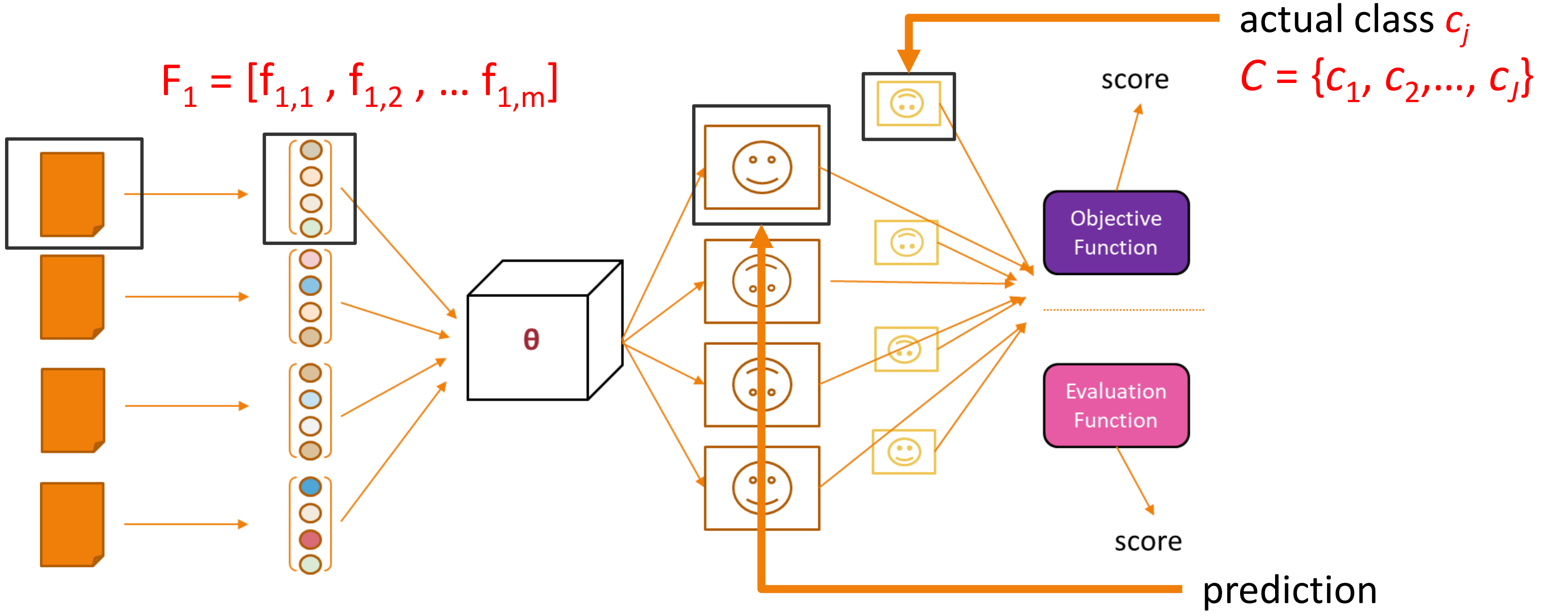
Word sense disambiguation (WSD) within or across languages

Accent restoration

…

Applications

features F_1 extracted from word w_1 and its surrounding words (context)



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

Example: Part of Speech Tagging

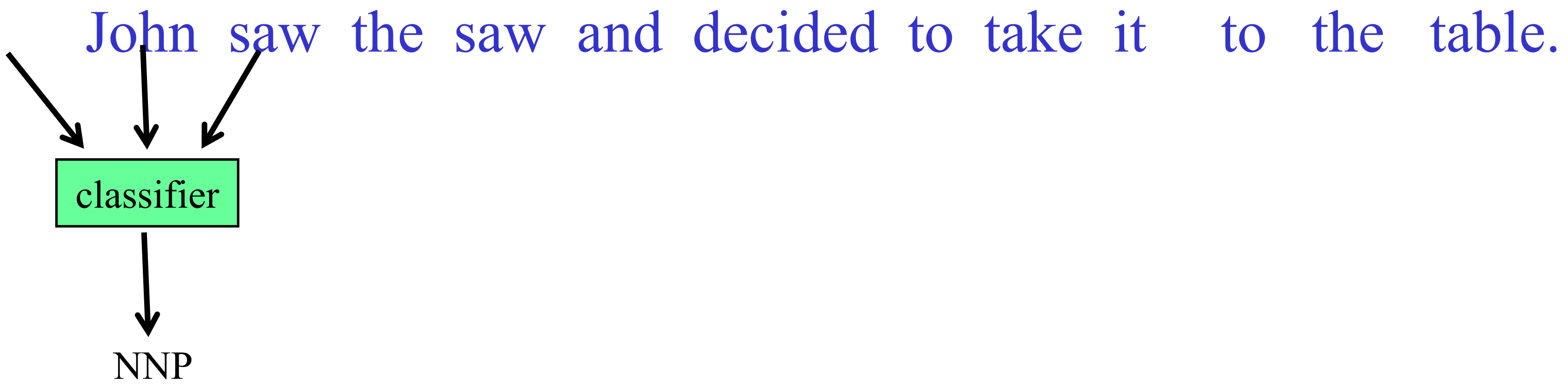
We could treat tagging as a token classification problem

