10-708 Probabilistic Graphical Models

Machine Learning Department
School of Computer Science
Carnegie Mellon University

Course Overview

Matt Gormley
Lecture 1
Feb. 01, 2021

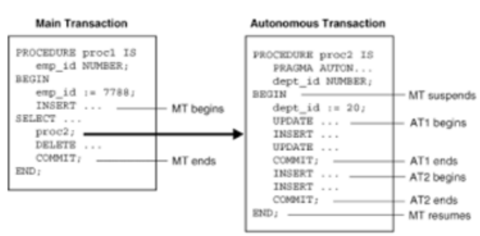
How to define a structured prediction problem

STRUCTURED PREDICTION

Structured vs. Unstructured Data

Structured Data Examples

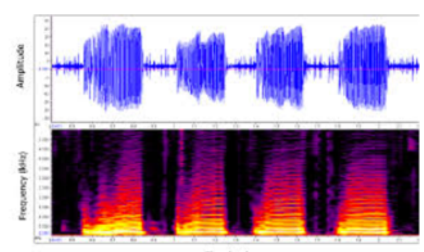
- database entries
- transactional information
- wikipedia infobox
- knowledge graphs
- hierarchies



Unstructured Data Examples

- written text
- images
- videos
- spoken language
- music
- sensor data

مساء الخير! مرحبا بكم في الدرجة



Structured Prediction

- The focus of most Intro ML courses is **classification**
  - Given observations: $\boldsymbol{x} = (x_1, x_2, \ldots, x_K)$
  - Predict a (binary) **label:** y
- Many real-world problems require **structured prediction**
  - Given observations: $\boldsymbol{x} = (x_1, x_2, \ldots, x_K)$
  - Predict a **structure:** $\boldsymbol{y} = (y_1, y_2, \ldots, y_J)$
- Some *classification* problems benefit from **latent structure**

Structured Prediction

Classification / Regression

1. Input can be semi-structured data
2. Output is a single number (integer / real)
3. In linear models, features can be arbitrary combinations of [input, output] pair
4. Output space is small
5. Inference is trivial

Structured Prediction

1. Input can be semi-structured data
2. Output is a sequence of numbers representing a structure
3. In linear models, features can be arbitrary combinations of [input, output] pair
4. Output space may be exponentially large in the input space
5. Inference problems are NP-hard or #P-hard in general and often require approximations

Structured Prediction Examples

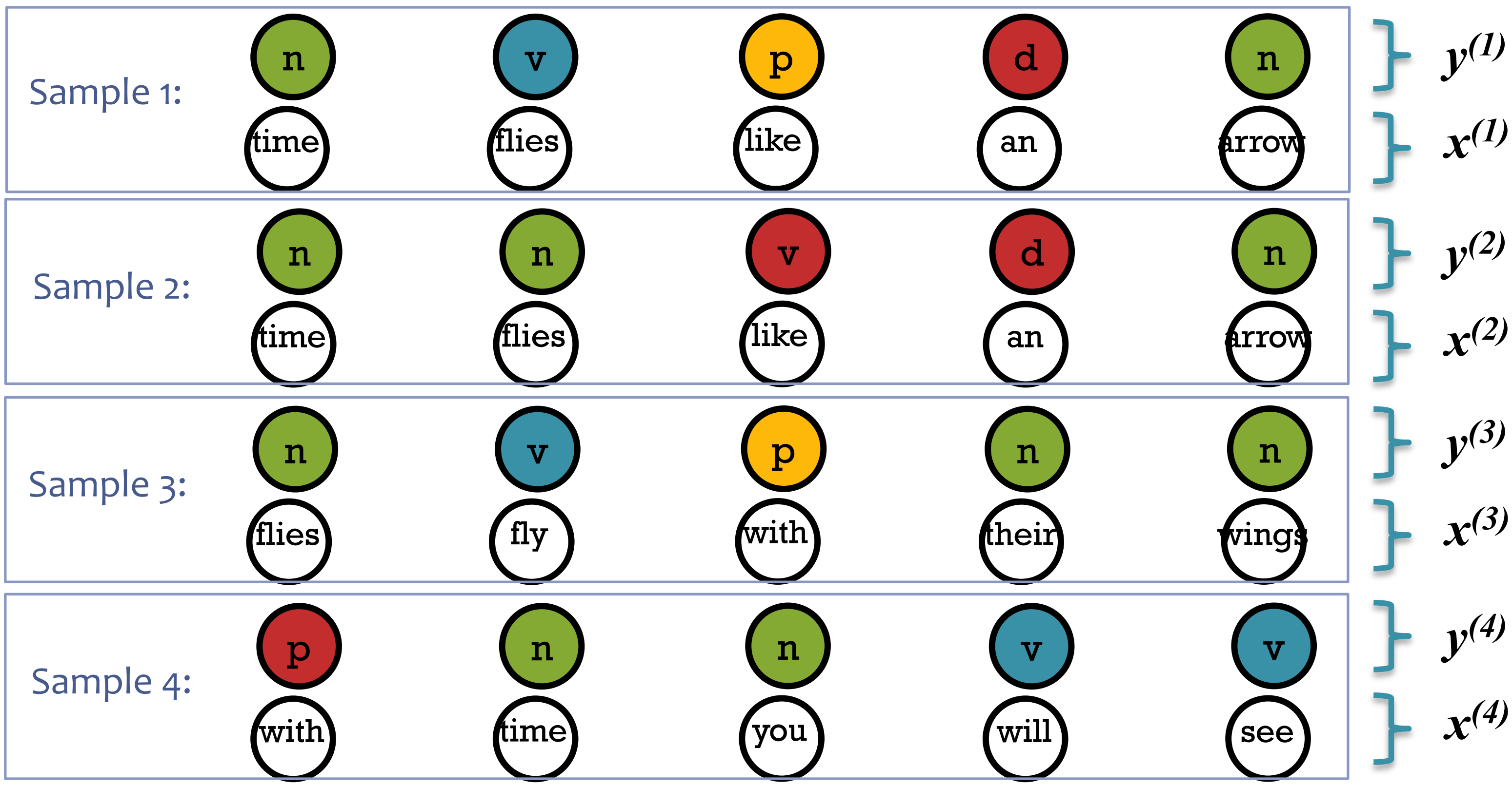
- **Examples of structured prediction**
  - Part-of-speech (POS) tagging
  - Handwriting recognition
  - Speech recognition
  - Object detection
  - Scene understanding
  - Machine translation
  - Protein sequencing

Part-of-Speech (POS) Tagging

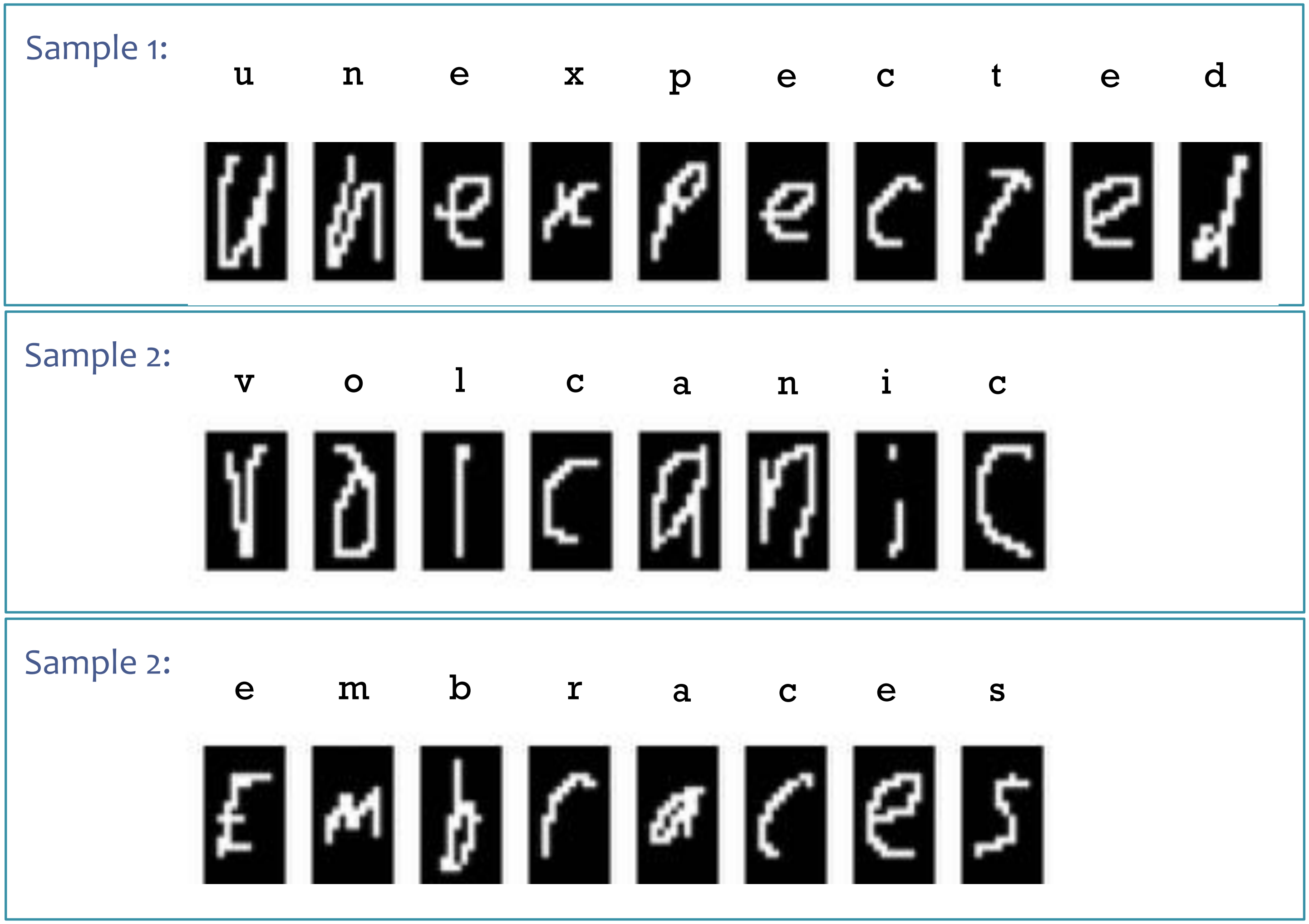
Sample 1:	n time	v flies	p like	d an	n arrow
Sample 2:	n time	n flies	v like	d an	n arrow
Sample 3:	n flies	v fly	p with	n their	n wings
Sample 4:	p with	n time	n you	v will	v see

Dataset for Supervised Part-of-Speech (POS) Tagging

Data: $\mathcal{D} = \{\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)}\}_{n=1}^{N}$



Handwriting Recognition



Figures from (Chatzis & Demiris, 2013)

