Q1-1: Can you successfully respond to BBCollaborate poll questions?

1. Yes
2. No

Q1-2: Which module topic from the course website https://happyharrycn.github.io/CS540-Fall20/schedule/ are you most excited about?

1. Search
2. Mathematical Foundation of AI
3. Game Theory
4. Machine Learning Part I
5. Machine Learning Part II
6. Applications and Ethics of AI

Q1-3: How can you ask questions during the synchronous lecture?

1. Email the instructor
2. Use BBCollaborate chat
3. Post on the Piazza live Q&A
4. Unmute yourself and ask verbally on BBCollaborate

Q1-4: Has the GPT-3 model achieved artificial general intelligence?

1. Yes
2. No
3. Not sure

Q1-5: What do you have permission to do with recorded lectures and discussions?

1. Share them with friends, as long as it is private
2. Post them to YouTube
3. Keep a local copy on your computer for your own viewing

Q1-6: Where can you go for help with CS 540 questions?

1. Attend instructor, TA, and peer mentor office hours
2. Post on Piazza
3. Post on Canvas
4. Email the instructors, TAs, or peer mentors
5. Only 1 and 2

Q2-1: Which of these is a valid Python function?

1.
```
def my_fun_1():
  print("inside function 1")
```

2.
```
def void my_fun_2():
  print("inside function 2")
```

3.
```
def my_fun_3()
  print("inside function 3")
```

4.
```
def my_fun_4():
print("inside function 4")
```

Q2-2: Which of these will print 0 1 2 3 (one number per line)

1. ```
   for i in range(3):
     print(i)
   ```

2. ```
   for i in range(4):
     print(i)
   ```

3. ```
   for i in range(4):
     print('i')
   ```

4. ```
   for i in range(3)
     print(i)
   ```

Q2-3: In Python, classes are…

1. Required before writing any function
2. No allowed
3. Optional

Q2-4: After running the following code, what is stored in `list2`?

```
list1 = [5, 3, 1]
list2 = [i - 2 for i in list1]
```

1. `[-1, 1, 3]`
2. `[2, 2]`
3. `[3, 1, -1]`

Q3-1: What is the minimum we need to define to run an uninformed search algorithm?

1. Initial states, goal states, state space, successor function
2. Initial states, goal states, state space, successor function, edge costs
3. Initial states, goal states, state space, successor function, edge costs, scores for non-goal states

Q3-2: How many minimum cost solutions are there to the farmer crossing the river problem?

1. 1
2. 2
3. 3
4. 4

Q3-3: Does the successor function always need to return a non-empty set of states reachable from s?

1. Yes
2. No

Q3-4: In the complete configuration of the 8-queens problem how many successor states does the initial state have?

1. 1
2. 8
3. 16
4. 32
5. 64

Q3-5: In the column-by-column configuration of the 8-queens problem how many successor states does the initial state have?

1. 1
2. 8
3. 16
4. 32
5. 64

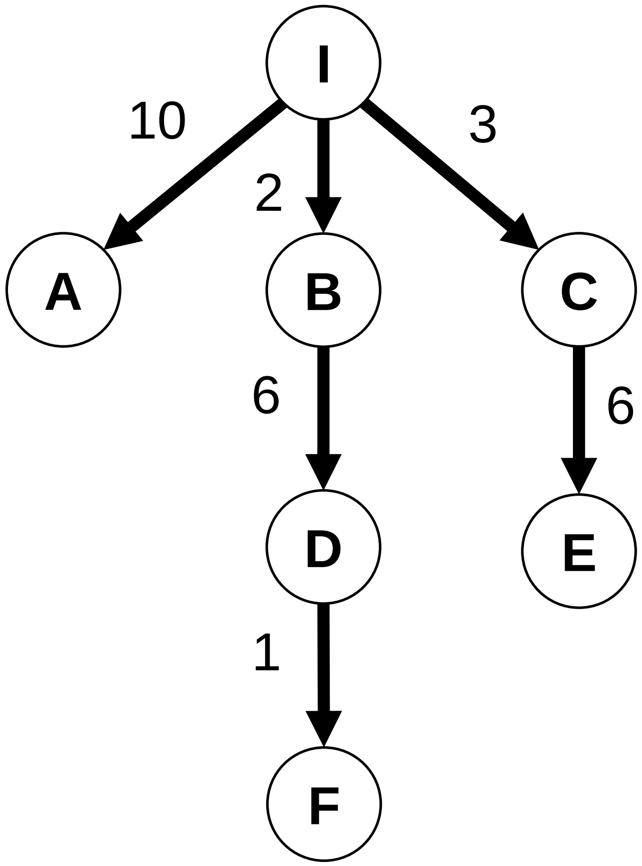
Q3-6: What is the fringe during a search?

1. The states we know nothing about
2. The states we have visited
3. The states that have been returned by the successor function that have not been visited

Q1-1: You are running BFS on a finite tree-structured state space graph that does not have a goal state. What is the behavior of BFS?