- 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

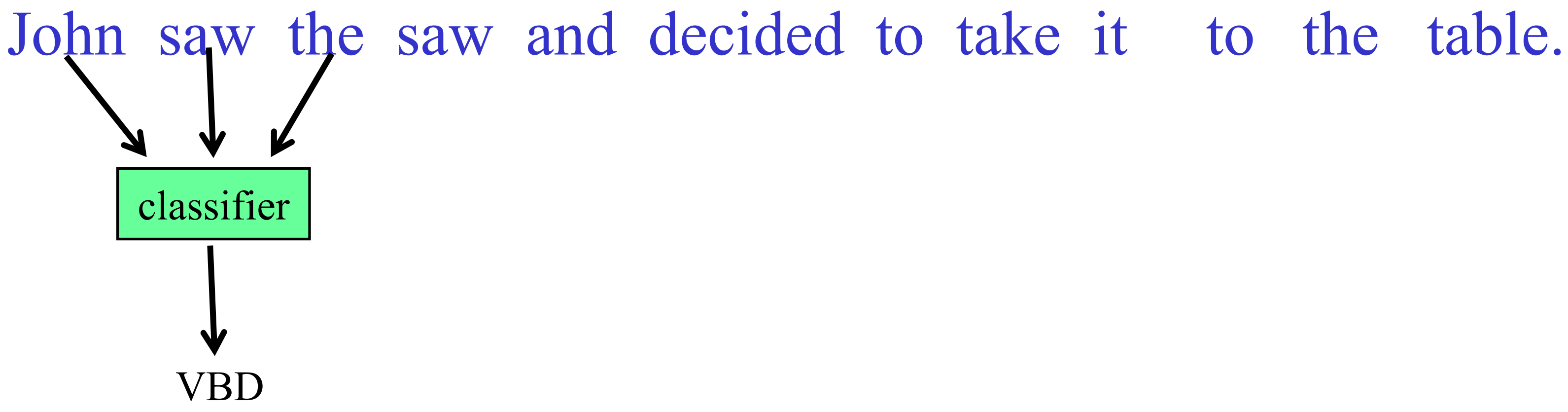
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

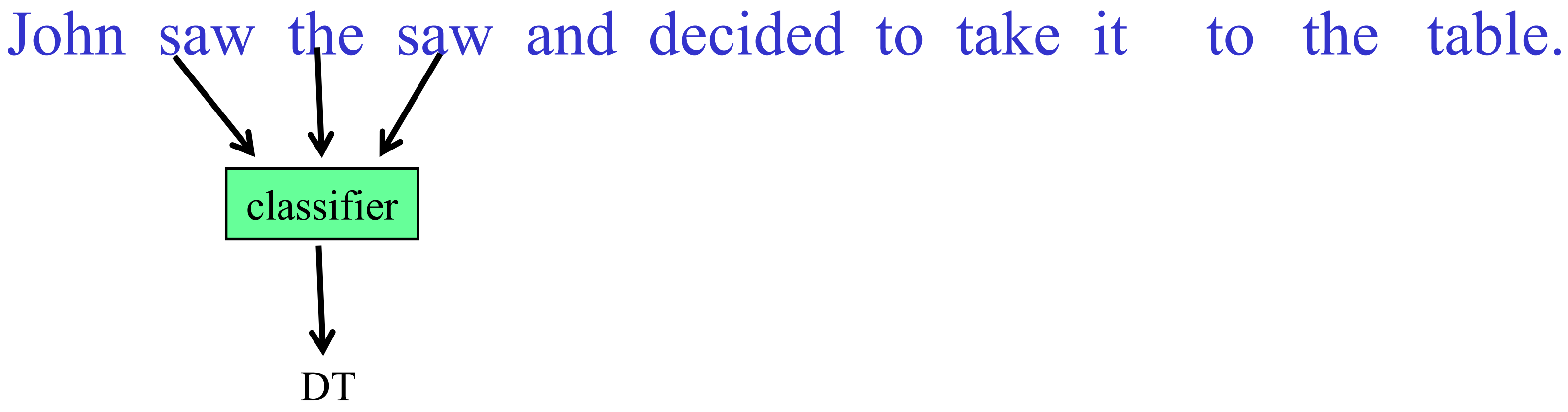
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

