10-607 Computational Foundations for Machine Learning

Machine Learning Department
School of Computer Science
Carnegie Mellon University

Data Structures

Matt Gormley
Lecture 9
Nov. 19, 2018

Reminders

- Homework B: Complexity & Recursion
  - Out: Thu, Nov. 8
  - Due: Tue, Nov. 20 at 11:59pm
- Quiz A: Logic & Proofs; Computation
  - Mon, Nov. 26, in-class
  - Covers Lectures 1 – 6

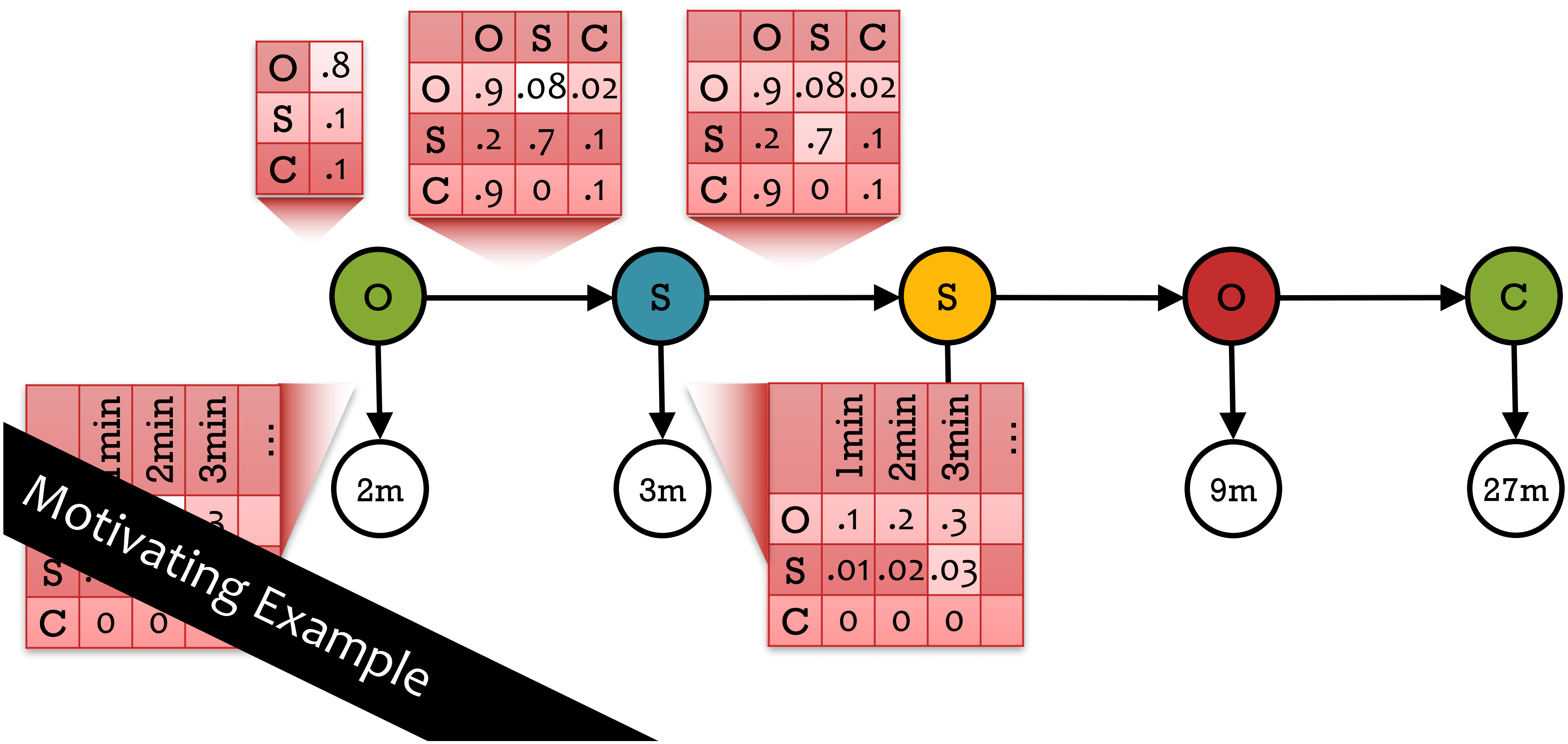
Q&A

DYNAMIC PROGRAMMING

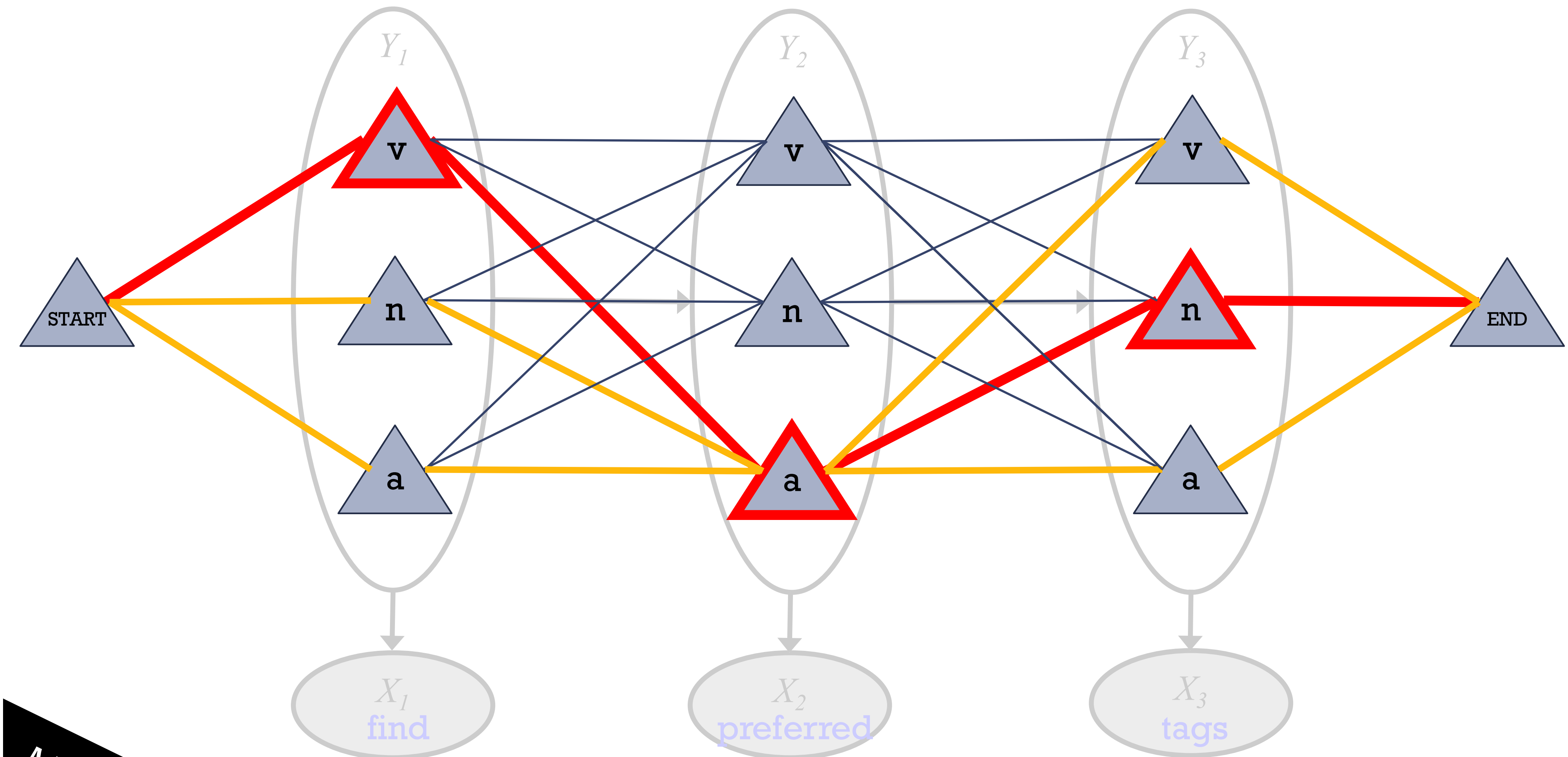
Hidden Markov Model

A Hidden Markov Model (HMM) provides a joint distribution over the the tunnel states / travel times with an assumption of dependence between adjacent tunnel states.