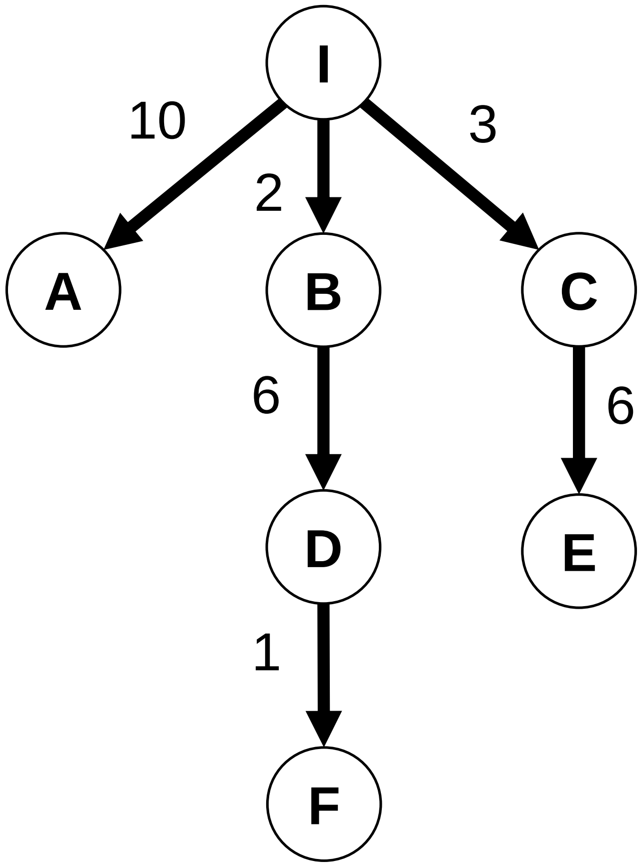
1. Visit all N nodes, then return one at random
2. Visit all N nodes, then return “failure”
3. Visit all N nodes, then return the node farthest from the initial state
4. Get stuck in an infinite loop

Q1-2: You are running UCS in the state space graph below. You just called the successor function on node **D**. What is the cost of node **F**?



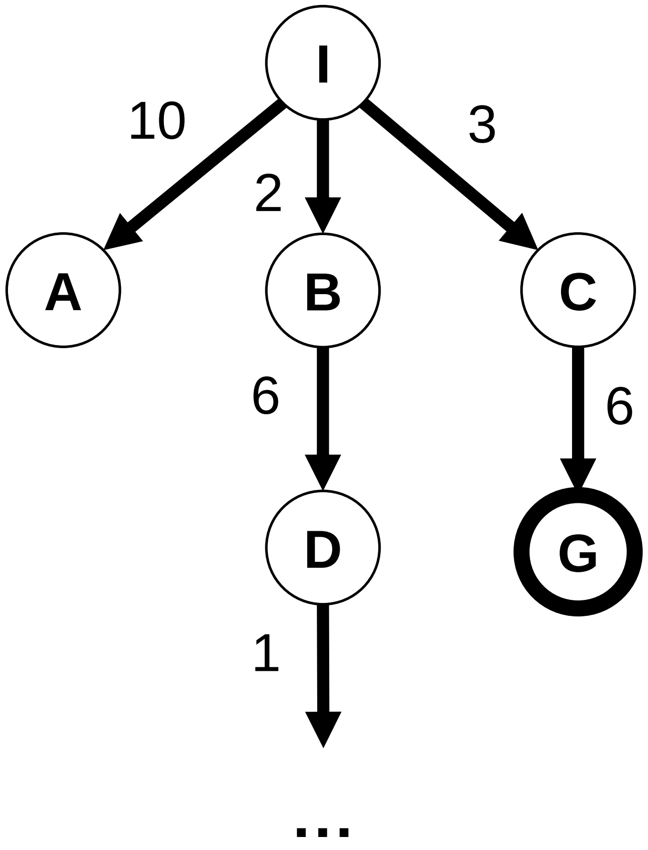
1. 2
2. 7
3. 8
4. 9

Q1-3: You are running UCS in the state space graph below. You just expanded (visited) node **C**. What node will the search expand next?



1. A
2. D
3. E
4. F

Q2-1: You are running DFS in the state space graph below. DFS expands nodes left to right. **G** is the goal state. The state space graph is infinite (the path after **D** does not terminate). What is the behavior of DFS?



1. Get stuck in an infinite loop
2. Return A
3. Return G
4. Return "failure"

Q2-2: You need to search a randomly generated state space graph with one goal, uniform edges costs, d=2, and m=100. Considering worst case behavior, do you select BFS or DFS for your search?

1. BFS

2. DFS

Q2-3: You need to search a randomly generated state space graph with one goal, uniform edges costs, d=25, and m=25. Considering worst case behavior, do you select BFS or DFS for your search?

1. BFS
2. DFS

Q2-4: You need to search a randomly generated state space graph with many goals, uniform edges costs, d=5, and m=10. Considering worst case behavior, do you select BFS or DFS for your search?

1. BFS

2. DFS

Q2-5: DFS uses a stack to maintain the fringe. You wish to implement DFS using the general state-space search algorithm that uses a min heap priority queue. What value should be stored in the queue for each node?