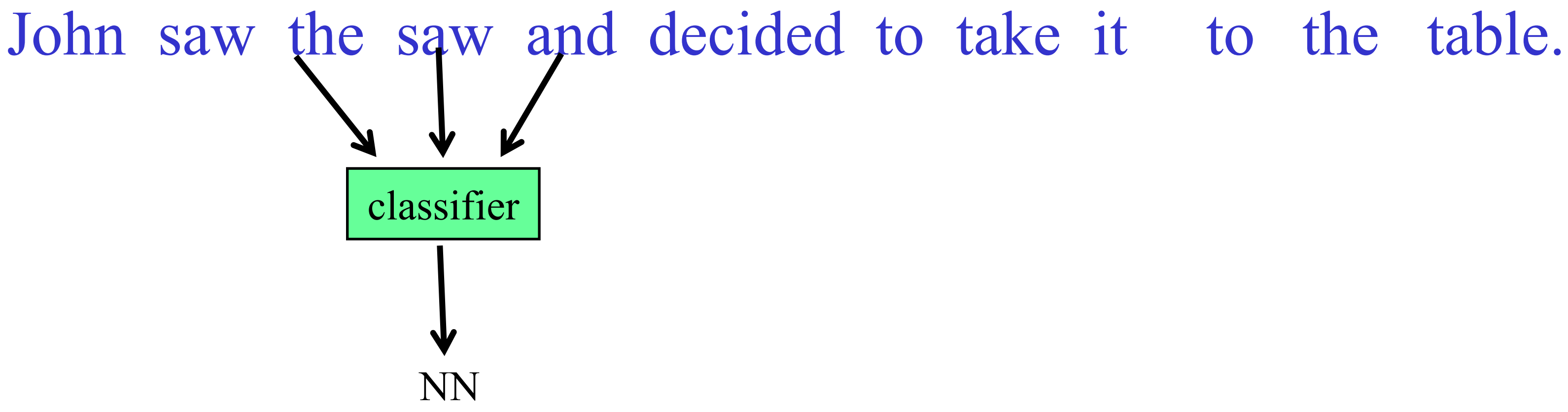
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

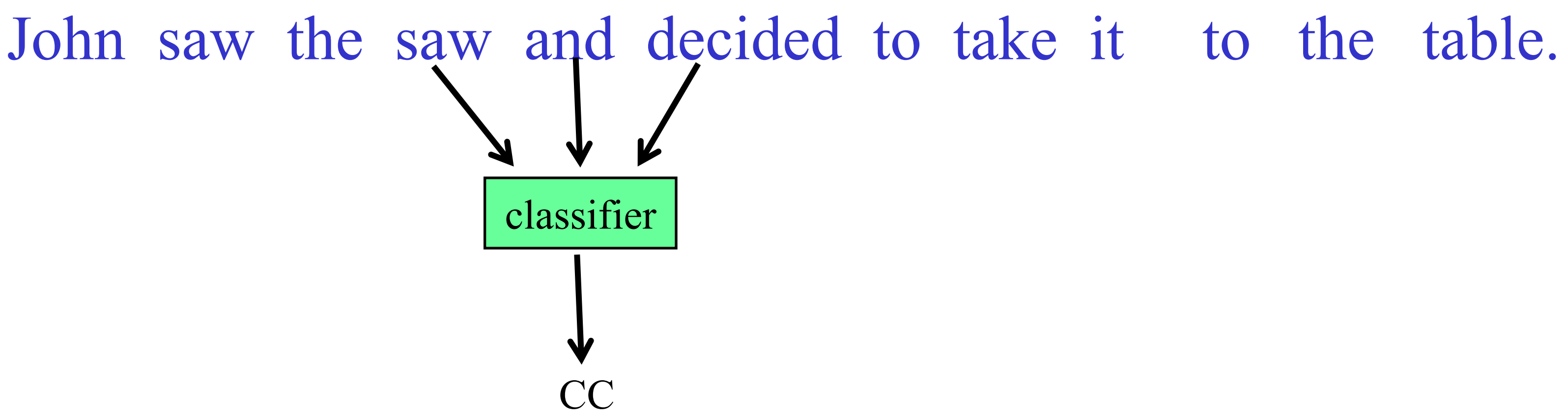
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