$$p(\text{O, S, S, O, C, 2m, 3m, 18m, 9m, 27m}) = (.8 * .08 * .2 * .7 * .03 * \ldots)$$



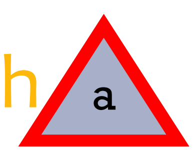
Forward-Backward Algorithm: Finds Marginals



Motivating Example

p(**v a n**) = (1/Z) * product weight of one path

- M l probability p(Y_2 = a)

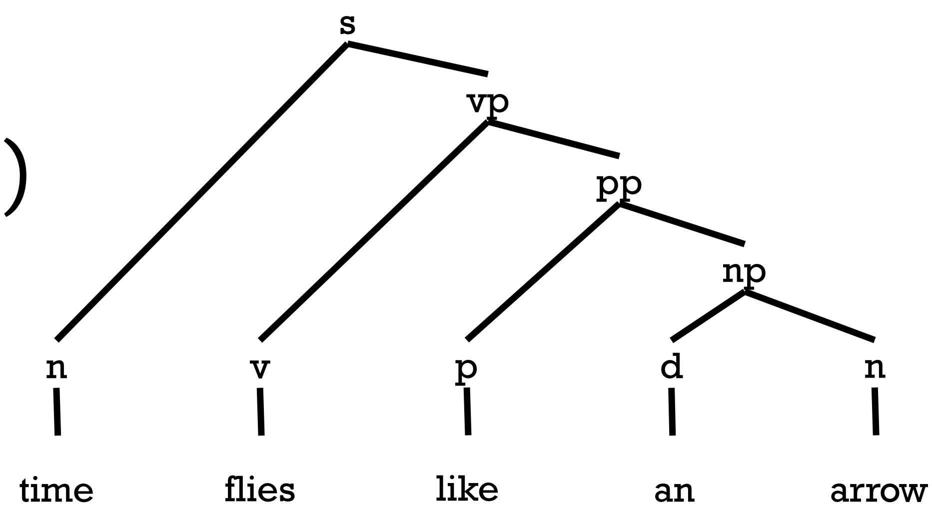
= (* total weight of *all* paths through 

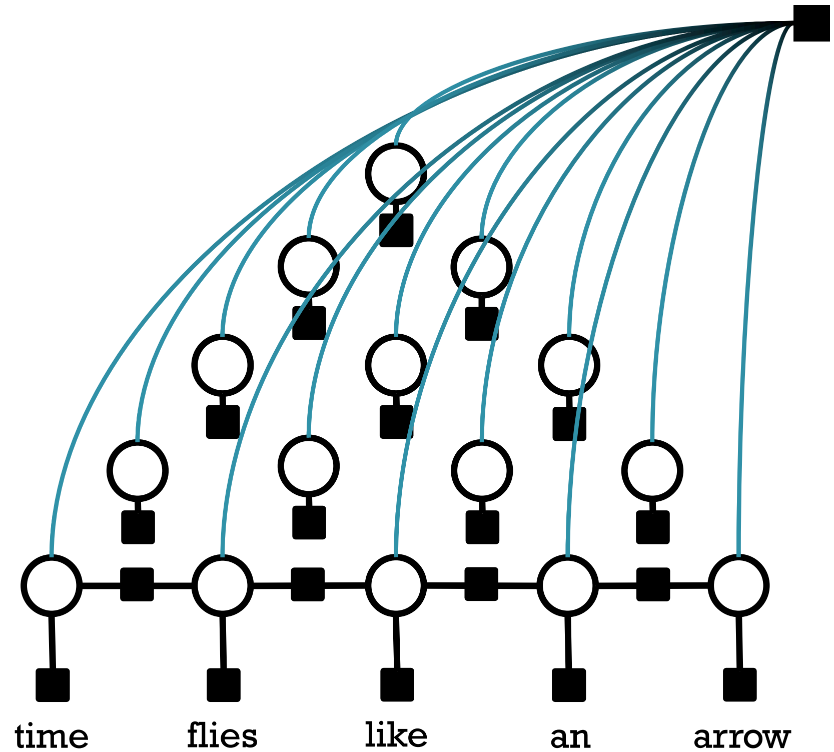
Example Task:

Constituency Parsing

- **Variables:**
  - Constituent type (or ∅) for each of $O(n^2)$ substrings
- **Interactions:**
  - Constituents must describe a binary tree
  - Tag bigrams
  - Nonterminal triples (parent, left-child, right-child)

[these factors not shown]





Motivating Example

(Naradowsky, Vi… …th, 2012)

Dynamic Programming

Key Idea: Divide a large problem into reusable subproblems and solve each subproblem, storing the result of each for later reuse

"Let's take a word that has an absolutely precise meaning, namely dynamic, in the classical physical sense. It also has a very interesting property as an adjective, and that is it's impossible to use the word, dynamic, in a pejorative sense. Try thinking of some combination that will possibly give it a pejorative meaning. [. . .] Thus, I thought **dynamic programming** was a good name. It was something not even a Congressman could object to. So I used it as an umbrella for my activities."

Richard Bellman, Autobiography (1984)

Dynamic Programming

Chalkboard:

- Big Idea: Dynamic Programming
- Example: Fibonacci with and without dynamic programming
  - Recursive Fibonacci's computational complexity
  - Dynamic programming Fibonacci's computational complexity
- Types of Dynamic Programming
  - Tabulation (bottom-up)
  - Memoization (top-down)
- Example: Matrix Product Parenthesization

DATA STRUCTURES FOR ML

Abstractions vs. Data Structures

Abstractions

- List
- Set
- Map
- Queue (FIFO)
- Stack (LIFO)
- Graph
- Priority Queue

Data Structures

- Array (fixed size)
- Array (variable size)
- Linked List
- Doubly-Linked List
- Multidimensional Array
- Tensor
- Hash Map
- Binary Search Tree
- Balanced Tree
- Trie
- Stack
- Heap
- Graph
- Bipartite Graph
- Sparse Vector
- Sparse Matrix

Data Structures for ML

Examples…

- Data:
  - Dense feature vector (array)
  - Sparse feature vector (sparse vector)
  - Design matrix (multidimensional array)
- Models:
  - Decision Trees (tree)
  - Bayesian Network (directed acyclic graph)
  - Factor Graph (bipartite graph)
- Algorithms:
  - Greedy Search (weighted graph)
  - A* Search (priority queue/heap)
  - Forward-backward for HMM (trellis)

Trees

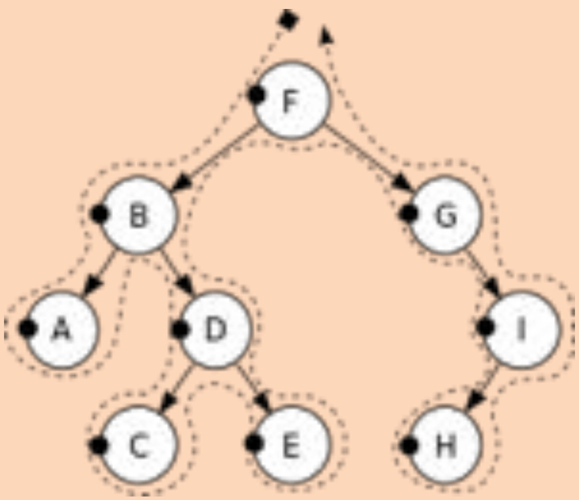
Chalkboard:

- Binary Tree
  - Representation
  - Depth First Search
    - pre-order traversal
    - in-order traversal
    - post-order traversal
  - Breadth First Search
- Decision Tree
  - Representation