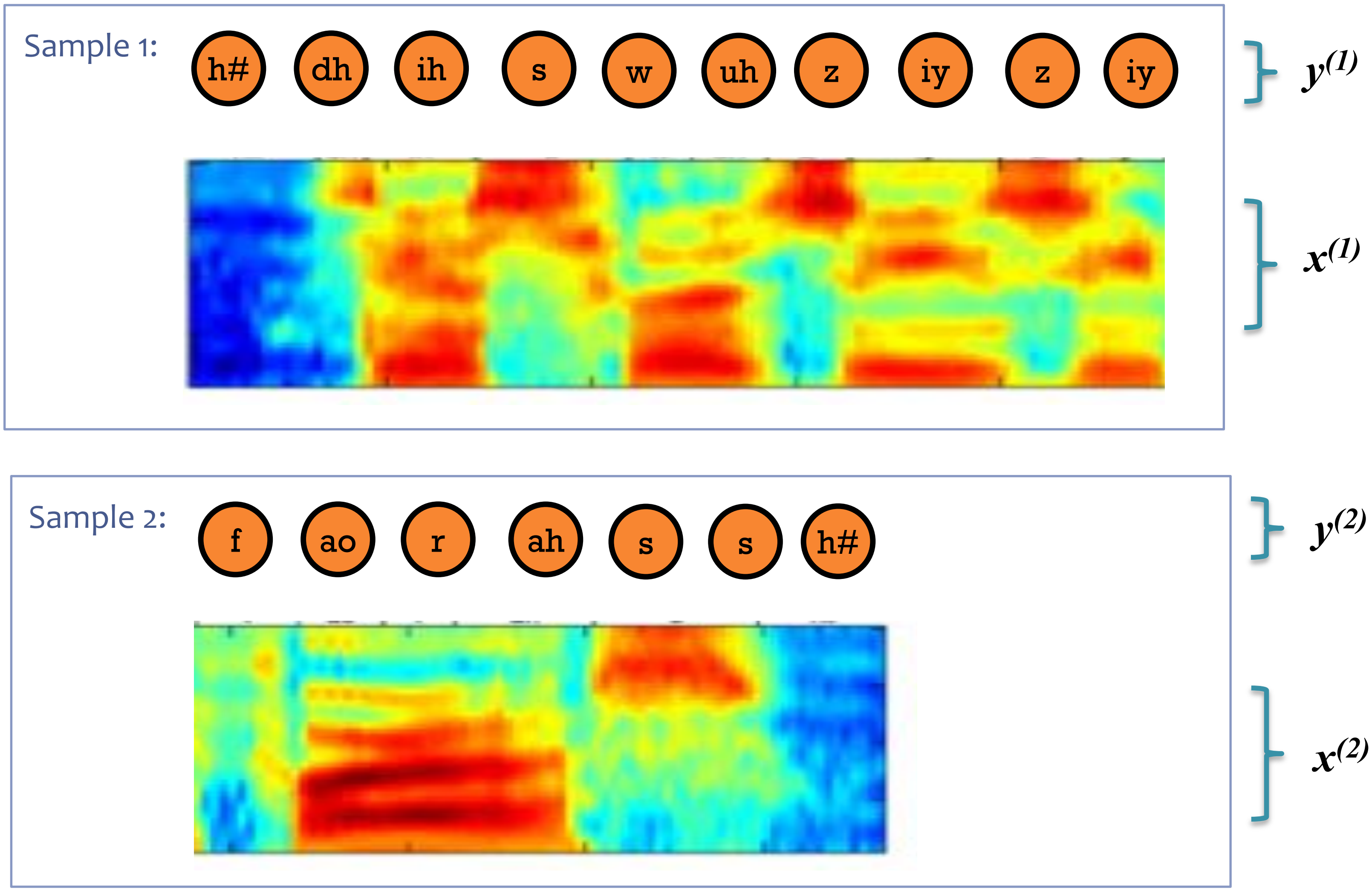
Dataset for Supervised Handwriting Recognition

Data: $\mathcal{D} = \{\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)}\}_{n=1}^{N}$

Figures from (Chatzis & Demiris, 2013)

Dataset for Supervised Phoneme (Speech) Recognition

Data: $\mathcal{D} = \{\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)}\}_{n=1}^{N}$



Figures from (Jansen & Niyogi, 2013)

Case Study: Object Recognition

Data consists of images $\boldsymbol{x}$ and labels y.

$\boldsymbol{x}^{(1)}$

pigeon $\quad y^{(1)}$

$\boldsymbol{x}^{(2)}$

rhinoceros $\quad y^{(2)}$

$\boldsymbol{x}^{(3)}$

leopard $\quad y^{(3)}$

$\boldsymbol{x}^{(4)}$

llama $\quad y^{(4)}$

Case Study: Object Recognition

Data consists of images x and labels y.

- Preprocess data into "patches"
- Posit a latent labeling z describing the object's parts (e.g. head, leg, tail, torso, grass)

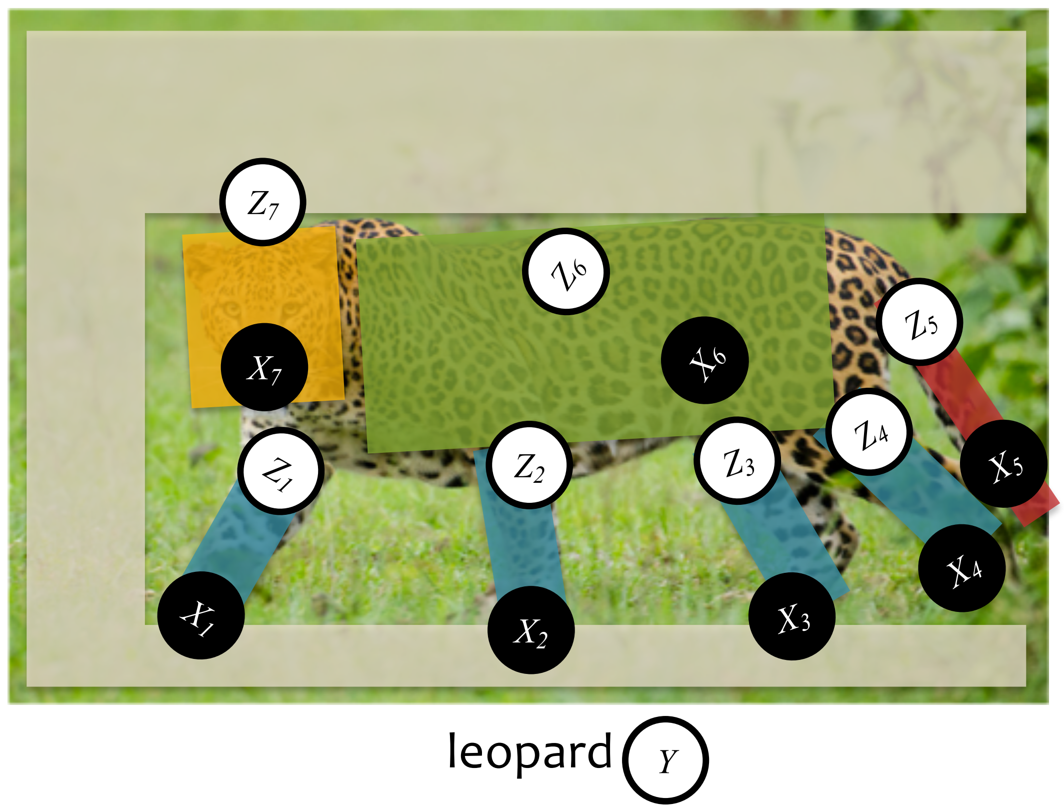
- Define graphical model with these latent variables in mind
- z is not observed at train or test time

leopard

Case Study: Object Recognition

Data consists of images x and labels y.

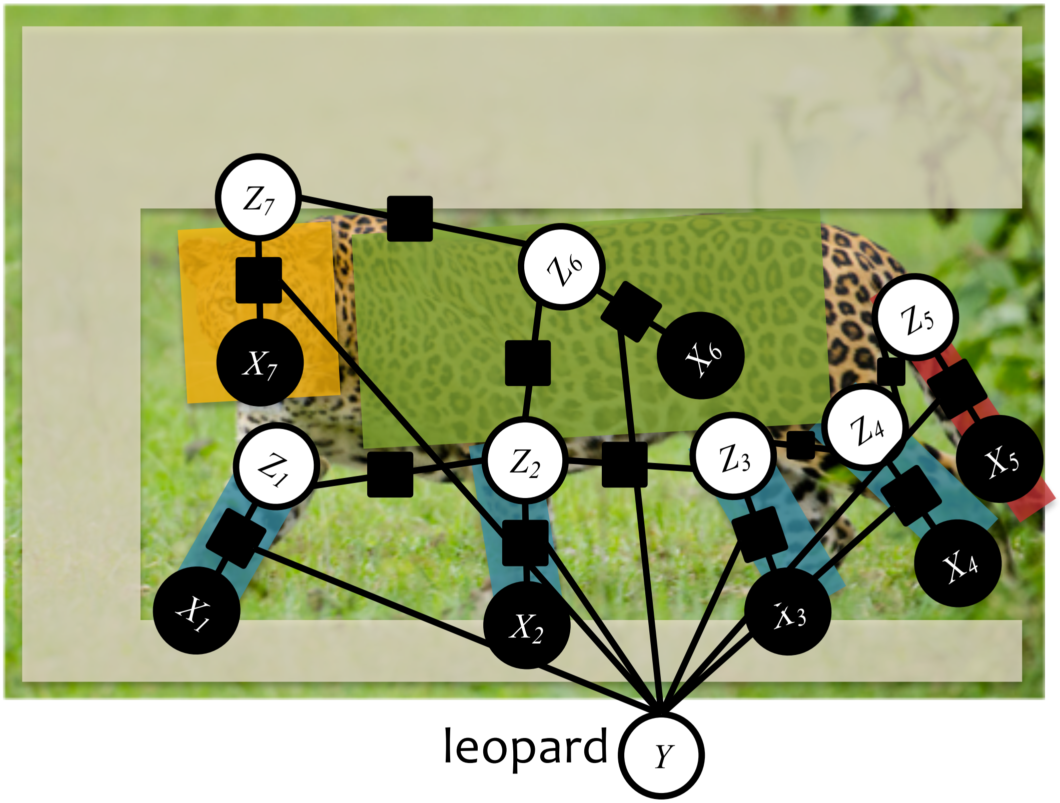
- Preprocess data into "patches"
- Posit a latent labeling z describing the object's parts (e.g. head, leg, tail, torso, grass)
- Define graphical model with these latent variables in mind
- z is not observed at train or test time



Case Study: Object Recognition

Data consists of images x and labels y.

- Preprocess data into "patches"
- Posit a latent labeling z describing the object's parts (e.g. head, leg, tail, torso, grass)
- Define graphical model with these latent variables in mind
- z is not observed at train or test time



Structured Prediction

Preview of challenges to come…

- Consider the task of finding the **most probable assignment** to the output

Classification

$$\hat{y} = \operatorname*{argmax}_{y} p(y|\mathbf{x})$$

where $y \in \{+1, -1\}$

Structured Prediction

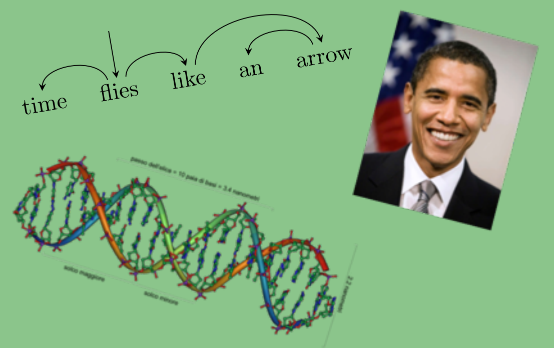
$$\hat{\mathbf{y}} = \operatorname*{argmax}_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})$$

where $\mathbf{y} \in \mathcal{Y}$

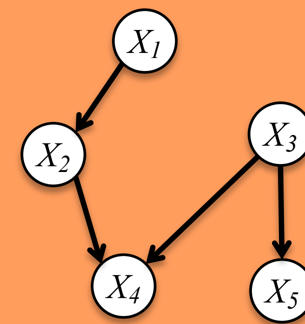
and $|\mathcal{Y}|$ is very large

Structured Prediction

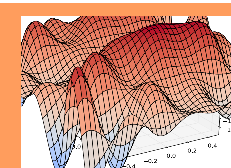
Data



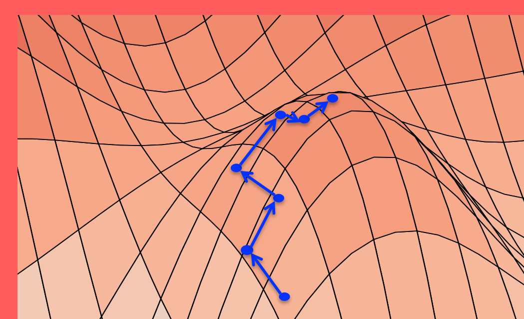
Model



Objective



Learning



Inference

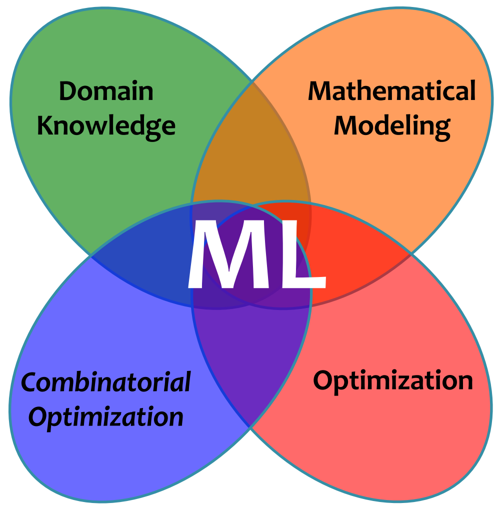
(**Inference** is usually called as a subroutine in learning)