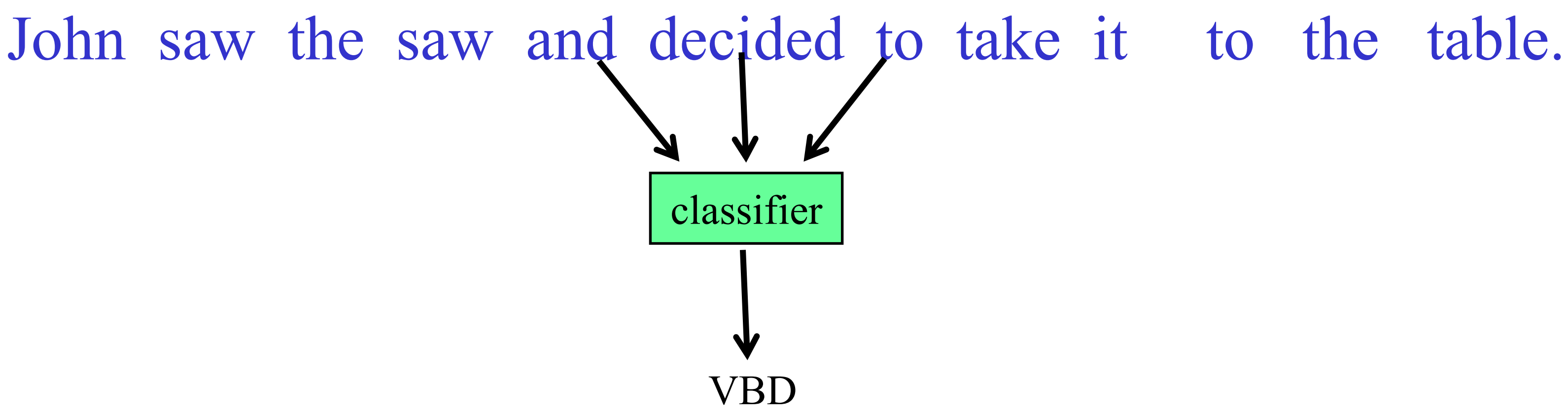
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

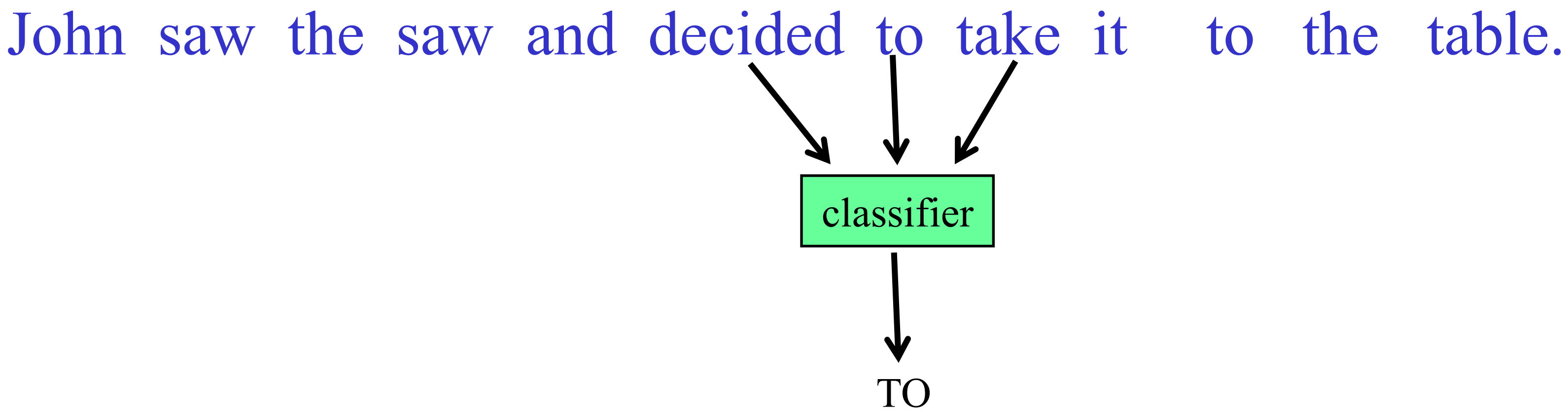
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

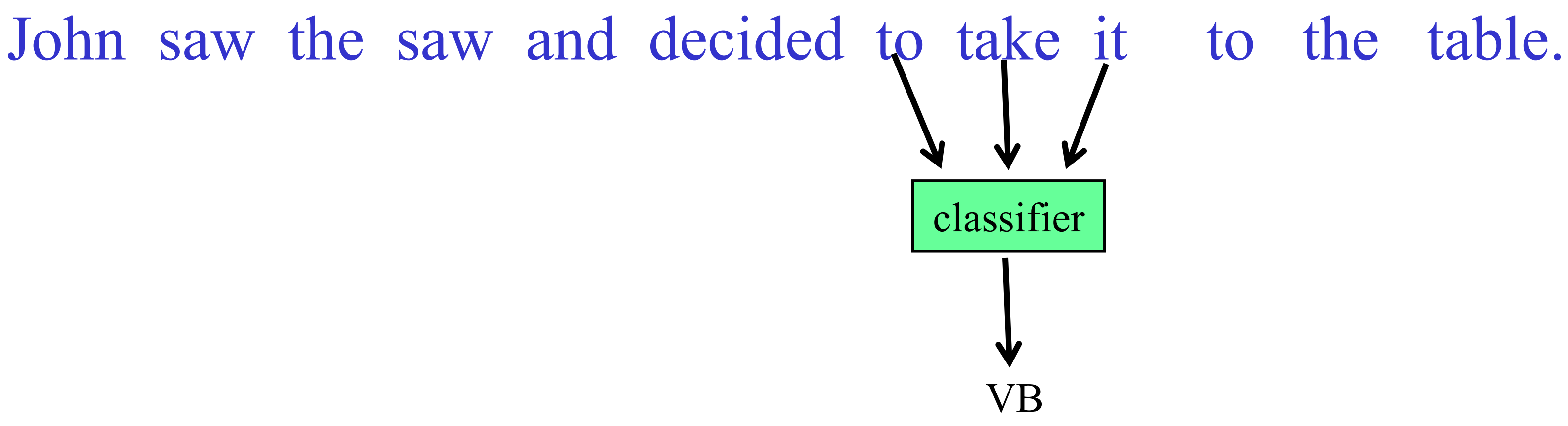
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