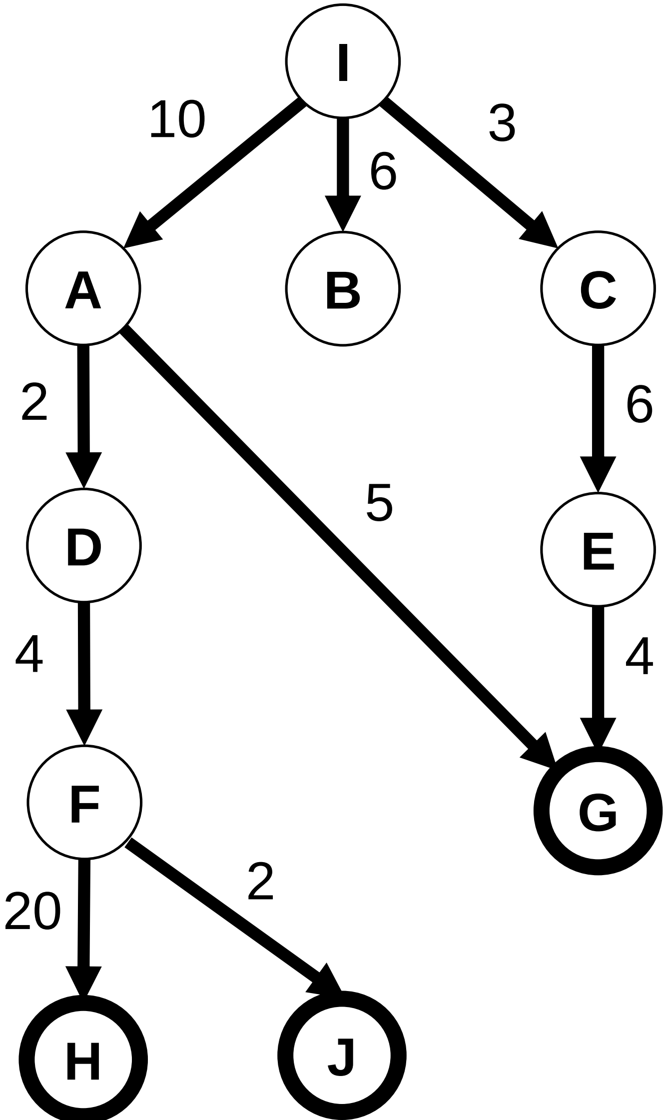
1. Node’s path cost from initial state

2. -1 * (node’s path cost from initial state)

3. Node’s depth from initial state

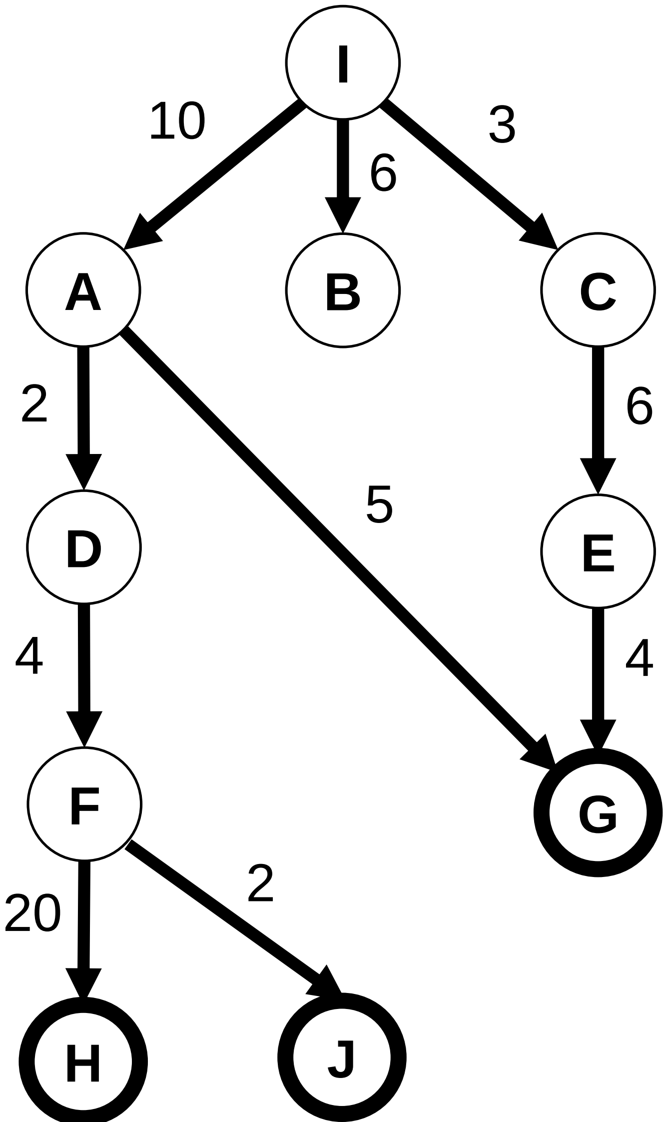
4. -1 * (node’s depth from initial state)

Q3-1: Consider the state space graph below. Goal states have **bold** borders. Nodes are expanded left to right when there are ties. What solution path is returned by BFS?