Structured Prediction

The **data** inspires the structures we want to predict

Our **model** defines a score for each structure

It also tells us what to optimize



Inference finds {best structure, marginals, partition function} for a new observation

Learning tunes the parameters of the model

(**Inference** is usually called as a subroutine in learning)

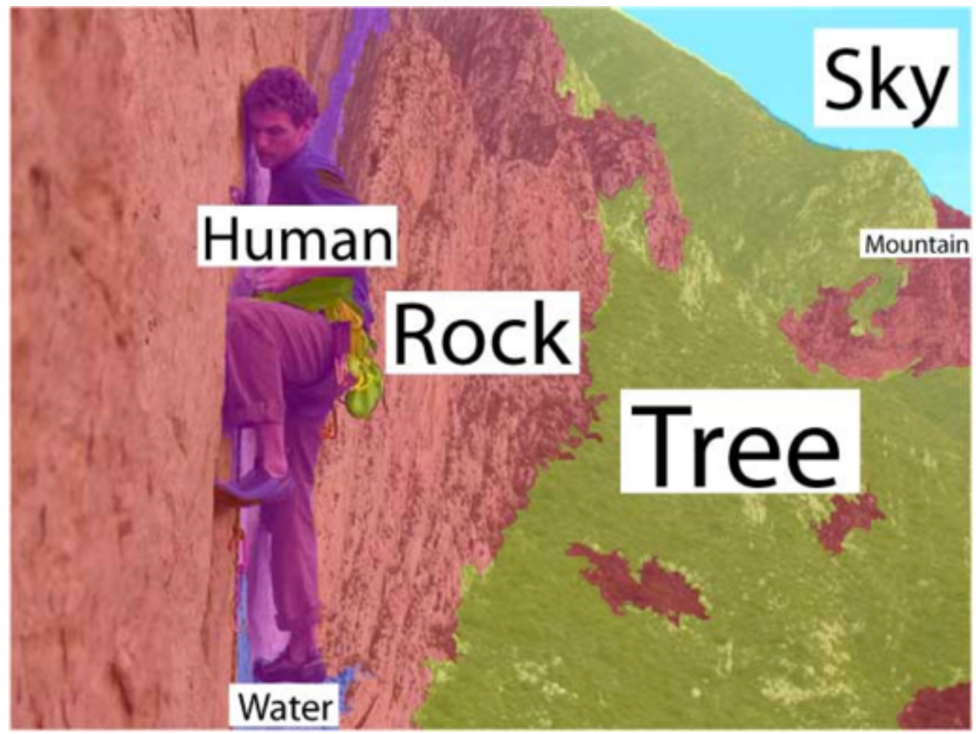
Decomposing a Structure into Parts

- Why divide a **structure** into its **pieces**?
  - amenable to **efficient inference**
  - enable natural **parameter sharing** during learning
  - easier definition of fine-grained **loss functions**
  - clearer depiction of **model's uncertainty**
  - easier specification of **interactions** between the parts
  - (may) lead to natural definition of a **search problem**
- A key step in **formulating a task as a structured prediction**

Scene Understanding

- **Variables:**
  - boundaries of image regions
  - tags of regions
- **Interactions:**
  - semantic plausibility of nearby tags
  - continuity of tags across visually similar regions (i.e. patches)

Labels **with** top-down information

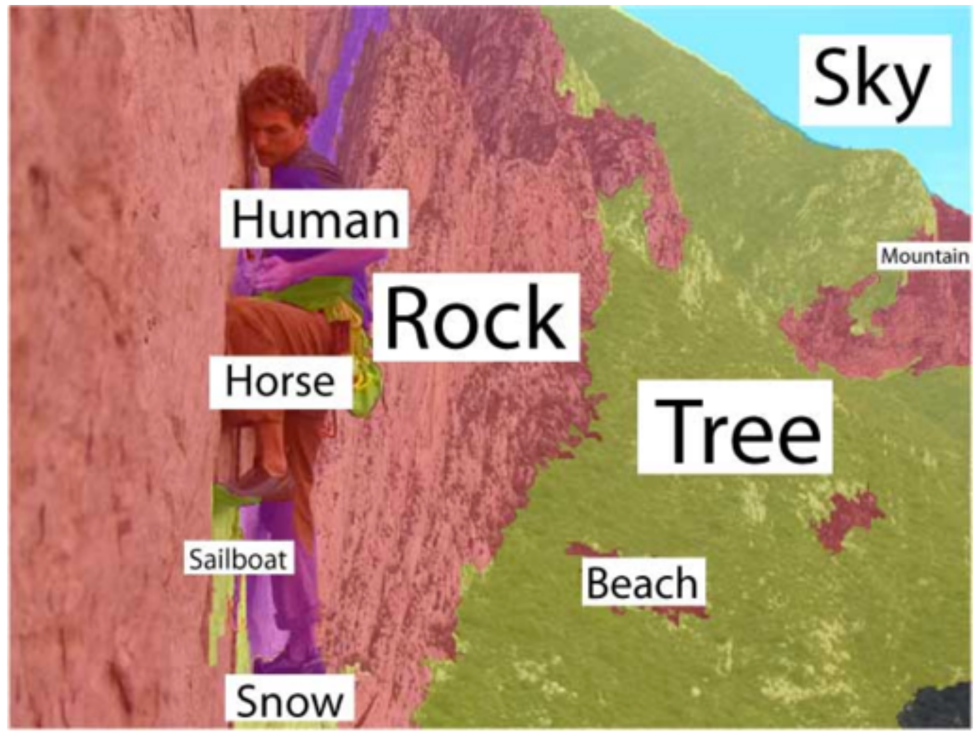


(Li et al., 2009)

Scene Understanding

- **Variables:**
  - boundaries of image regions
  - tags of regions
- **Interactions:**
  - semantic plausibility of nearby tags
  - continuity of tags across visually similar regions (i.e. patches)

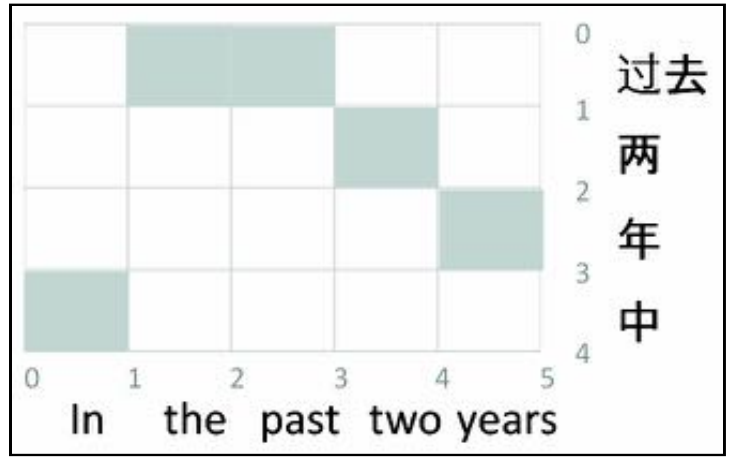
Labels **without** top-down information

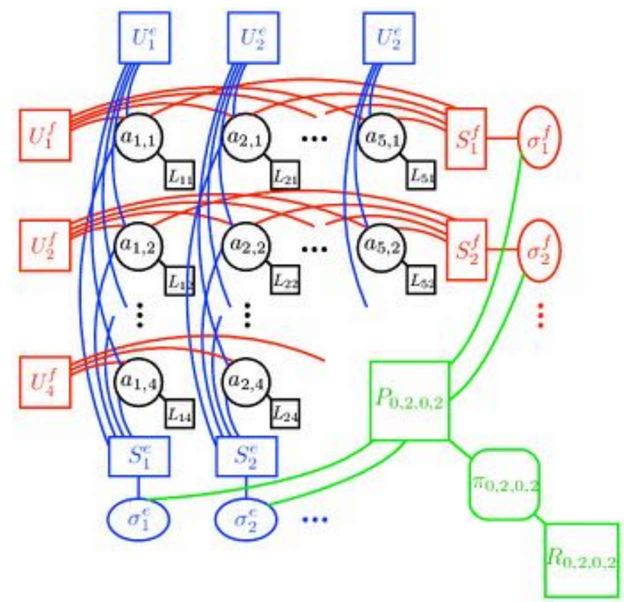


(Li et al., 2009)

Word Alignment / Phrase Extraction

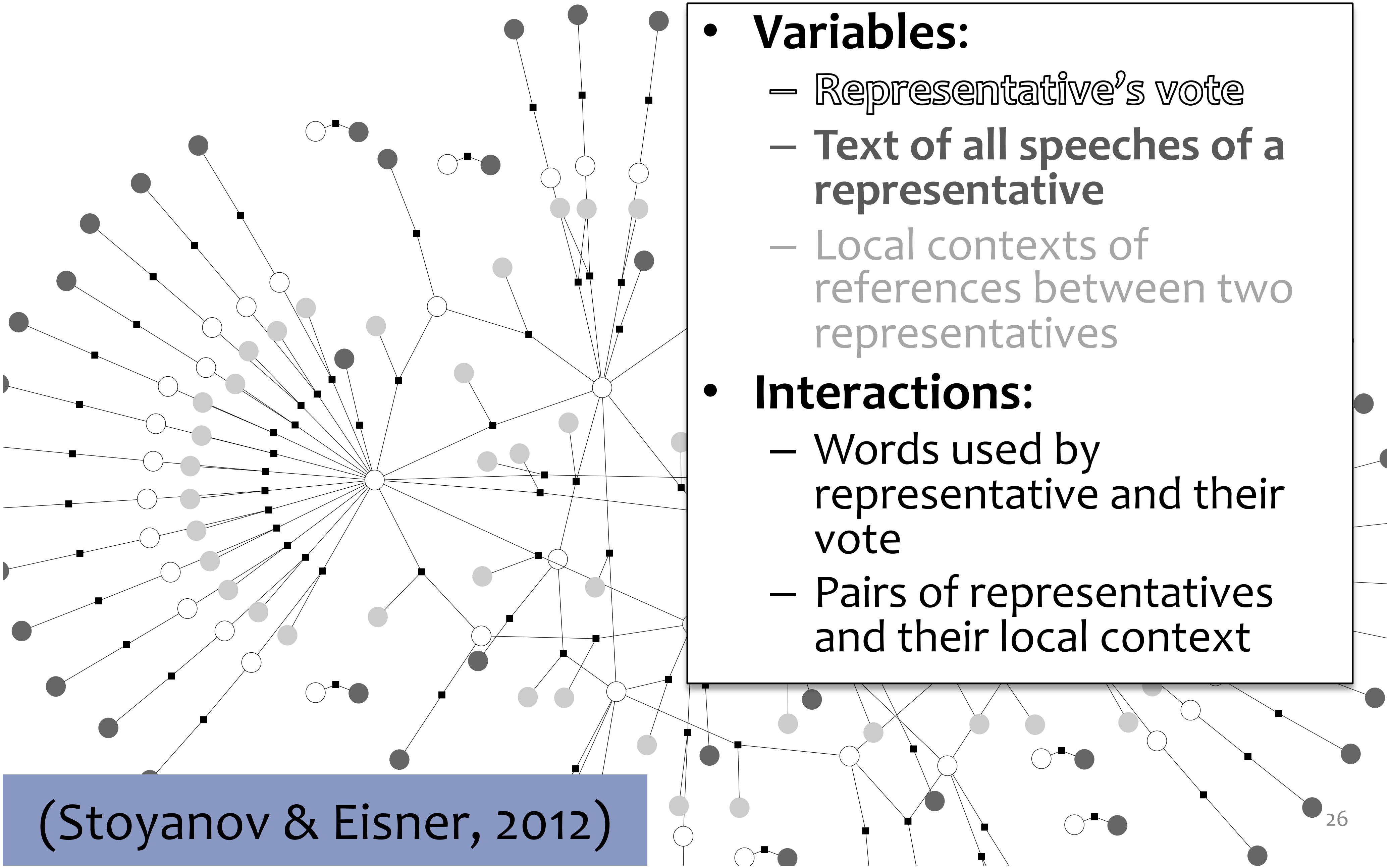
- **Variables (boolean):**
  - For each (Chinese phrase, English phrase) pair, are they linked?
- **Interactions:**
  - Word fertilities
  - Few “jumps” (discontinuities)
  - Syntactic reorderings
  - “ITG contraint” on alignment
  - Phrases are disjoint (?)