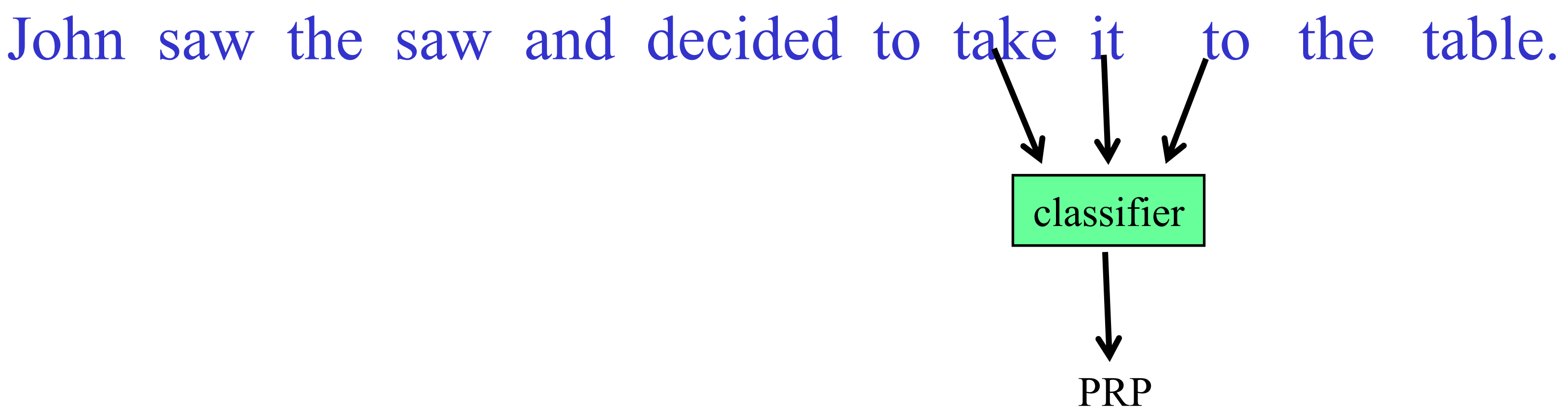
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

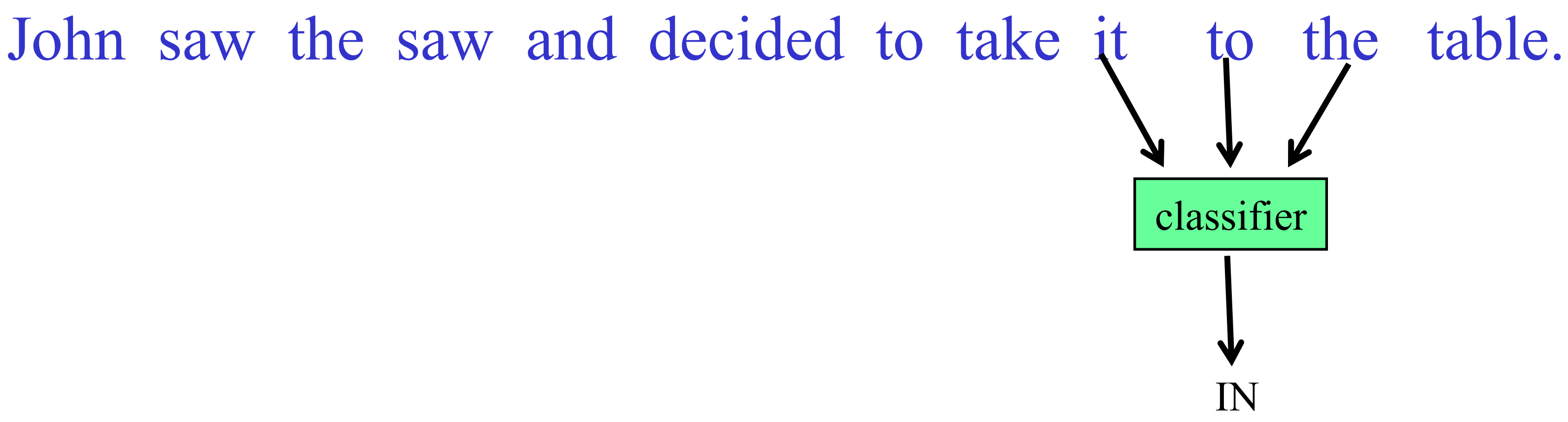
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