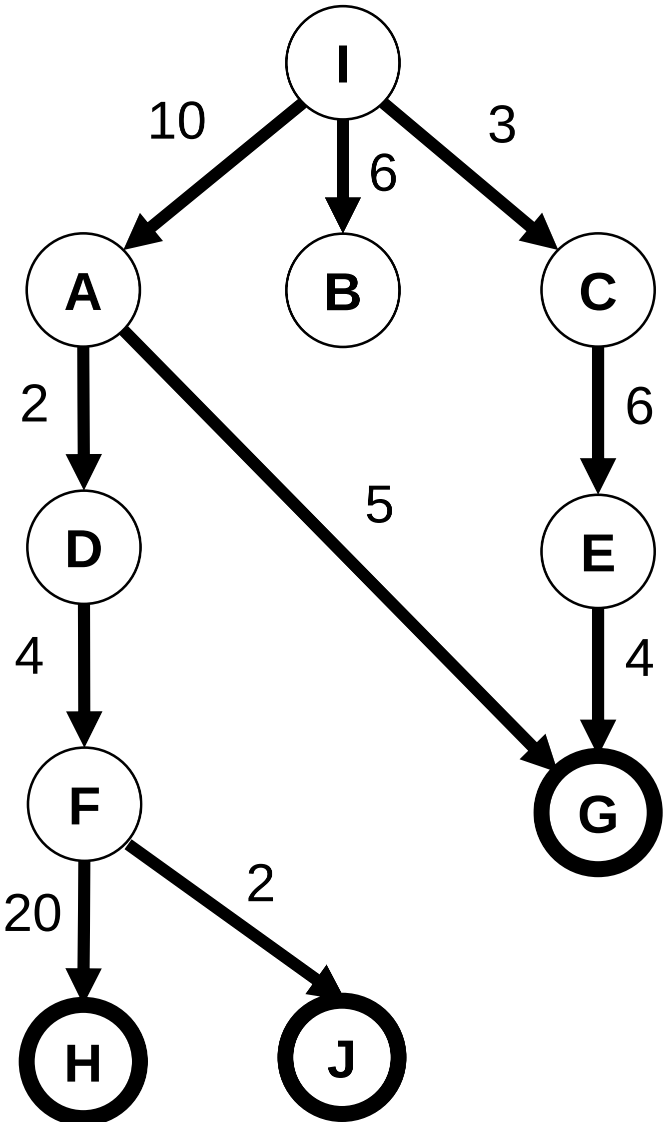
1. IADFH
2. IADFJ
3. IAG
4. ICEG

Q3-2: Consider the state space graph below. Goal states have **bold** borders. Nodes are expanded left to right when there are ties. What solution path is returned by UCS?



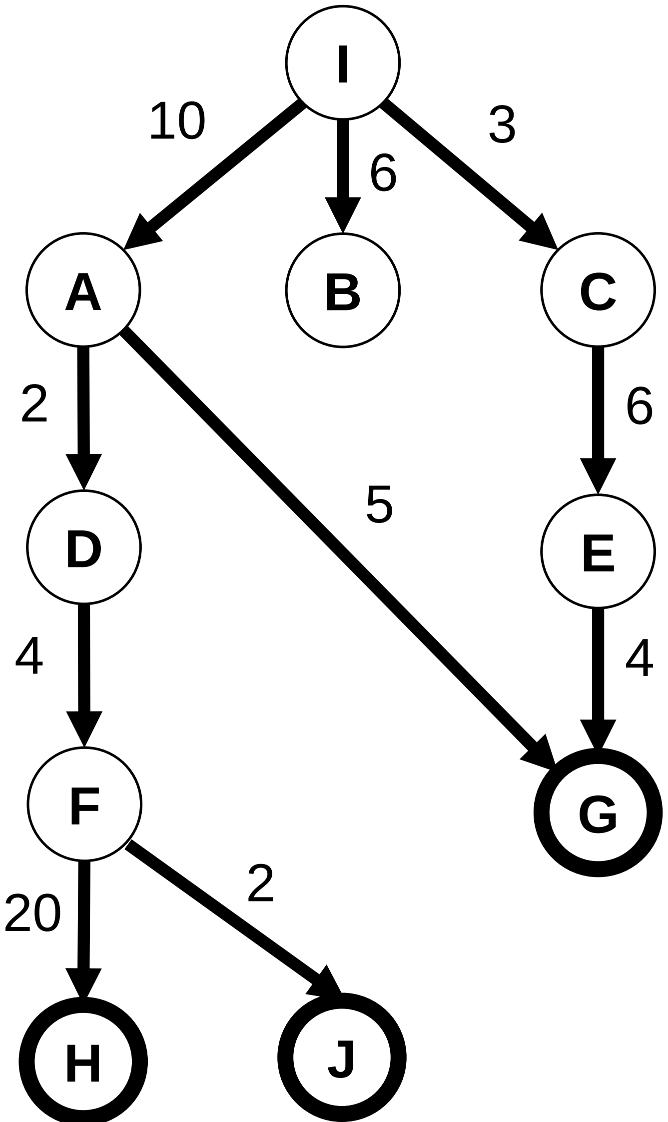
1. IADFH
2. IADFJ
3. IAG
4. ICEG

Q3-3: Consider the state space graph below. Goal states have **bold** borders. Nodes are expanded left to right when there are ties. What solution path is returned by DFS?