(Burkett & Klein, 2012)

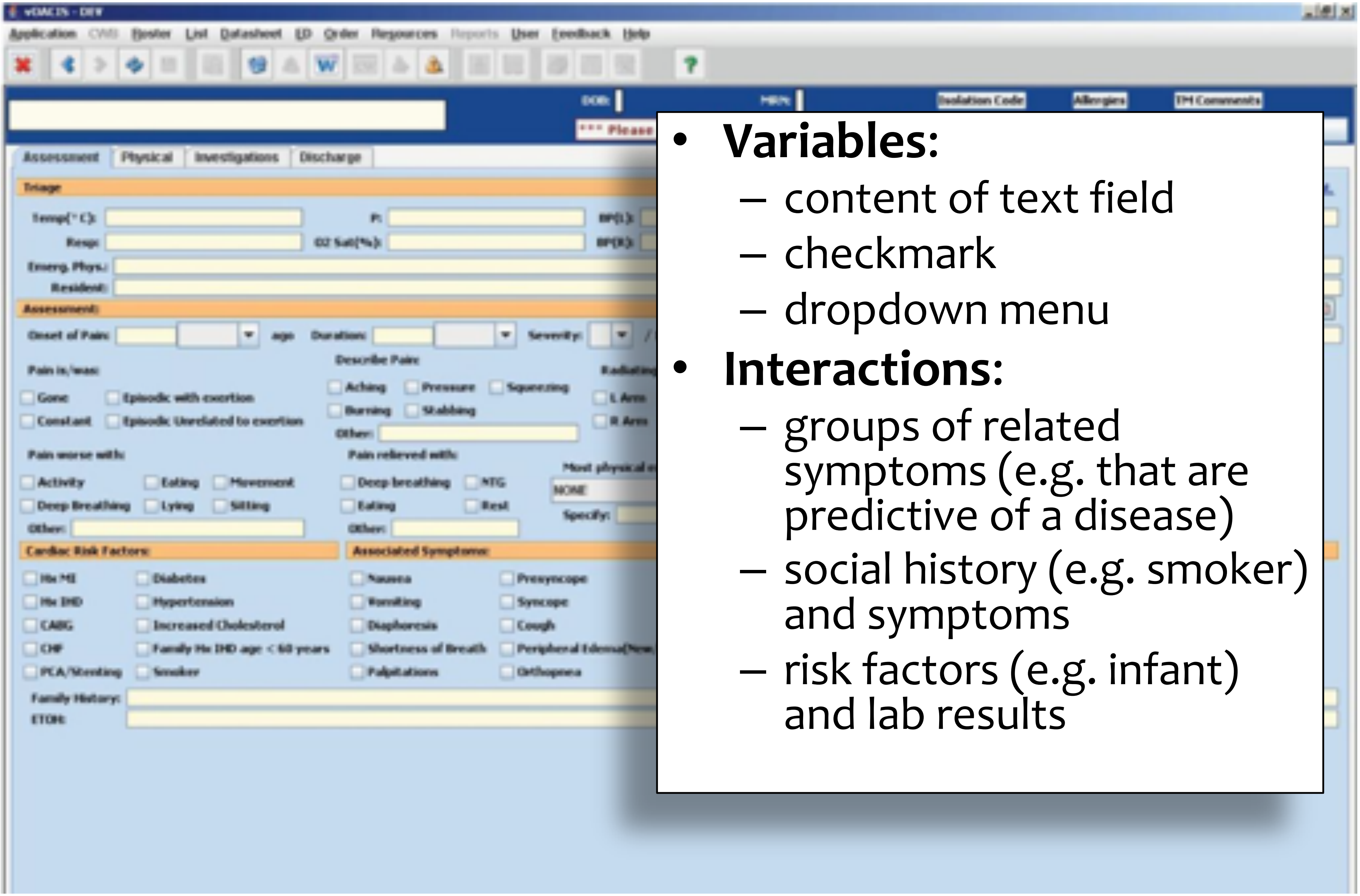
Congressional Voting



- **Variables:**
  - Representative's vote
  - **Text of all speeches of a representative**
  - Local contexts of references between two representatives
- **Interactions:**
  - Words used by representative and their vote
  - Pairs of representatives and their local context

(Stoyanov & Eisner, 2012)

Medical Diagnosis



- **Variables:**
  - content of text field
  - checkmark
  - dropdown menu
- **Interactions:**
  - groups of related symptoms (e.g. that are predictive of a disease)
  - social history (e.g. smoker) and symptoms
  - risk factors (e.g. infant) and lab results

Wikipedia Infoboxes

Gryan Miers	
Personal information	
Date of birth	30 March 1999 (age 20)
Original team(s)	Grovedale (GFL)
Draft	No. 57, 2017 national draft
Debut	Round 1, 2019, Geelong vs. Collingwood, at the MCG
Height	178 cm (5 ft 10 in)
Weight	78 kg (172 lb)
Position(s)	Small forward
Club information	
Current club	Geelong
Number	32
Playing career[1]	

Years	**Club**	**Games (Goals)**
2019–	Geelong	19 (19)

[1] Playing statistics correct to the end of round 17, 2019.

Career highlights

- AFL Rising Star nominee: 2019

Sources: AFL Tables, AustralianFootball.com

Pather Panchali

A poster of *Pather Panchali*

Directed by	Satyajit Ray
Screenplay by	Satyajit Ray
Based on	*Pather Panchali* by Bibhutibhushan Bandyopadhyay
Starring	Subir Banerjee Kanu Banerjee Karuna Banerjee Uma Dasgupta Chunibala Devi Tulsi Chakrabarti
Music by	Ravi Shankar
Cinematography	Subrata Mitra
Edited by	Dulal Dutta
Production company	Government of West Bengal
Distributed by	Aurora Film Corporation (1955) Edward Harrison (1958) Merchant Ivory Productions Sony Pictures Classics (1995)[a]
Release date	26 August 1955 (India)

Changsha Kingdom
長沙國

203/202 BC–AD 33

Silk map unearthed from Mawangdui, showing Changsha and the neighboring kingdom of Nanyue.

Capital	Linxiang (present-day Changsha)
Government	Monarchy
History	
• Established	203/202 BC
• Extinction of the Wu family line	157 BC
• Reestablishment under the Liu family	155 BC
• Dissolution under Wang Mang	AD 9
• Restoration	AD 26
• Disestablished	AD 33

Space Invaders

Promotional flyer

Developer(s)	Taito
Publisher(s)	JP: Taito NA: Midway EU: Midway[1] AU: Leisure & Allied Industries[2] Atari, Inc. (home)
Designer(s)	Tomohiro Nishikado
Platform(s)	Arcade, Atari 2600, Atari 5200, Atari 8-bit, MSX
Release	JP: June 1978[3] NA: July 1978
Genre(s)	Fixed shooter
Mode(s)	Single-player, 2 players alternating
Cabinet	Upright, cocktail[4]
Arcade system	Taito 8080[5]
CPU	8080 @ 2 MHz[5]
Sound	SN76477 @ 1.9968 MHz
Display	Fujitsu MB14241,[6] monochrome raster, vertical orientation, 224×256 resolution[5]

Exercise: Wikipedia Infoboxes

Question:

Suppose you want to populate missing infobox fields. What model would you pick for the job, and why?

A. Multiclass classifier

B. RNN

C. Graphical model

Answer:

Central Park Conservancy

From Wikipedia, the free encyclopedia

Coordinates: 40.76424°N 73.97169°W

The **Central Park Conservancy** is a private, nonprofit park conservancy that manages Central Park under a contract with the City of New York and NYC Parks. The conservancy employs most maintenance and operations staff in the park. It effectively oversees the work of both the private and public employees under the authority of the publicly appointed Central Park administrator, who reports to the parks commissioner and the conservancy's president.[1]

The Central Park Conservancy was founded in 1980 in the aftermath of Central Park's decline in the 1960s and 1970s.[2] Initially devoted to fundraising for projects to restore and improve the park, it took over the park's management duties in 1998.[3] The organization has invested more than $800 million toward the restoration and enhancement of Central Park since its founding.[4] With an endowment of over $200 million, consisting of contributions from residents, corporations, and foundations,[5] the Conservancy provides 75 percent of the Park's $65 million annual operating budget and is responsible for all basic care of the park.[6] The Conservancy also provides maintenance support and staff training programs for other public parks in New York City, and has assisted with the development of new parks, such as the High Line and Brooklyn Bridge Park.[7]:45–46

Central Park Conservancy

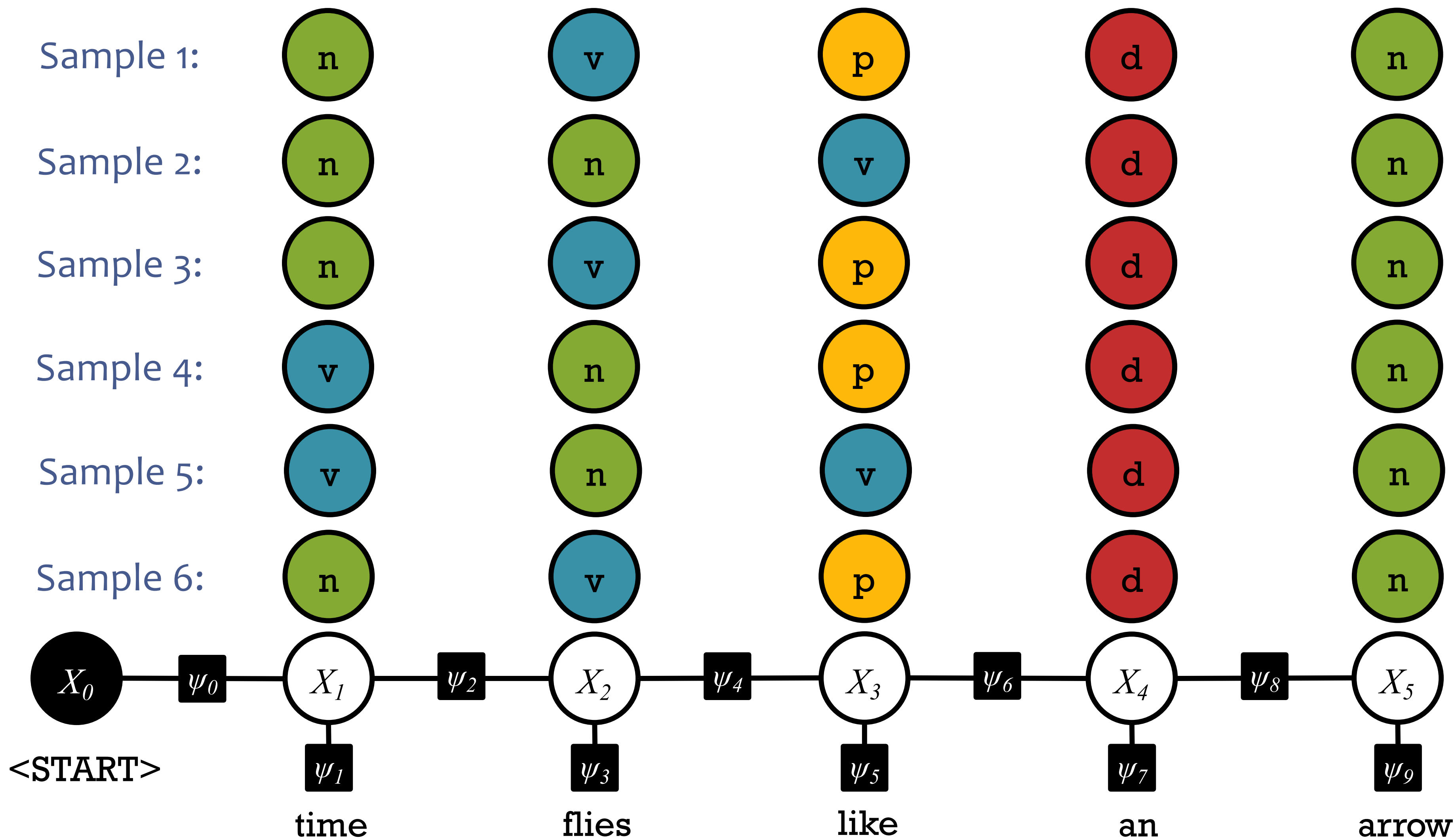
Predecessor	Central Park Task Force, Central Park Community Fund
Founded	1980
Founders	Elizabeth Barlow Rogers, William Sperry Beinecke, Ed Koch, Gordon Davis, Andrew Stein (*ex officio*)
Tax ID no.	13-3022855
Location	New York City, U.S.
Coordinates	40.76424°N 73.97169°W
Area served	Central Park
Key people	Elizabeth W. Smith (President & CEO)
Website	centralparknyc.org