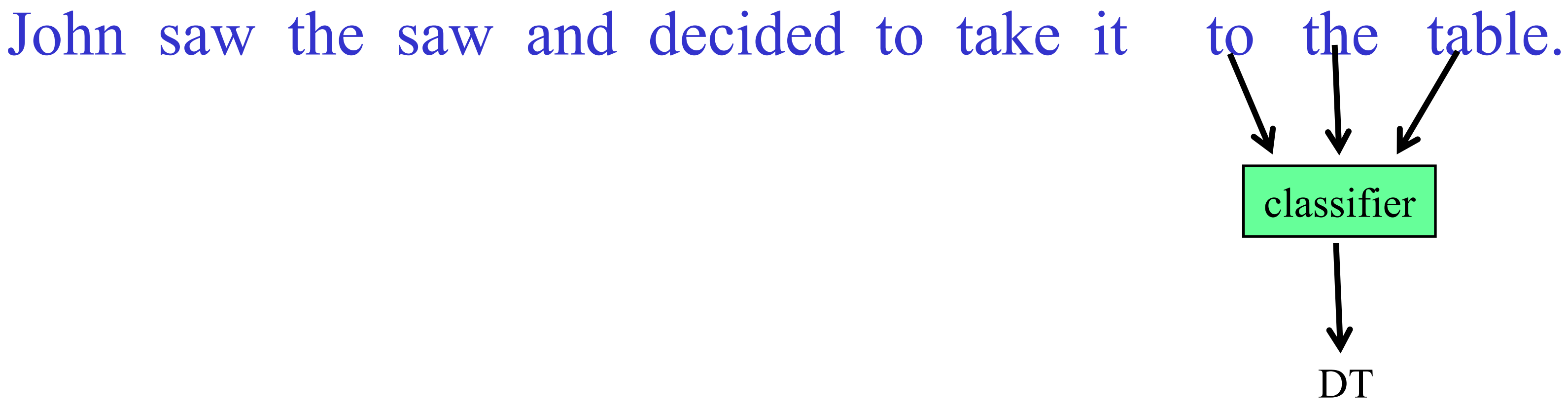
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



Slide courtesy Ray Mooney, with mild edits

Sequence Labeling as Classification

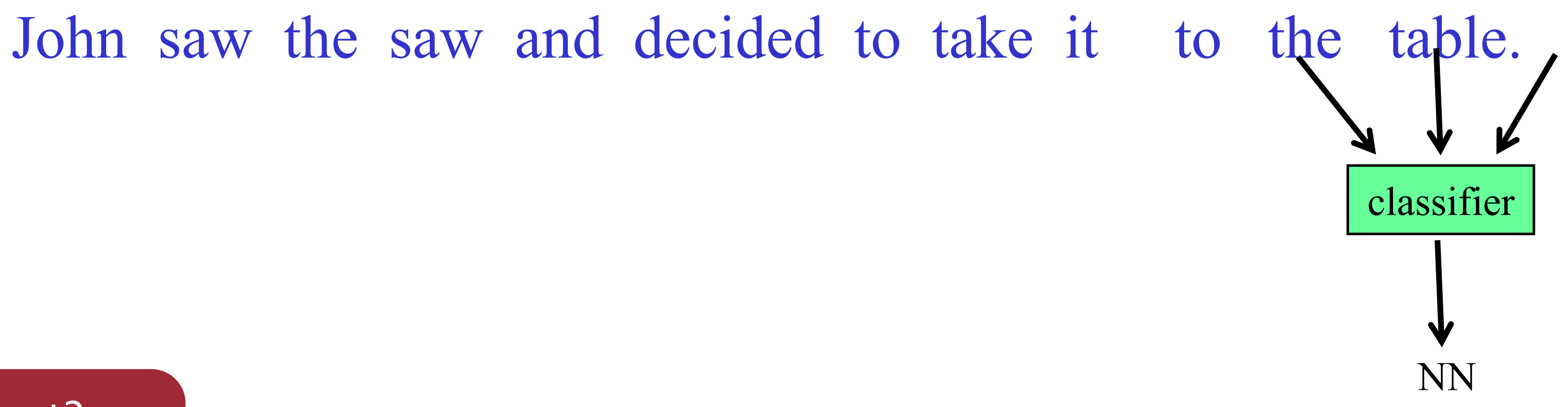
Classify each token independently but use as input features, information about the surrounding tokens (sliding window).