1. IADFH
2. IADFJ
3. IAG
4. ICEG

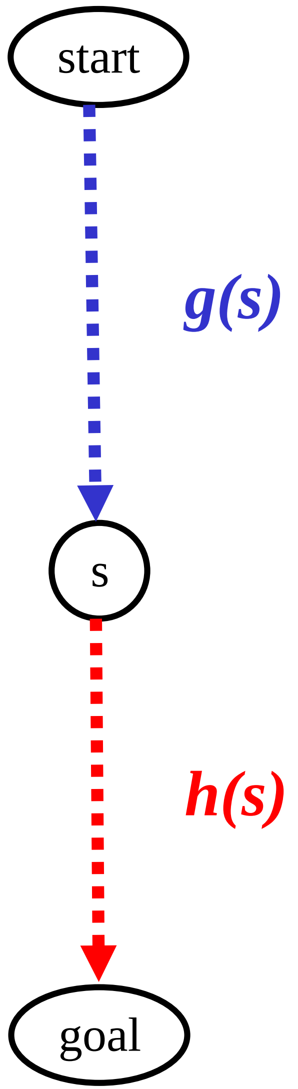
Q3-4: Consider the state space graph below. Goal states have **bold** borders. Nodes are expanded left to right when there are ties. What solution path is returned by IDS?



1. IADFH
2. IADFJ
3. IAG
4. ICEG

Q1-1: Your search algorithm uses a priority queue to track the state with the min score to pop next. If $h(s)$ is an admissible heuristic, what score should you use in the priority queue? Rank them from best to worst.

1. $g(s)+\mathrm{h}(s)$, $h(s)$, $g(s)$
2. $g(s)+\mathrm{h}(s)$, $g(s)$, $h(s)$
3. $g(s)$, $h(s)$, $g(s)+\mathrm{h}(s)$
4. $g(s)$, $g(s)+\mathrm{h}(s)$, $h(s)$



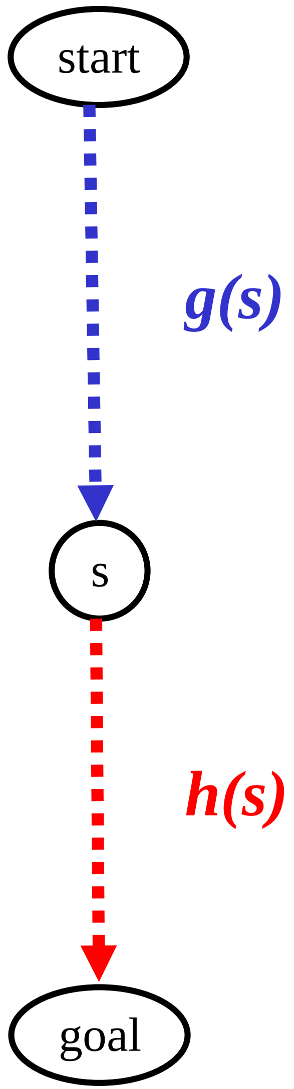
Q1-1: Your search algorithm uses a priority queue to track the state with the min score to pop next. If $h(s)$ is an admissible heuristic, what score should you use in the priority queue? Rank them from best to worst.

1. $g(s)$+h(s), $h(s)$, $g(s)$

2. $g(s)$+h(s), $g(s)$, $h(s)$

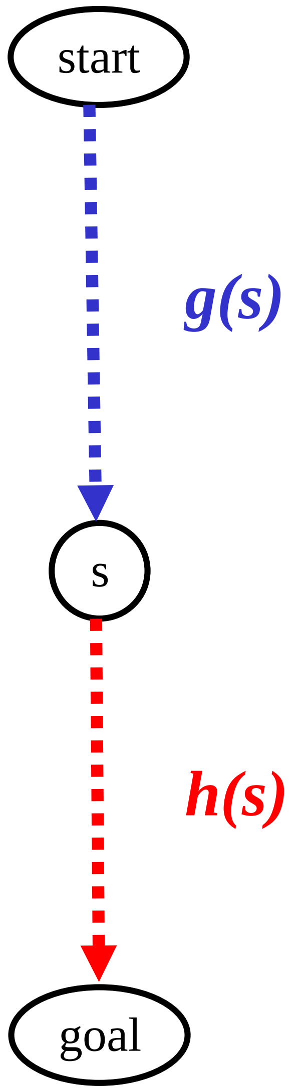
3. $g(s)$, $h(s)$, $g(s)$+h(s)