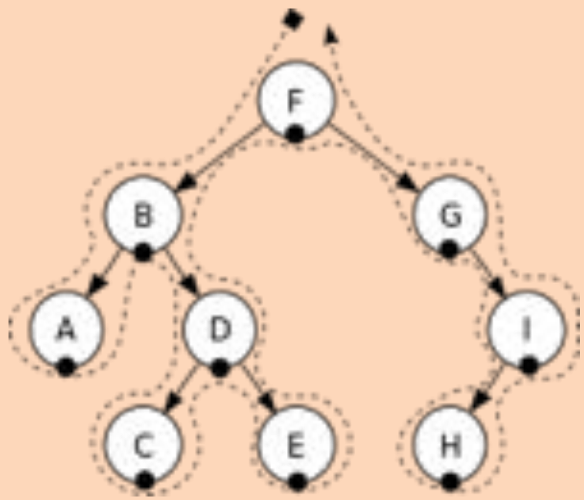
Tree Traversals

Depth First Search

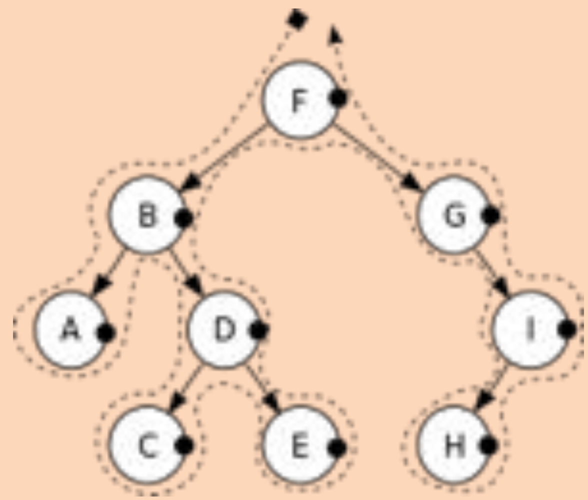
Pre-order



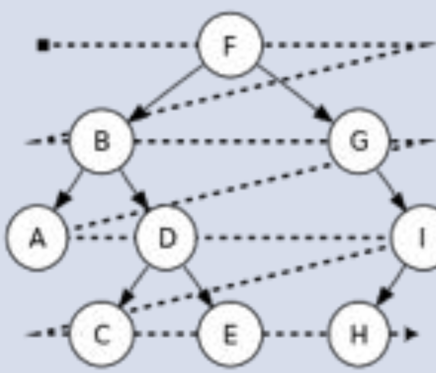
In-order



Post-order



Breadth First Search



Figures from Wikipedia

Graphs

Chalkboard:

- Undirected Graphs
  - Representation
  - Breadth First Search

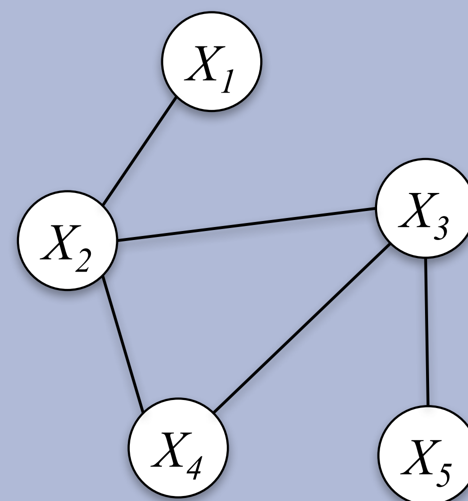
Decision Tree Learning Example

In-Class Exercise

- Suppose we now have an undirected (possibly cyclic) graph
- You need a breadth-first traversal of the graph from some query node
- Your friend suggests you use the same BFS algorithm we defined for binary trees

1. *What goes wrong?*
2. *How can you fix it?*

Example Undirected Graph:



Sparse Vectors

Chalkboard:

- Sparse Vector
  - Representation
  - Sparse Dot Product
  - Addition of dense vector and sparse vector

Data Structures & Algorithms

Chalkboard:

- Weighted Directed Acyclic Graph
  - Representation
  - Greedy Search
  - Dijkstra’s Algorithm
  - A* Search
- Binary Search Tree
  - Representation
  - Average vs. Worst Case Time Complexity
  - Search
  - Insertion
  - Deletion