CS440/ECE448
Lecture 11: Exam 1 Review

3/3/2021

Topics covered

- Lecture 2: Search
- Lecture 3: A*
- Lecture 4: Heuristics
- Lecture 5: Probability
- Lecture 6: Naïve Bayes
- Lecture 7: Classifiers

Lecture 2: Search

- Search in general:
  - State: enough info to decide if you're at the goal state
  - Node: state + information about the path taken to get here (tree search)
  - Frontier
  - Explored Set/Explored Dict
- Breadth-first search (BFS):
  - Frontier is a FIFO queue
  - Time complexity and space complexity are both $O\{b^d\}$.
  - Optimal, if each action has the same cost.
- Depth-first search (DFS):
  - Frontier is a LIFO stack
  - Time complexity is $O\{b^m\}$, but space complexity is only $O\{mb\}$.
  - Not optimal. Not even complete.

Lecture 3: A*

- Uniform Cost Search: Like BFS, but for variable-cost actions
  - Frontier is a priority queue, sorted by g(n)
  - Finds the optimal path
- Greedy Search
  - Frontier is a priority queue, sorted by h(n)
  - Not optimal. Not even complete.
- A* Search:
  - Frontier is a priority queue, sorted by f(n)=g(n)+h(n)
  - Optimal and complete, as long as h(n) is admissible

Lecture 4: Heuristics

- Consistent
  - If heuristic is consistent, A* works with an explored set
  - With an inconsistent heuristic, A* works (1) with an explored dict, or (2) with neither an explored set nor an explored dict.
- Zero = UCS
- Dominant
- Designing a heuristic by simplifying the problem
- Dominant heuristic as the max of many heuristics

Lecture 5: Probability

- Axioms of probability: non-negative, max 1, probability of union
- Events
- Random variables
- Conditional probability
- Marginal probability
- Independence
- Conditional Independence

Lecture 6: Naïve Bayes

- Class labels and observations
- Using Bayes' rule to estimate the most probable class label
- The naïve Bayes assumption: observations conditionally independent given the class label
- Maximum likelihood estimation of the model parameters
- Laplace smoothing