4. $g(s)$, $g(s)$+h(s), $h(s)$



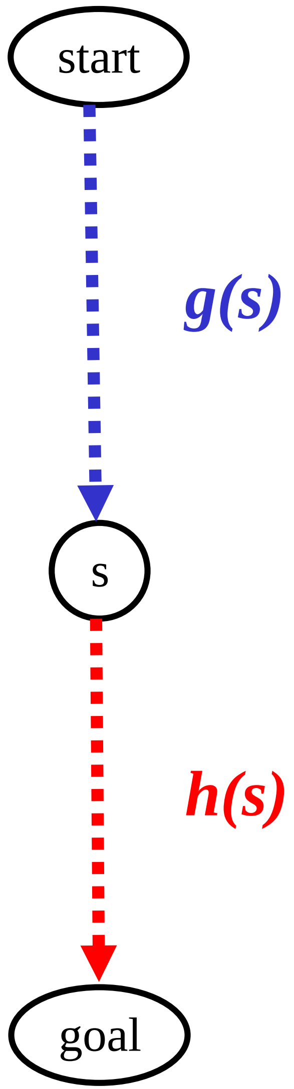
Q1-2: Your search algorithm uses a priority queue to track the state with the min score to pop next. If $h(s)$ is a heuristic (<u>not necessarily admissible</u>), what score should you use in the priority queue? Rank them from best to worst.

1. $g(s)+h(s)$, $h(s)$, $g(s)$
2. $g(s)+h(s)$, $g(s)$, $h(s)$
3. $g(s)$, $h(s)$, $g(s)+h(s)$
4. $g(s)$, $g(s)+h(s)$, $h(s)$



Q1-2: Your search algorithm uses a priority queue to track the state with the min score to pop next. If $h(s)$ is a heuristic (not necessarily admissible), what score should you use in the priority queue? Rank them from best to worst.

1. $g(s)+h(s)$, $h(s)$, $g(s)$
2. $g(s)+h(s)$, $g(s)$, $h(s)$
3. $g(s)$, $h(s)$, $g(s)+h(s)$
4. $g(s)$, $g(s)+h(s)$, $h(s)$



Q1-3: Consider finding the fastest driving route from one US city to another. Measure cost as the number of hours driven when driving at the speed limit. Let $h(s)$ be the number of hours needed to ride a bike from city s to your destination. $h(s)$ is

1. Not a valid heuristic

2. A valid heuristic but not admissible

3. An admissible heuristic

Q1-3: Consider finding the fastest driving route from one US city to another. Measure cost as the number of hours driven when driving at the speed limit. Let $h(s)$ be the number of hours needed to ride a bike from city s to your destination. $h(s)$ is

1. Not a valid heuristic

2. A valid heuristic but not admissible

3. An admissible heuristic

Q1-4: Consider finding the fastest driving route from one US city to another. Measure cost as the number of hours driven when driving at the speed limit. Let $h(s)$ be the current temperature in city s. $h(s)$ is

1. Not a valid heuristic

2. A valid heuristic but not admissible

3. An admissible heuristic

Q1-4: Consider finding the fastest driving route from one US city to another. Measure cost as the number of hours driven when driving at the speed limit. Let $h(s)$ be the current temperature in city s. $h(s)$ is

1. Not a valid heuristic

2. A valid heuristic but not admissible

3. An admissible heuristic

Q1-5: Consider finding the fastest driving route from one US city to another. Measure cost as the number of hours driven when driving at the speed limit. Let $h(s)$ be the number of hours to drive from s to your destination if you ignore speed limits. $h(s)$ is

1. Not a valid heuristic

2. A valid heuristic but not admissible

3. An admissible heuristic

Q1-5: Consider finding the fastest driving route from one US city to another. Measure cost as the number of hours driven when driving at the speed limit. Let $h(s)$ be the number of hours to drive from s to your destination if you ignore speed limits. $h(s)$ is

1. Not a valid heuristic

2. A valid heuristic but not admissible

3. An admissible heuristic ⬅

Q2-1: Consider two heuristics for the 8 puzzle problem. *h1* is the number of tiles in wrong position. *h2* is Manhattan distance between tile and its goal location. How do *h1* and *h2* relate?

1. *h1* dominates *h2*

2. *h2* dominates *h1*

3. Neither heuristic dominates the other

Q2-1: Consider two heuristics for the 8 puzzle problem. *h1* is the number of tiles in wrong position. *h2* is Manhattan distance between tile and its goal location. How do *h1* and *h2* relate?