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?
What are the features?
What types of applications?

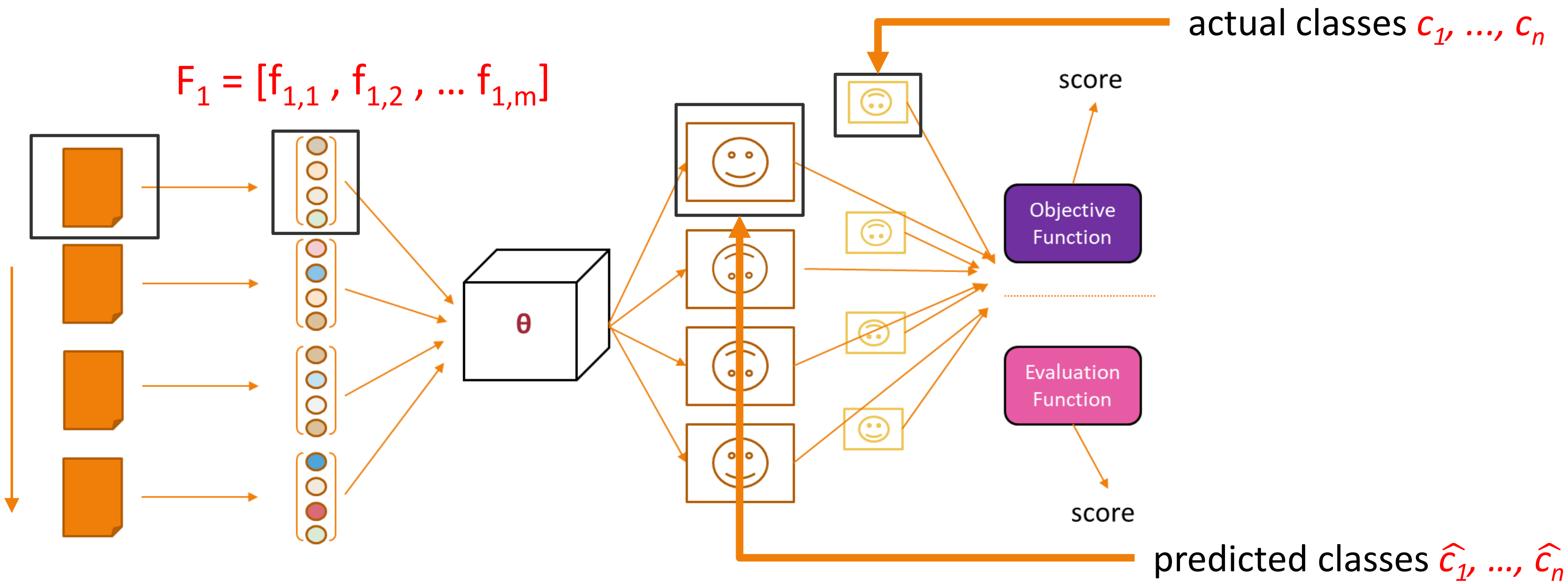
Slide courtesy Ray Mooney, with mild edits

Token Classification in a Sequence Input/Output

p(class | token in context, classes of surrounding words)

features F_1 extracted from word w_1 and its surrounding words (context)

Sliding window

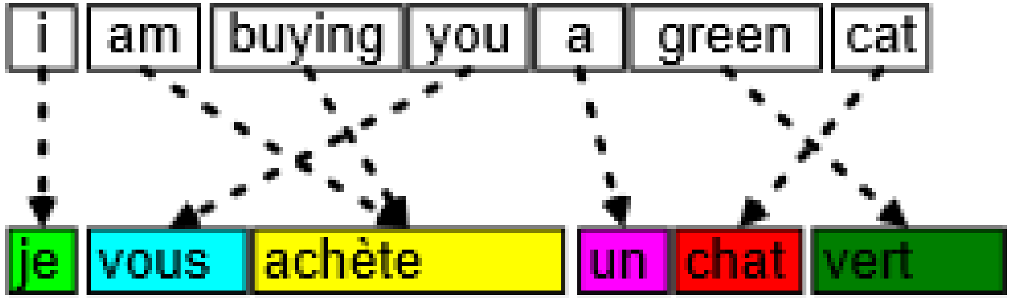


Token Classification in a Sequence Applications

Part of speech tagging

Word alignment

Machine Translation: Word Alignment



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

- Person names
- Organizations (companies, government organisations, committees, etc.)
- Locations (cities, countries, rivers, etc.)
- Date and time expressions
- Measures (percent, money, weight, etc.),
- email addresses, web addresses, street addresses, etc.
- Domain-specific: names of drugs, medical conditions,
- names of ships, bibliographic references etc.