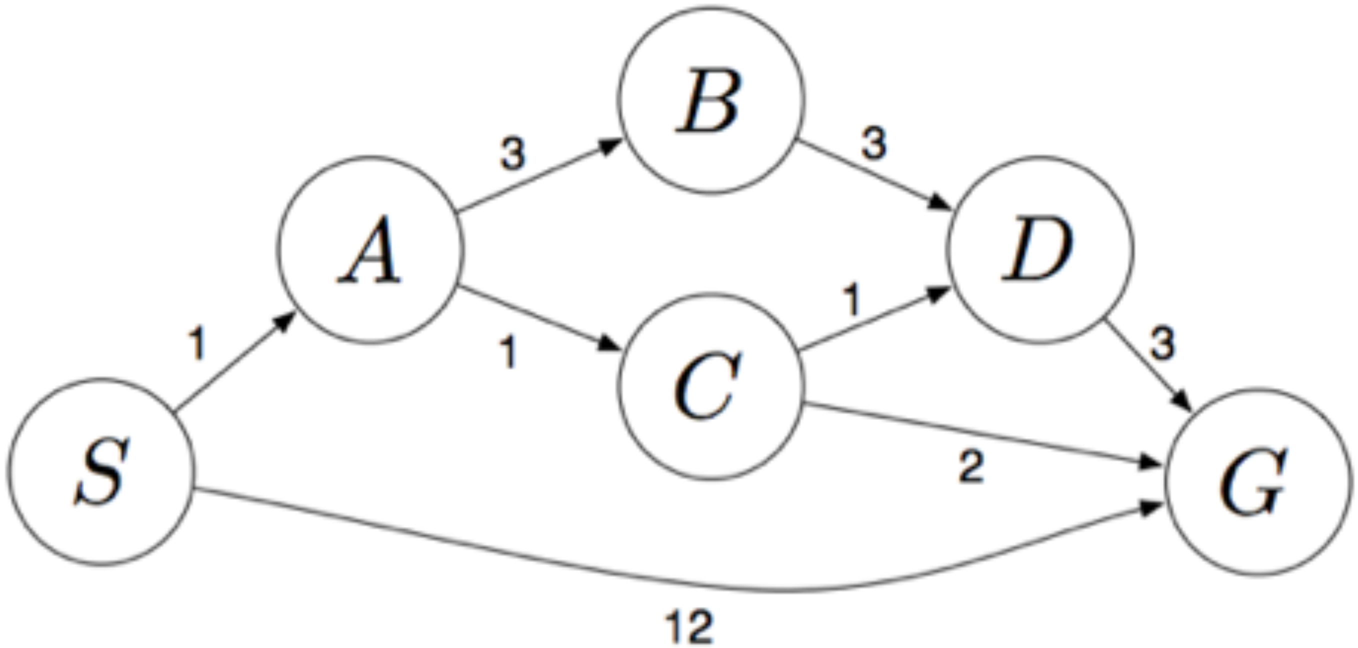
Lecture 7: Classifiers

- The Bayesian classifier: MAP = MPE
- False alarms, missed detections, and confusion matrix
- Training a classifier, choosing a classifier, evaluating a classifier
- Nearest-neighbor classifier
- Linear classifiers
- Implementation of symbolic logic using a linear classifier

Some sample problems, from the practice exam

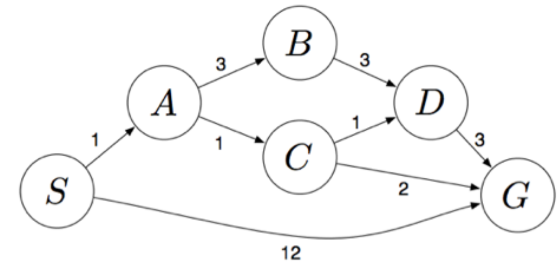
- Question 5: BFS, DFS, UCS, A*
- Question 8: Axioms of probability
- Question 6: Naïve Bayes

Question 5: BFS, DFS, UCS and A*



S denotes the start state, G denotes the goal state, step costs are written next to each arc. Assume that ties are broken alphabetically.

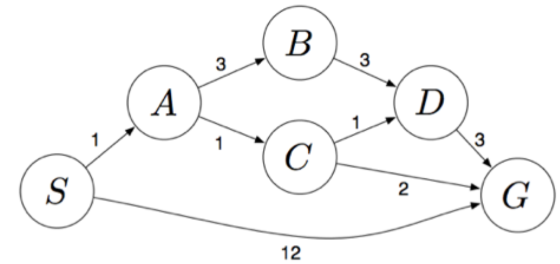
5(a): What path does BFS return?



- Frontier starts with {S}
- S is popped, A and G are inserted, so it contains {A,G}
- A is popped, B and C are inserted, so it contains {G,B,C}
- G is popped. It is the goal state.

Answer: S,G

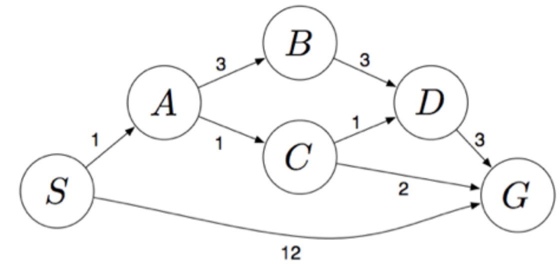
5(b): What path does DFS return?



- Frontier starts with {S}
- S is popped, A and G are inserted, so it contains {A,G}
- A is popped, B and C are inserted, so it contains {B,C,G}
- B is popped, D is inserted, so it contains {D,C,G}
- D is popped, G is inserted, so it contains {G,D,C,G}
- G is popped. It is the goal state.

Answer: S,A,B,D,G

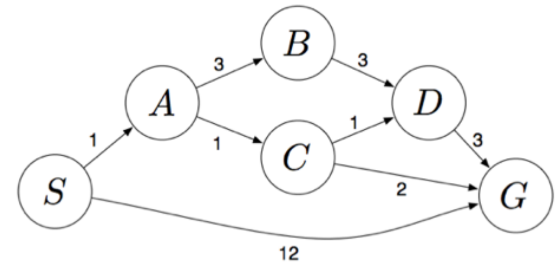
5(c): What path does UCS return?