A Visual Language for Variables and Interactions

GRAPHICAL MODELS

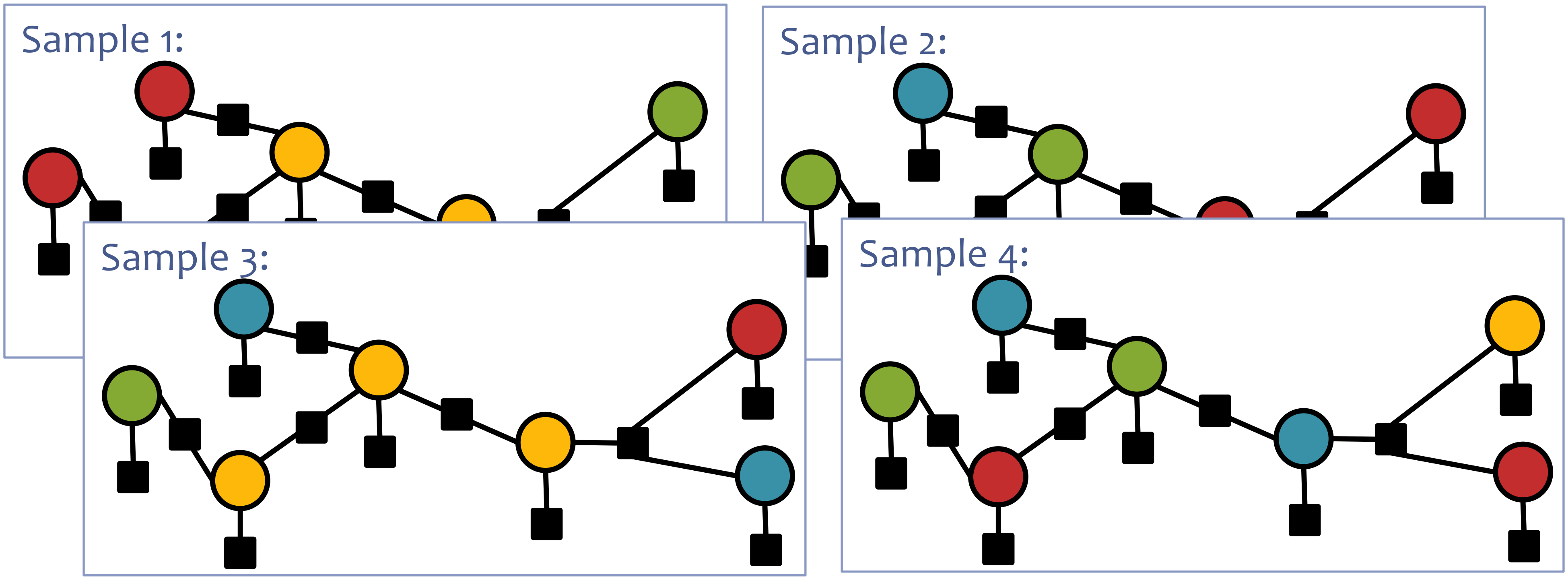
Sampling from a Joint Distribution

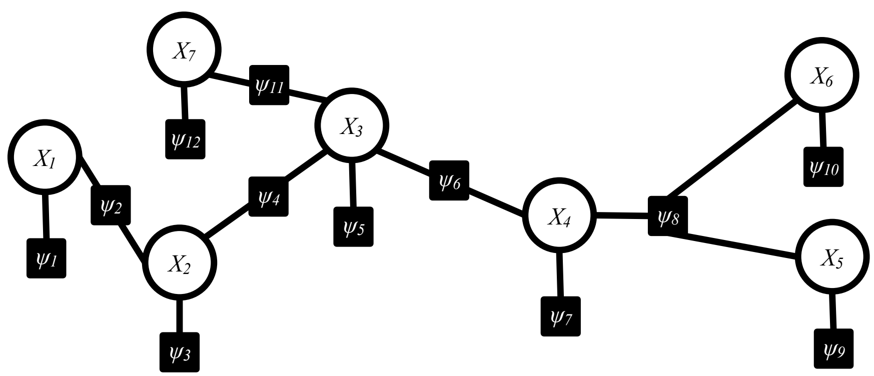
A **joint distribution** defines a probability $p(\boldsymbol{x})$ for each assignment of values $\boldsymbol{x}$ to variables $\boldsymbol{X}$. This gives the **proportion** of samples that will equal $\boldsymbol{x}$.



Sampling from a Joint Distribution

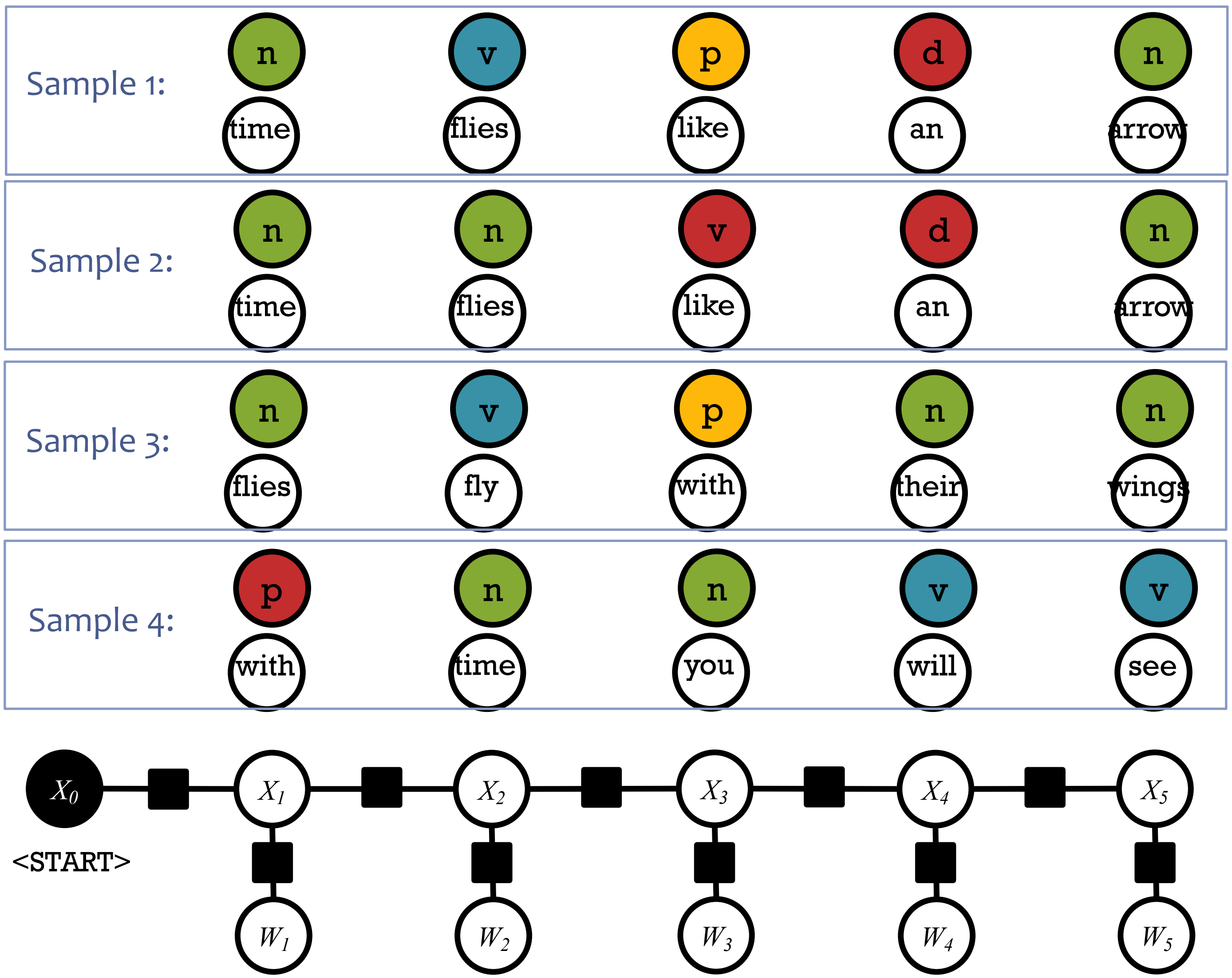
A **joint distribution** defines a probability $p(\boldsymbol{x})$ for each assignment of values $\boldsymbol{x}$ to variables $\boldsymbol{X}$.
This gives the **proportion** of samples that will equal $\boldsymbol{x}$.





Sampling from a Joint Distribution

A **joint distribution** defines a probability $p(\boldsymbol{x})$ for each assignment of values $\boldsymbol{x}$ to variables $\boldsymbol{X}$. This gives the **proportion** of samples that will equal $\boldsymbol{x}$.



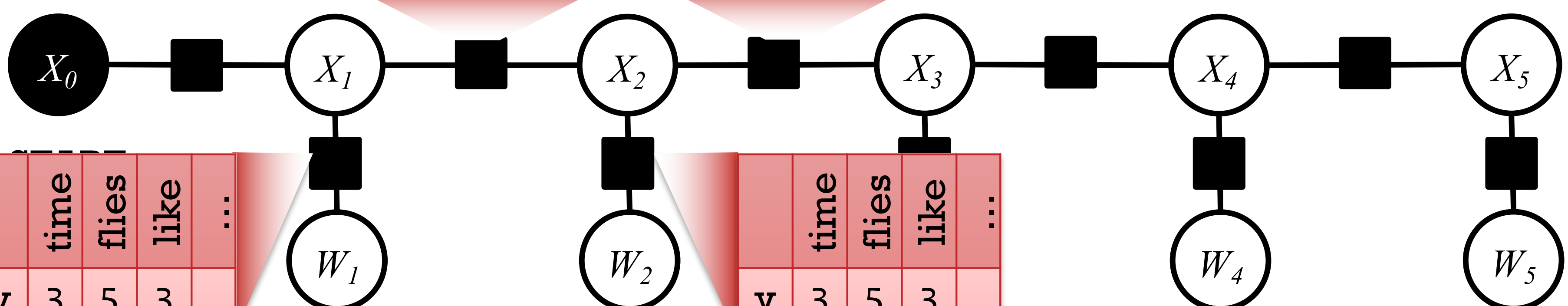
Factors have local opinions (≥ 0)

Each black box looks at *some* of the tags X_i and words W_i

	v	n	p	d
v	1	6	3	4
n	8	4	2	0.1
p	1	3	1	3
d	0.1	8	0	0

	v	n	p	d
v	1	6	3	4
n	8	4	2	0.1
p	1	3	1	3
d	0.1	8	0	0

Note: We chose to reuse the same factors at different positions in the sentence.



	time	flies	like	…
v	3	5	3	
n	4	5	2	
p	0.1	0.1	3	
d	0.1	0.2	0.1	

	time	flies	like	…
v	3	5	3	
n	4	5	2	
p	0.1	0.1	3	
d	0.1	0.2	0.1	

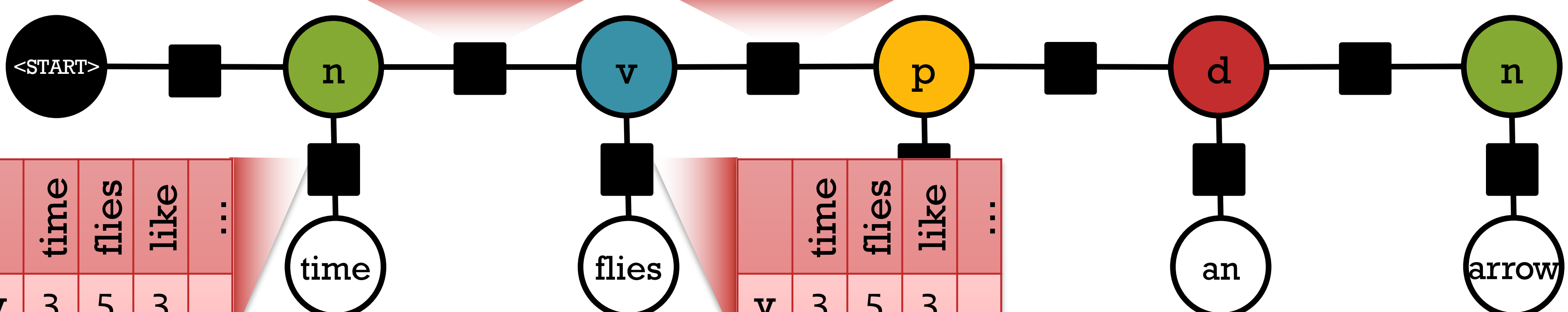
Factors have local opinions (≥ 0)