Cunningham and Bontcheva (2003, RANLP Tutorial)

NE Types

Type	Tag	Sample Categories
People	PER	Individuals, fictional characters, small groups
Organization	ORG	Companies, agencies, political parties, religious groups, sports teams
Location	LOC	Physical extents, mountains, lakes, seas
Geo-Political Entity	GPE	Countries, states, provinces, counties
Facility	FAC	Bridges, buildings, airports
Vehicles	VEH	Planes, trains, and automobiles

Type	Example
People	*Turing* is often considered to be the father of modern computer science.
Organization	The *IPCC* said it is likely that future tropical cyclones will become more intense.
Location	The *Mt. Sanitas* loop hike begins at the base of *Sunshine Canyon*.
Geo-Political Entity	*Palo Alto* is looking at raising the fees for parking in the University Avenue district.
Facility	Drivers were advised to consider either the *Tappan Zee Bridge* or the *Lincoln Tunnel*.
Vehicles	The updated *Mini Cooper* retains its charm and agility.

Slide courtesy Jim Martin

Named Entity Recognition

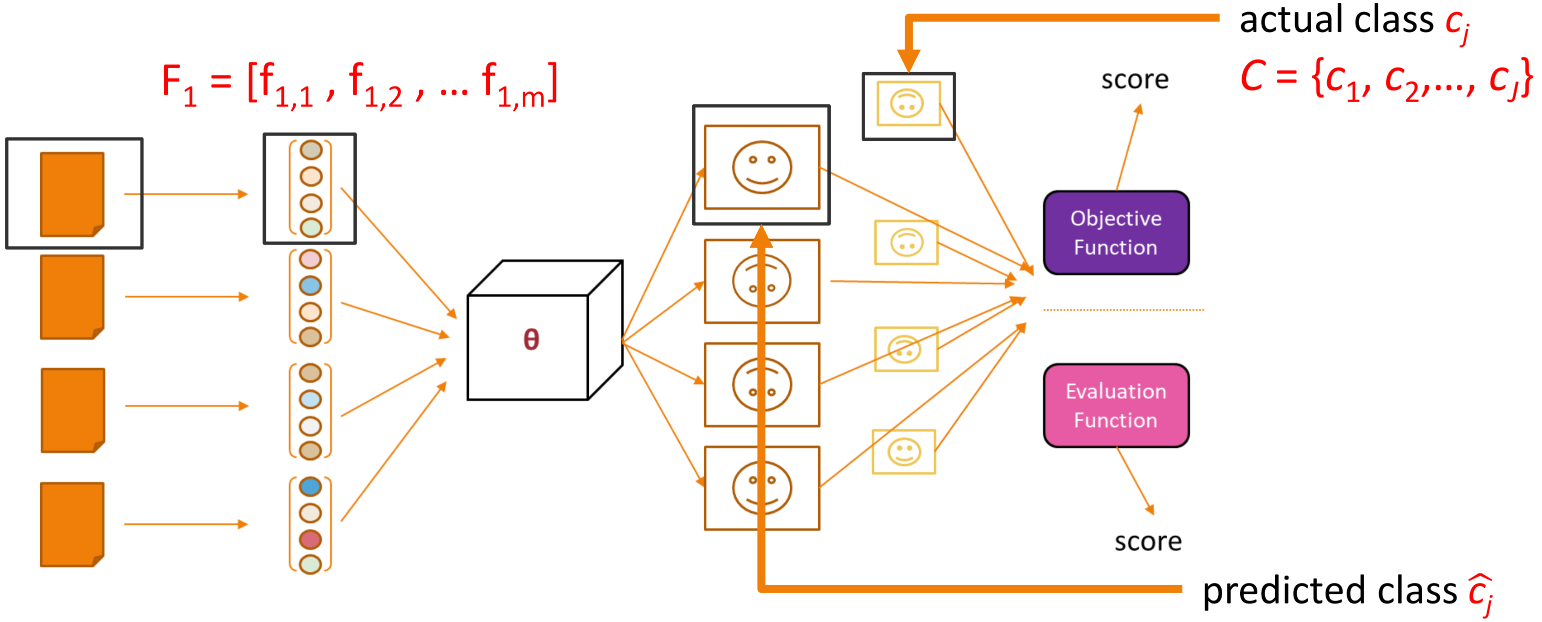
CHICAGO (AP) — Citing high fuel prices, United Airlines said Friday it has increased fares by $6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit AMR, immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL, said the increase 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.

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

Slide courtesy Jim Martin

Chunking Input/Output

features F_1 extracted from phrase p_1 and its surrounding context



p(class | phrase in context)