- Frontier starts with {0:S}
- S is popped, A and G are inserted, so it contains {1:A,12:G}
- A is popped, B and C are inserted, so it contains {2:C,4:B,12:G}
- C is popped, D and G are inserted, so frontier contains {3:D,4:B,4:G,12:G}
- D is popped, G is inserted, so frontier contains {4:B,4:G,6:G,12:G}
- B is popped, and if there is no explored set, D is inserted, so frontier contains {4:G, 6:G, 7:D, 12:G}
- G is popped. It is the goal.

Answer: S,A,C,G – the optimal path

5(d): Heuristic h1



Heuristic h1 has the following values:

h1(S)=5, h1(A)=3, h1(B)=6, h1(C)=2, h1(D)=3, h1(G)=0

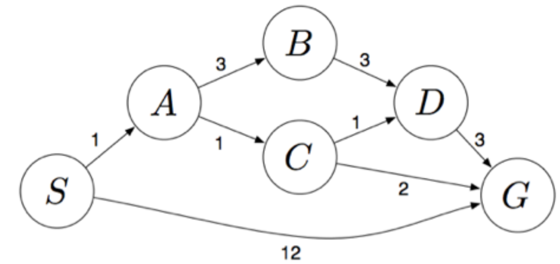
- Is it admissible?

No. h1(S) = 5, but d(S)=4.

- Is it consistent?

No. An inadmissible heuristic is never consistent.

5(d): Heuristic h2



Heuristic h2 has the following values:

h2(S)=4, h2(A)=2, h2(B)=6, h2(C)=1, h2(D)=3, h2(G)=0

- Is it admissible?

Yes. h2(n) <= d(n) for all nodes n.

- Is it consistent?

No. d(S)-d(A)=1, but h2(S)-h2(A)=2.

Question 8: Axioms of probability

Use the axioms of probability to prove that $P(\neg A) = 1 - P(A)$.

Question 8: Axioms of probability

The axioms of probability are:

1. Non-negative: $P(A) \geq 0$ for any event A, with zero probability for impossible events.
2. Max 1: If Ω is the union of all possible events, $P(\Omega) = 1$.
3. Probability of union: $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Use the axioms of probability to prove that P(¬A) = 1−P(A).

Question 8: Axioms of probability

The axioms of probability are:

1. Non-negative: $P(A) \geq 0$ for any event A, with zero probability for impossible events.
2. Max 1: If Ω is the union of all possible events, $P(\Omega) = 1$.
3. Probability of union: $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Use the axioms of probability to prove that P(¬A) = 1−P(A).

Step 1: Either A occurs, or ¬A occurs. Therefore the union $A \vee \neg A$ is the union of all possible events, therefore $\mathrm{P}(A \vee \neg A) = 1$.

Question 8: Axioms of probability

The axioms of probability are:

1. Non-negative: $P(A) \geq 0$ for any event A, with zero probability for impossible events.
2. Max 1: If Ω is the union of all possible events, $P(\Omega) = 1$.
3. Probability of union: $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Use the axioms of probability to prove that P(¬A) = 1−P(A).

Step 1: Either A occurs, or ¬A occurs. Therefore the union $A \vee \neg A$ is the union of all possible events, therefore $\mathrm{P}(A \vee \neg A) = 1$.

Step 2: $\mathrm{P}(A \vee \neg A) = P(A) + P(\neg A) - P(A \wedge \neg A)$.

Question 8: Axioms of probability

The axioms of probability are:

1. Non-negative: $P(A) \geq 0$ for any event A, with zero probability for impossible events.
2. Max 1: If Ω is the union of all possible events, $P(\Omega) = 1$.
3. Probability of union: $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Use the axioms of probability to prove that P(¬A) = 1−P(A).

Step 1: Either A occurs, or ¬A occurs. Therefore the union $A \vee \neg A$ is the union of all possible events, therefore $\mathrm{P}(A \vee \neg A) = 1$.

Step 2: $\mathrm{P}(A \vee \neg A) = P(A) + P(\neg A) - P(A \wedge \neg A)$.