Each black box looks at *some* of the tags X_i and words W_i

$$p(\text{n, v, p, d, n, time, flies, like, an, arrow}) \quad = \quad ?$$

	v	n	p	d
v	1	6	3	4
n	8	4	2	0.1
p	1	3	1	3
d	0.1	8	0	0

	v	n	p	d
v	1	6	3	4
n	8	4	2	0.1
p	1	3	1	3
d	0.1	8	0	0



	time	flies	like	…
v	3	5	3	
n	4	5	2	
p	0.1	0.1	3	
d	0.1	0.2	0.1	

	time	flies	like	…
v	3	5	3	
n	4	5	2	
p	0.1	0.1	3	
d	0.1	0.2	0.1	

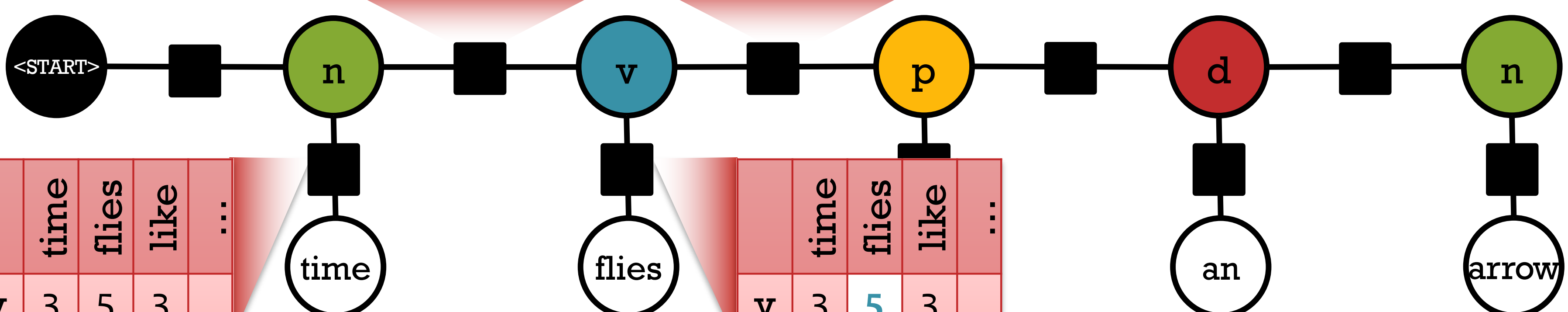
Global probability = product of local opinions

Each black box looks at *some* of the tags X_i and words W_i

$$p(\text{n, v, p, d, n, time, flies, like, an, arrow}) = \frac{1}{Z}(4 * 8 * 5 * 3 * \ldots)$$

Uh-oh! The probabilities of the various assignments sum up to Z > 1.

So divide them all by Z.



	v	n	p	d
v	1	6	3	4
n	8	4	2	0.1
p	1	3	1	3
d	0.1	8	0	0

	v	n	p	d
v	1	6	3	4
n	8	4	2	0.1
p	1	3	1	3
d	0.1	8	0	0

	time	flies	like	…
v	3	5	3	
n	4	5	2	
p	0.1	0.1	3	
d	0.1	0.2	0.1	

	time	flies	like	…
v	3	5	3	
n	4	5	2	
p	0.1	0.1	3	
d	0.1	0.2	0.1	

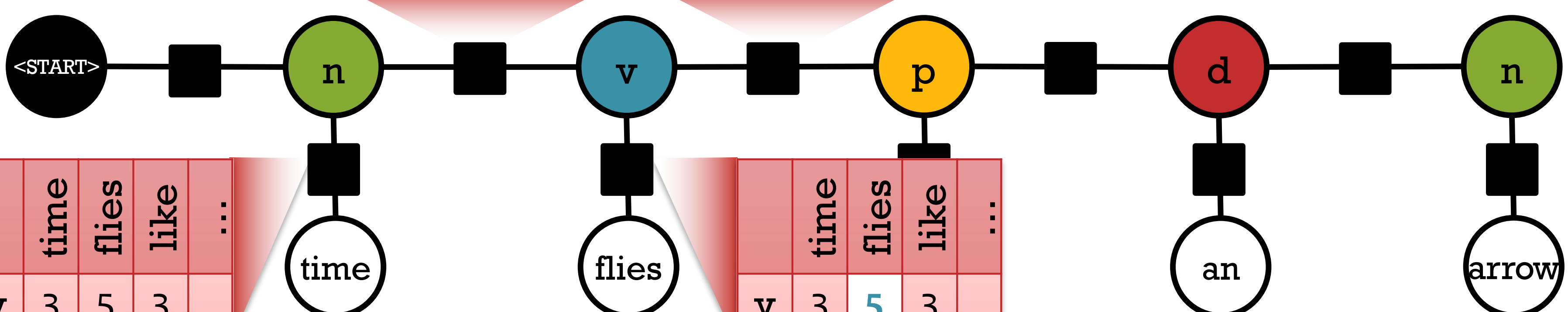
Markov Random Field (MRF)

Joint distribution over tags X_i and words W_i
The individual factors aren't *necessarily* probabilities.

$$p(\text{n, v, p, d, n, time, flies, like, an, arrow}) = \frac{1}{Z}(4 * 8 * 5 * 3 * \ldots)$$

	v	n	p	d
v	1	6	3	4
n	8	4	2	0.1
p	1	3	1	3
d	0.1	8	0	0

	v	n	p	d
v	1	6	3	4
n	8	4	2	0.1
p	1	3	1	3
d	0.1	8	0	0



	time	flies	like	…
v	3	5	3	
n	4	5	2	
p	0.1	0.1	3	
d	0.1	0.2	0.1	

	time	flies	like	…
v	3	5	3	
n	4	5	2	
p	0.1	0.1	3	
d	0.1	0.2	0.1	

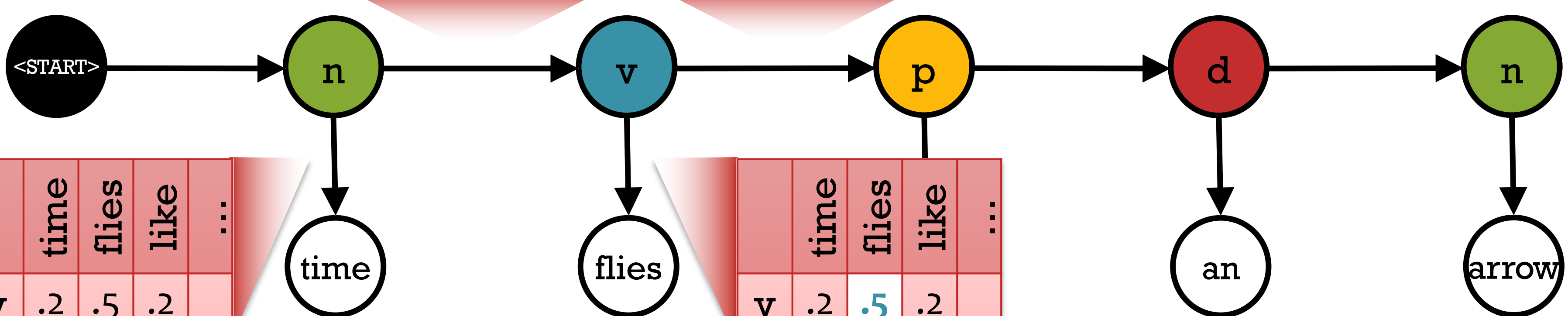
Hidden Markov Model

But sometimes we *choose* to make them probabilities.
Constrain each row of a factor to sum to one. Now $Z = 1$.

$$p(\text{n, v, p, d, n, time, flies, like, an, arrow}) = \cancel{\frac{1}{Z}}(.3 * .8 * .2 * .5 * \ldots)$$

	v	n	p	d
v	.1	.4	.2	.3
n	.8	.1	.1	0
p	.2	.3	.2	.3
d	.2	.8	0	0

	v	n	p	d
v	.1	.4	.2	.3
n	.8	.1	.1	0
p	.2	.3	.2	.3
d	.2	.8	0	0



	time	flies	like	…
v	.2	.5	.2	
n	.3	.4	.2	
p	.1	.1	.3	
d	.1	.2	.1	

	time	flies	like	…
v	.2	.5	.2	
n	.3	.4	.2	
p	.1	.1	.3	
d	.1	.2	.1	

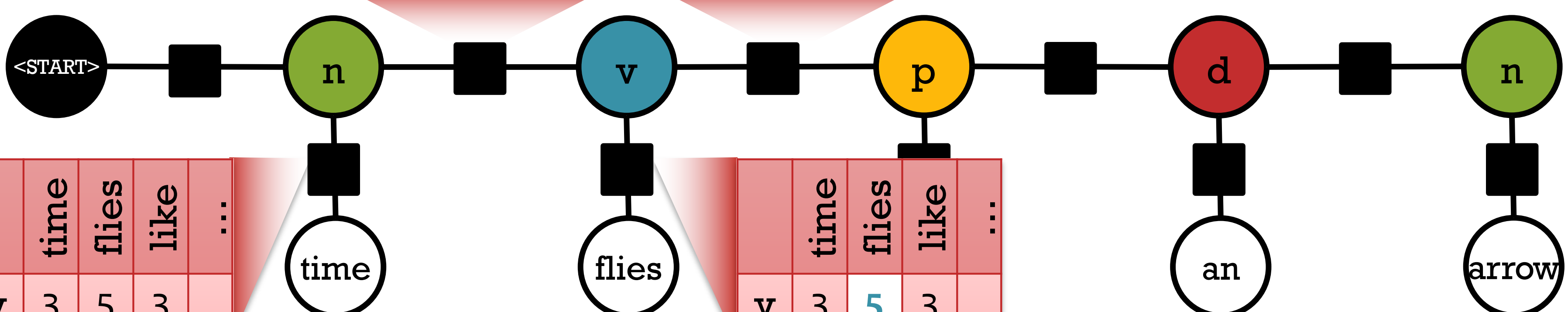
Markov Random Field (MRF)

Joint distribution over tags X_i and words W_i

$$p(\text{n, v, p, d, n, time, flies, like, an, arrow}) = \frac{1}{Z}(4 * 8 * 5 * 3 * \ldots)$$

	v	n	p	d
v	1	6	3	4
n	8	4	2	0.1
p	1	3	1	3
d	0.1	8	0	0

	v	n	p	d
v	1	6	3	4
n	8	4	2	0.1
p	1	3	1	3
d	0.1	8	0	0



	time	flies	like	…
v	3	5	3	
n	4	5	2	
p	0.1	0.1	3	
d	0.1	0.2	0.1	

	time	flies	like	…
v	3	5	3	
n	4	5	2	
p	0.1	0.1	3	
d	0.1	0.2	0.1	

Conditional Random Field (CRF)

Conditional distribution over tags X_i given words w_i.
The factors and Z are now specific to the sentence w.