1. *h1* dominates *h2*

2. *h2* dominates *h1*

3. Neither heuristic dominates the other

Q2-2: Consider finding the fastest driving route from one US city to another, as before. Let $h(s)$ be the minimum number of hours to drive from s to your destination, returned from 1000s of physics simulations of driving in different weather conditions. $h(s)$ is

1. Not a valid heuristic

2. A valid heuristic but not admissible

3. An admissible heuristic

4. An admissible heuristic but a poor choice for A*

Q2-2: Consider finding the fastest driving route from one US city to another, as before. Let $h(s)$ be the <u>minimum</u> number of hours to drive from s to your destination, returned from 1000s of physics simulations of driving in different weather conditions. $h(s)$ is

1. Not a valid heuristic

2. A valid heuristic but not admissible

3. An admissible heuristic

4. An admissible heuristic but a poor choice for A*

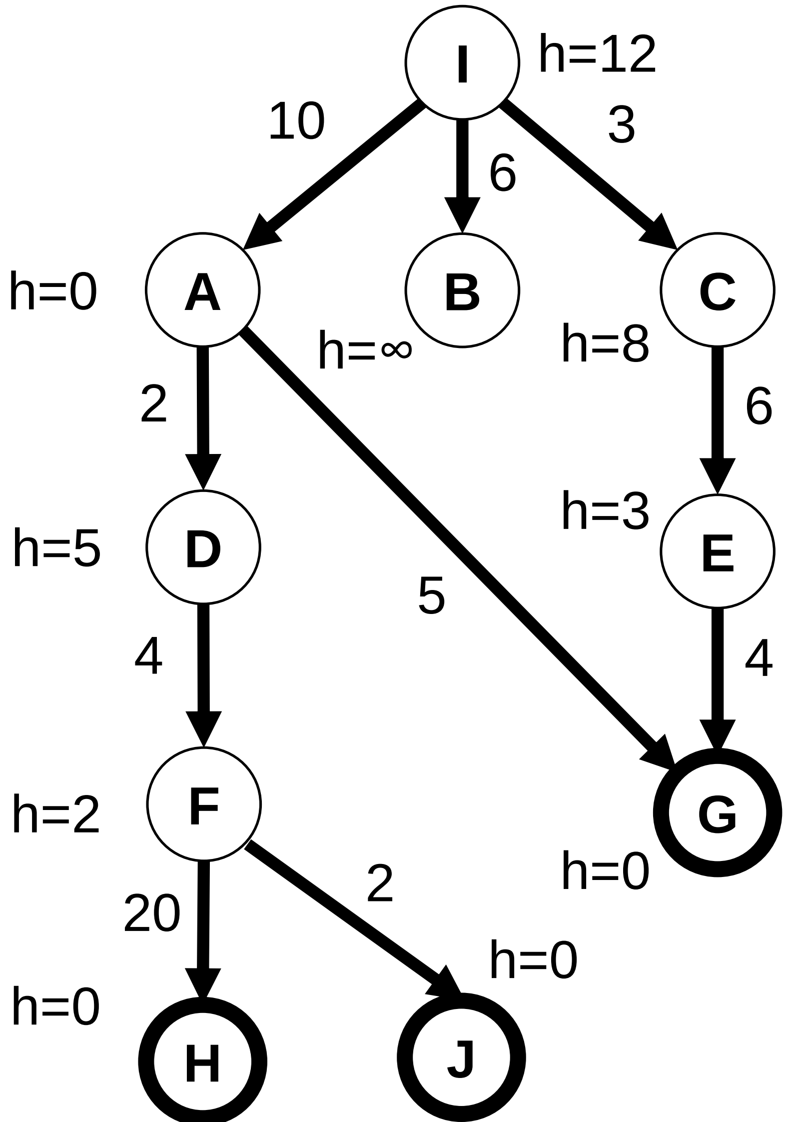
Q2-3: A* search places expanded states in the CLOSED data structure. What is the space complexity of CLOSED?

1. O(1)
2. O(depth of optimal goal)
3. O(log number of states)
4. O(number of states)

Q2-3: A* search places expanded states in the CLOSED data structure. What is the space complexity of CLOSED?

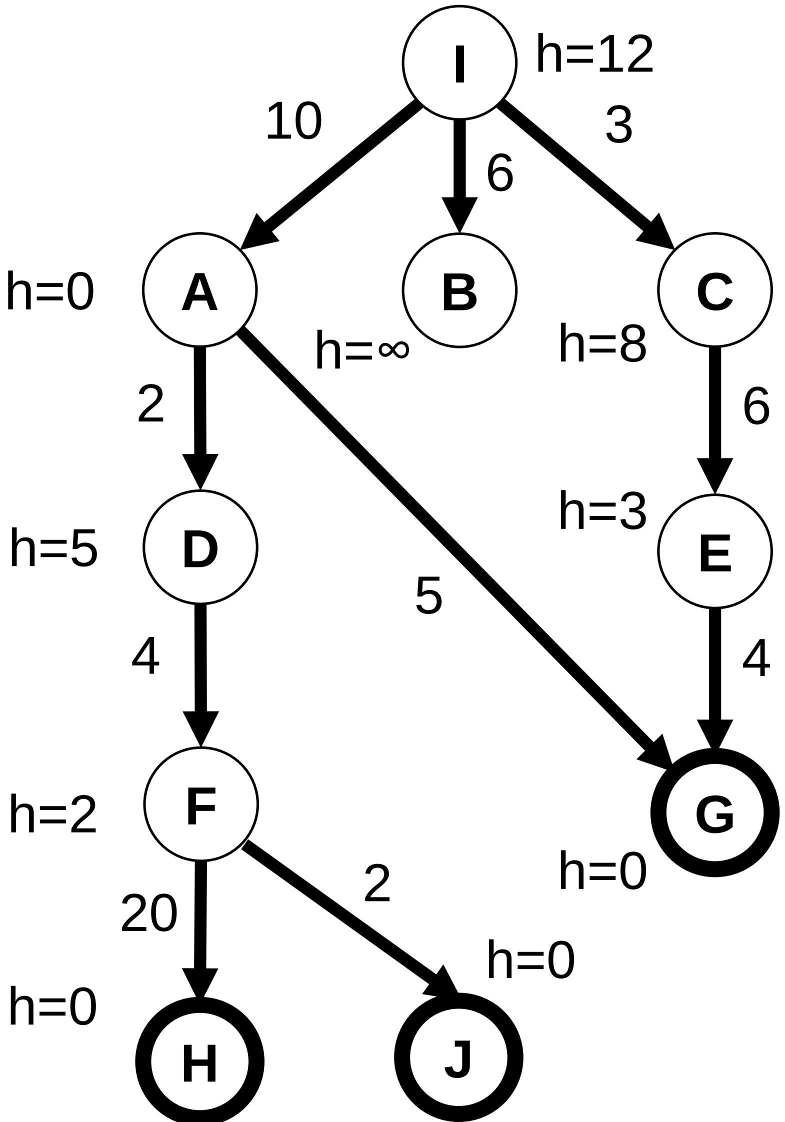
1. O(1)
2. O(depth of optimal goal)
3. O(log number of states)
4. O(number of states)