Step 3: A and ¬A is impossible, so $P(A \wedge \neg A) = 0$, therefore

$$\mathrm{P}(A \vee \neg A) = P(A) + P(\neg A)$$

Question 8: Axioms of probability

The axioms of probability are:

1. Non-negative: $P(A) \geq 0$ for any event A, with zero probability for impossible events.
2. Max 1: If Ω is the union of all possible events, $P(\Omega) = 1$.
3. Probability of union: $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Use the axioms of probability to prove that P(¬A) = 1−P(A).

Step 1: Either A occurs, or ¬A occurs. Therefore the union $A \vee \neg A$ is the union of all possible events, therefore $\mathrm{P}(A \vee \neg A) = 1$.

Step 2: $\mathrm{P}(A \vee \neg A) = P(A) + P(\neg A) - P(A \wedge \neg A)$.

Step 3: A and ¬A is impossible, so $P(A \wedge \neg A) = 0$.

Step 4: $1 = P(A) + P(\neg A)$, i.e., P(¬A) = 1−P(A).

Question 6: Naïve Bayes

You're creating a sentiment classifier. Let Y=1 for positive sentiment, Y=0 for negative sentiment. You have a training corpus with four movie reviews:

Index	Sentiment	Review
1	1	what a great movie
2	1	I love this film
3	0	what a horrible movie
4	0	I hate this film

Question 6(a-b)

Index	Sentiment	Review
1	1	what a great movie
2	1	I love this film
3	0	what a horrible movie
4	0	I hate this film

(a) What's the maximum likelihood estimate of P(Y=1)?

Solution: 2/4

(b) What are maximum likelihood estimates of P(W|Y=0) and P(W|Y=1)?

Solution: each part of the corpus has 8 words, so ML estimates are:

	P(W|Y)									
Y	what	a	movie	I	this	film	great	love	horrible	hate
1	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8	0	0
0	1/8	1/8	1/8	1/8	1/8	1/8	0	0	1/8	1/8

Question 6(c)

Index	Sentiment	Review
1	1	what a great movie
2	1	I love this film
3	0	what a horrible movie
4	0	I hate this film

Use Laplace smoothing, with k=1.

Solution: add 10 to each denominator, and 1 to each numerator:

	P(W|Y)									
Y	what	a	movie	I	this	film	great	love	horrible	hate
1	2/18	2/18	2/18	2/18	2/18	2/18	2/18	2/18	1/18	1/18
0	2/18	2/18	2/18	2/18	2/18	2/18	1/18	1/18	2/18	2/18

Question 6(d)

Using methods unknown to your, your professor has come up with the following estimates:

	P(W|Y			
Y	great	love	horrible	hate
1	0.01	0.01	0.005	0.005
0	0.005	0.005	0.01	0.01

…and P(Y=1)=0.5. All other words are "out of vocabulary;" you can treat them as if they had P(W|Y=0)=P(W|Y=1)=1. Under these assumptions, what is the probability that the following review is a positive review:

I'm horrible fond of this movie, and I hate anyone who insults it.

Question 6(d) Solution

	P(W|Y			
Y	great	love	horrible	hate
1	0.01	0.01	0.005	0.005
0	0.005	0.005	0.01	0.01

I'm horrible fond of this movie, and I hate anyone who insults it.

Solution:

The only words not "out of vocabulary" are "horrible" and "hate." We have

P(Y=0,horrible,hate)=P(Y=0)P(horrible|Y=0)P(hate|Y=0) = 0.5(0.01)(0.01)

P(Y=1,horrible,hate)=P(Y=1)P(horrible|Y=1)P(hate|Y=1) = 0.5(0.005)(0.005)

Using Bayes' rule:

$$P(Y = 1|\text{horrible}, \text{hate}) = \frac{0.5(0.005)(0.005)}{0.5(0.005)(0.005) + 0.5(0.01)(0.01)}$$

Topics covered

- Lecture 2: Search
- Lecture 3: A*
- Lecture 4: Heuristics
- Lecture 5: Probability
- Lecture 6: Naïve Bayes
- Lecture 7: Classifiers