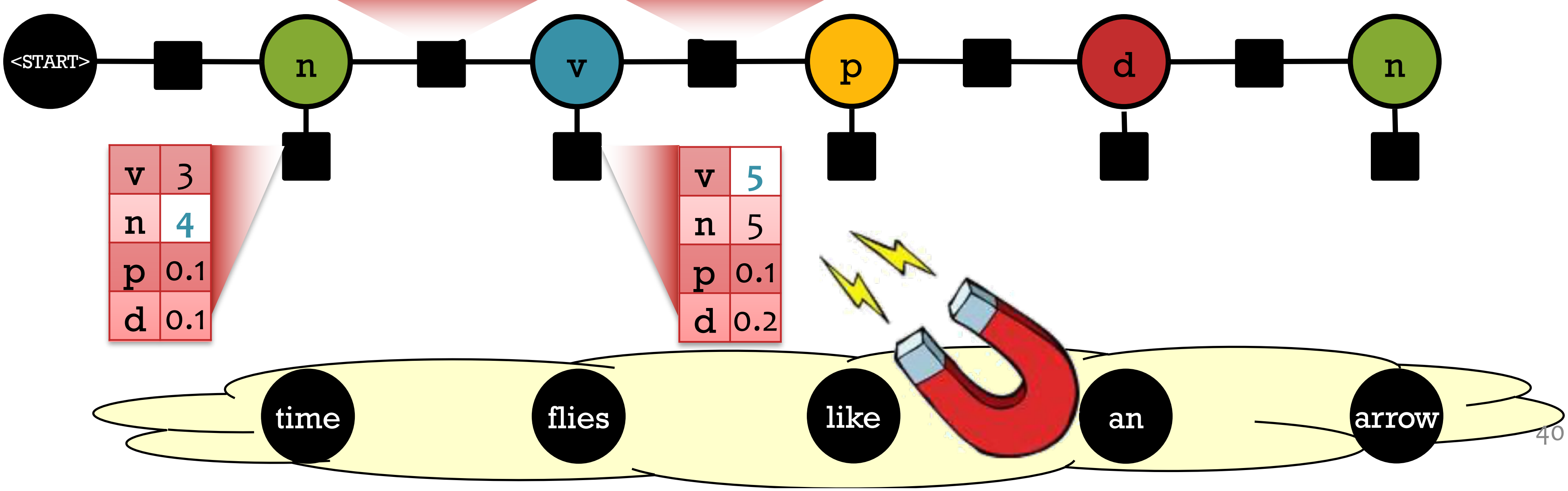
$$p(\text{n, v, p, d, n} \mid \text{time, flies, like, an, arrow}) = \frac{1}{Z}(4 * 8 * 5 * 3 * \ldots)$$

	v	n	p	d
v	1	6	3	4
n	8	4	2	0.1
p	1	3	1	3
d	0.1	8	0	0

	v	n	p	d
v	1	6	3	4
n	8	4	2	0.1
p	1	3	1	3
d	0.1	8	0	0

v	3
n	4
p	0.1
d	0.1

v	5
n	5
p	0.1
d	0.2



Exercise: Wikipedia Infoboxes

Question:

Suppose you want to populate missing infobox fields.

1. What are the variables?
2. What are the interactions?

Answer:

Central Park Conservancy

From Wikipedia, the free encyclopedia

Coordinates: 40.76424°N 73.97169°W

The **Central Park Conservancy** is a private, nonprofit park conservancy that manages Central Park under a contract with the City of New York and NYC Parks. The conservancy employs most maintenance and operations staff in the park. It effectively oversees the work of both the private and public employees under the authority of the publicly appointed Central Park administrator, who reports to the parks commissioner and the conservancy's president.[1]

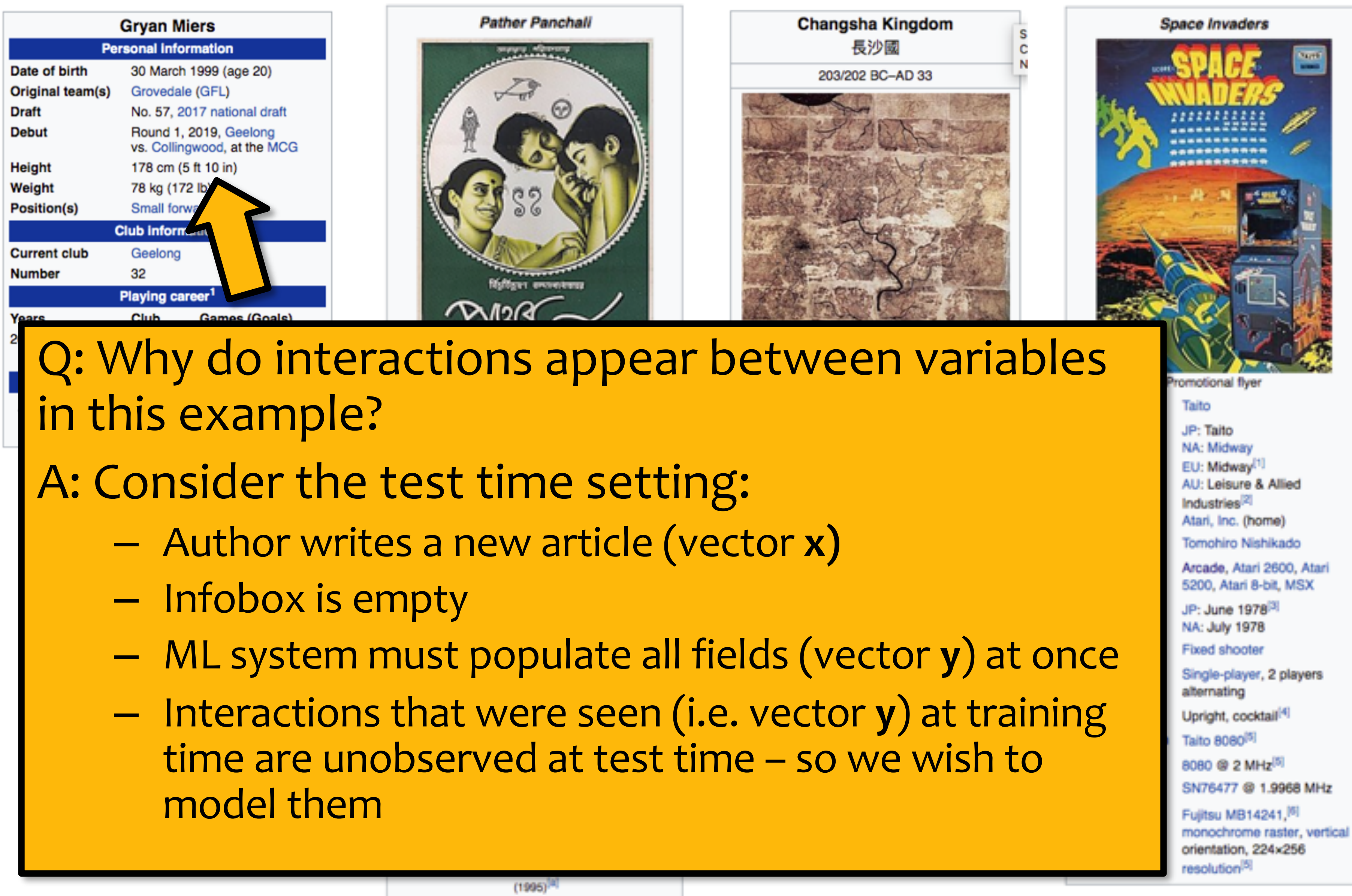
The Central Park Conservancy was founded in 1980 in the aftermath of Central Park's decline in the 1960s and 1970s.[2] Initially devoted to fundraising for projects to restore and improve the park, it took over the park's management duties in 1998.[3] The organization has invested more than $800 million toward the restoration and enhancement of Central Park since its founding.[4] With an endowment of over $200 million, consisting of contributions from residents, corporations, and foundations,[5] the Conservancy provides 75 percent of the Park's $65 million annual operating budget and is responsible for all basic care of the park.[6] The Conservancy also provides maintenance support and staff training programs for other public parks in New York City, and has assisted with the development of new parks, such as the High Line and Brooklyn Bridge Park.[7]:45–46

Central Park Conservancy

Predecessor	Central Park Task Force, Central Park Community Fund
Founded	1980
Founders	Elizabeth Barlow Rogers, William Sperry Beinecke, Ed Koch, Gordon Davis, Andrew Stein (*ex officio*)
Tax ID no.	13-3022855
Location	New York City, U.S.
Coordinates	40.76424°N 73.97169°W
Area served	Central Park
Key people	Elizabeth W. Smith (President & CEO)
Website	centralparknyc.org

Exercise: Wikipedia Infoboxes

Populating Wikipedia Infoboxes



Q: Why do interactions appear between variables in this example?

A: Consider the test time setting:

- Author writes a new article (vector **x**)
- Infobox is empty
- ML system must populate all fields (vector **y**) at once
- Interactions that were seen (i.e. vector **y**) at training time are unobserved at test time – so we wish to model them

INFERENCE PROBLEMS

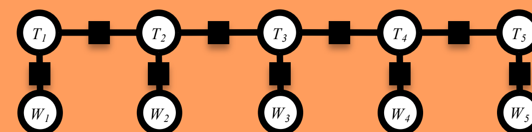
Inference

1. Data

$$\mathcal{D} = \{\boldsymbol{x}^{(n)}\}_{n=1}^{N}$$

2. Model

$$p(\boldsymbol{x} \mid \boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \prod_{C \in \mathcal{C}} \psi_C(\boldsymbol{x}_C)$$



3. Objective

$$\ell(\theta; \mathcal{D}) = \sum_{n=1}^{N} \log p(\boldsymbol{x}^{(n)} \mid \boldsymbol{\theta})$$

5. Inference

1. Marginal Inference

$$p(\boldsymbol{x}_C) = \sum_{\boldsymbol{x}' : \boldsymbol{x}'_C = \boldsymbol{x}_C} p(\boldsymbol{x}' \mid \boldsymbol{\theta})$$

2. Partition Function

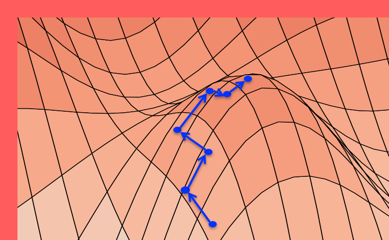
$$Z(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}} \prod_{C \in \mathcal{C}} \psi_C(\boldsymbol{x}_C)$$

3. MAP Inference

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmax}}\ p(\boldsymbol{x} \mid \boldsymbol{\theta})$$

4. Learning

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \ell(\boldsymbol{\theta}; \mathcal{D})$$



5. Inference

1. Marginal Inference (#P-Hard)

Compute marginals of variables and cliques

$$p(x_i) = \sum_{\boldsymbol{x}' : x'_i = x_i} p(\boldsymbol{x}' \mid \boldsymbol{\theta}) \qquad p(\boldsymbol{x}_C) = \sum_{\boldsymbol{x}' : \boldsymbol{x}'_C = \boldsymbol{x}_C} p(\boldsymbol{x}' \mid \boldsymbol{\theta})$$

2. Partition Function (#P-Hard)

Compute the normalization constant

$$Z(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}} \prod_{C \in \mathcal{C}} \psi_C(\boldsymbol{x}_C)$$

3. MAP Inference (NP-Hard)

Compute variable assignment with highest probability

$$\hat{\boldsymbol{x}} = \operatorname*{argmax}_{\boldsymbol{x}} \; p(\boldsymbol{x} \mid \boldsymbol{\theta})$$

4. Sampling (cf. convergence, variance)

Draw a sample variable assignment

$$\mathbf{x} \sim p(\cdot | \boldsymbol{\theta})$$

Q&A

Q: But in **deep learning** we don’t need to solve these inference problems, right?

A: Wrong…it’s not that we don’t *need* to solve them, it’s that we can’t!

Questions you *could* ask your RNN-LM or seq2seq model:

1. What is the probability of the 7^{th} token being ‘zebra’ (marginal inference)
2. For an unnormalized model, what is the normalization constant? (partition function)
3. What is the most probable output sequence? (MAP inference)
4. Give me 10 samples from the distribution.

SYLLABUS HIGHLIGHTS

Syllabus Highlights

The syllabus is located on the course webpage:

http://708.mlcourse.org → ... cs.cmu.edu...

The **course policies** are **required** reading.

Syllabus Highlights

- **Grading**: 45% homework, 35% project, 15% quizzes, 5% participation
- **Quizzes**: ~3 quizzes, during lecture/recitation time
- **Homework**: ~5 assignments
  - 6 grace days for homework assignments
  - Late submissions: 80% day 1, 60% day 2, 40% day 3, 20% day 4
  - No submissions accepted after 4 days w/o extension
  - Extension requests: see syllabus
- **Recitations**: Fridays, same time/place as lecture (optional, interactive sessions)
- **Readings**: required, online PDFs, recommended for after lecture
- **Technologies**:
  - Piazza (discussion),
  - Gradescope (homework),
  - Google Forms (polls),
  - Gather.Town (office hours),
  - Zoom (livestream),
  - Panopto (video recordings)
- **Academic Integrity**:
  - Collaboration encouraged, but must be documented
  - Solutions must always be written independently
  - No re-use of found code / past assignments
  - Severe penalties (i.e.. failure)
- **Office Hours**: posted on Google Calendar on "People" page

Lectures

- You should ask lots of questions
  - Interrupting (by raising a hand, turning on your video, and waiting to be called on) to ask your question is strongly encouraged
  - Use the chat to ask questions in real time (TAs will be monitoring the chat and will either answer or interrupt the instructor)
  - Asking questions later on Piazza is also great
- When I ask a question…
  - I want you to answer
  - Even if you don't answer, think it through as though I'm about to call on you
- Interaction improves learning (both in-class and at my office hours)

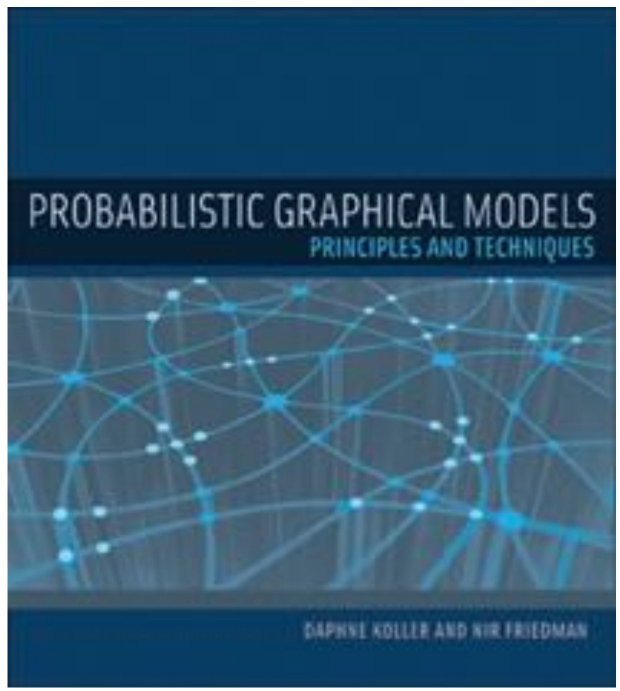
Homework

There will be 5 homework assignments during the semester. The assignments will consist of both conceptual and programming problems.

	Main Topic	Implementation	Application Area	Type
HW1	Properties of graphical models	NA	NA	written
HW2	Marginal inference and MLE	RNN + Tree CRF	natural language processing	written + programming
HW3	MAP inference and structured SVM	CNN + Ising model	computer vision	written + programming
HW4	MCMC	word embeddings + Gibbs sampler	topic modeling	written + programming
HW5	Variational Inference	variational inference	TBD	written + programming

Textbooks

You are not *required* to read a textbook, but Koller & Friedman is a thorough reference text that includes a lot of the topics we cover.

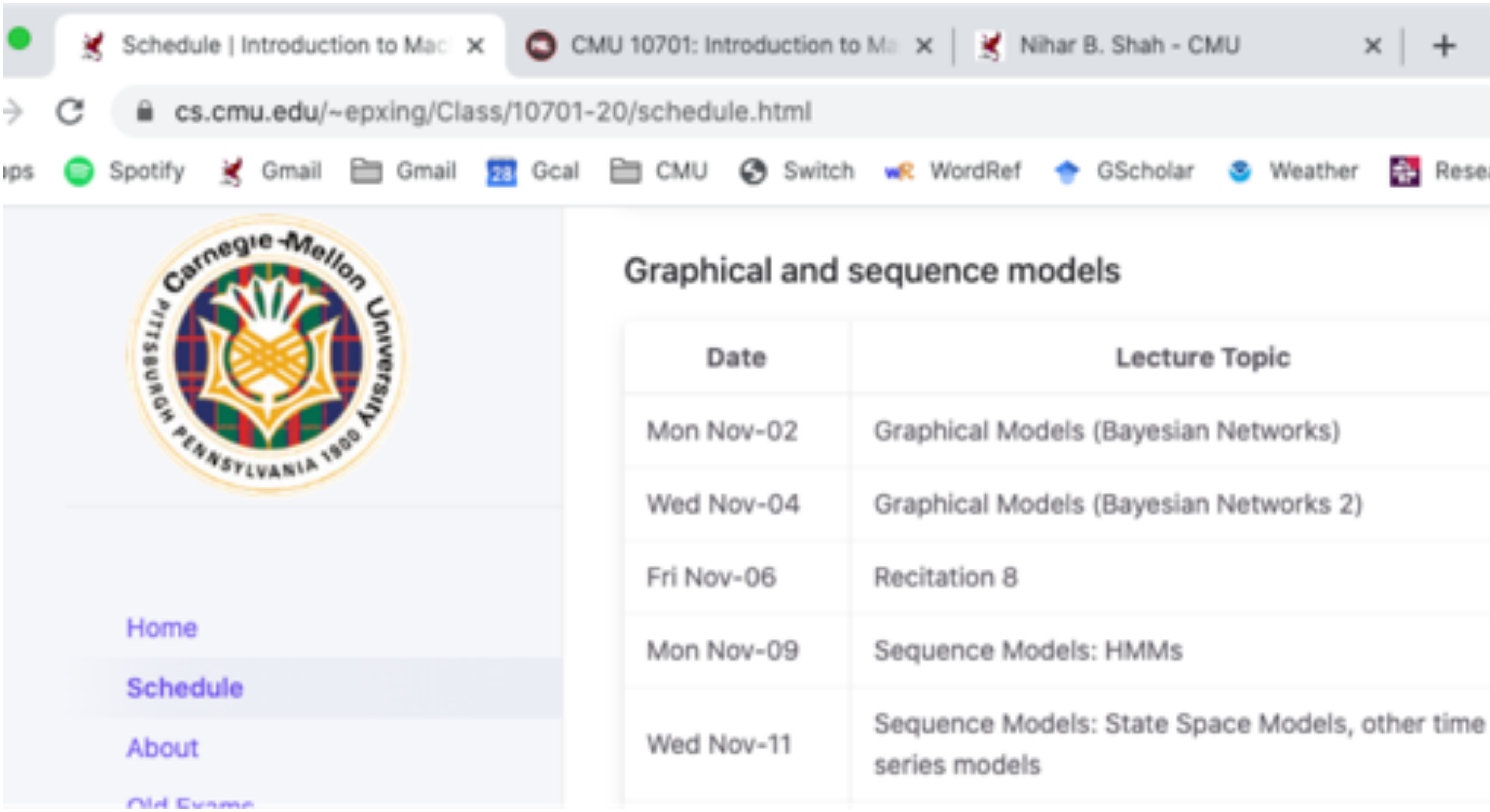


Prerequisites

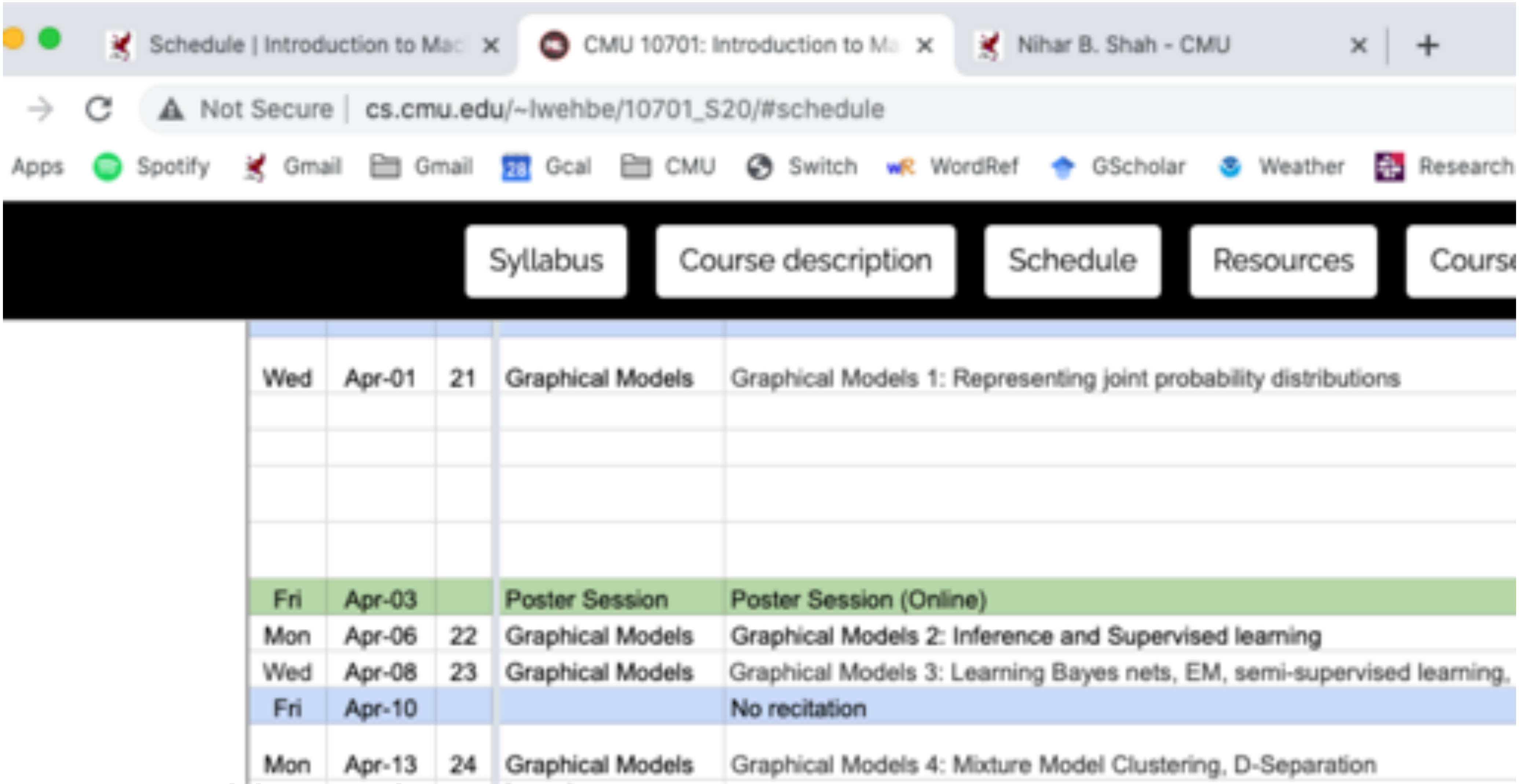
What they are:

1. Introductory machine learning. (i.e. 10-701, 10-715)
2. Significant experience programming in a general programming language.
   – Some homework may require you to use Python, so you will need to at least be **proficient in the basics of Python**.
3. College-level probability, calculus, linear algebra, and discrete mathematics.

Prerequisites



Prerequisites



Prerequisites

Schedule | Introduction to Mac × | CMU 10701: Introduction to Ma × | Nihar B. Shah - CMU ×

cs.cmu.edu/~nihars/teaching/10715-Fa20/index.html

Apps | Spotify | Gmail | Gmail | Gcal | CMU | Switch | WordRef | GScholar | Weather | Res

9	Boosting	SB C
Nov 11	Online learning	SB C
Nov 16	Semi-supervised learning, Active learning, Multi-armed bandits	Trans Multi-
Nov 18	Reinforcement learning	Surve
Nov 23	Graphical models	Grapl
Nov 25	No class (Thanksgiving break)	

Project

- Goals:
  - Explore an interesting problem in your domain of interest
  - Employ graphical models for a structured prediction task or latent variable modeling
  - For example:
    - compare models under the same inference technique
    - compare inference methods on the same model
    - compare learning methods on the same model
  - Deeper understanding of methods in real-world application
  - Work in teams of 3 – 4 students
- Milestones: (*due in 2nd half of semester*)
  1. Team Formation
  2. Proposal
  3. Midway Report
  4. Final Report
  5. Poster Presentation

Q&A