Q2-4: Consider the state space graph below. Goal states have **bold** borders. *h*(*s*) is show next to each node. What node will be expanded by A* after the initial state **I**?



1. A
2. B
3. C

Q2-4: Consider the state space graph below. Goal states have **bold** borders. *h*(*s*) is show next to each node. What node will be expanded by A* after the initial state **I**?



1. A
2. B
3. C

Q3-1: How does IDA* save space compared to A*?

1. IDA* only tracks the best *k* nodes in the priority queue

2. IDA* will not get stuck exploring deep, suboptimal paths

3. IDA* uses a better heuristic and expands fewer nodes

Q3-1: How does IDA* save space compared to A*?

1. IDA* only tracks the best *k* nodes in the priority queue

2. IDA* will not get stuck exploring deep, suboptimal paths

3. IDA* uses a better heuristic and expands fewer nodes

Q3-2: You are running IDA* in a state space graph with positive real-valued *f*(*s*) scores increasing threshold *t* as described. When *t* increases and IDA* restarts, now many new nodes are expanded?

1. 0 or 1 new nodes

2. Exactly 1 new node

3. 1 or more new nodes

Q3-2: You are running IDA* in a state space graph with positive real-valued *f*(*s*) scores increasing threshold *t* as described. When *t* increases and IDA* restarts, now many new nodes are expanded?

1. 0 or 1 new nodes
2. Exactly 1 new node
3. 1 or more new nodes

Q1-1: What are disadvantages of IDA* search?

1. IDA* has no disadvantages
2. IDA* sometimes returns a suboptimal solution
3. IDA* can visit the same state multiple times during the same iteration
4. When IDA* restarts, it discards all information except the next threshold
5. 3 and 4

Q1-1: What are disadvantages of IDA* search?

1. IDA* has no disadvantages
2. IDA* sometimes returns a suboptimal solution
3. IDA* can visit the same state multiple times during the same iteration
4. When IDA* restarts, it discards all information except the next threshold
5. 3 and 4

Q1-2: Consider the Beam search example with k = 2 and what would happen if we increased k to 3. Under what conditions will Beam search return the optimal solution?

1. When no states are evicted from OPEN because the capacity k has been reached

2. When the same solution is returned for k and k+1

3. When the heuristic used is admissible

Q1-2: Consider the Beam search example with k = 2 and what would happen if we increased k to 3. Under what conditions will Beam search return the optimal solution?

1. When no states are evicted from OPEN because the capacity k has been reached

2. When the same solution is returned for k and k+1

3. When the heuristic used is admissible

Q1-3: Consider proof that A* is optimal. Why can we not use this proof by contradiction to prove that Beam search is optimal?

1. Beam search may not find any path to the goal
2. If Beam search finds a suboptimal path to the goal, it may not have stored unexpanded node *n* on the optimal path
3. Beam search no longer guarantees f(n)=g(n)+h(n)
4. 1 and 2
5. 2 and 3

- Suppose A* finds a suboptimal path ending in goal *G'*, where $f(G') > f^*$ = cost of optimal path
- Let's look at the first unexpanded node *n* on the optimal path (*n* exists, otherwise the optimal goal would have been found)
- $f(n) \geq f(G')$, otherwise we would have expanded *n*
- $f(n) = g(n)+h(n)$

  $= g^*(n)+h(n)$

  $\leq g^*(n)+h^*(n)$

  $= f^*$
- $f^* \geq f(n) \geq f(G')$, contradicting the assumption at top

Q1-3: Consider proof that A* is optimal. Why can we not use this proof by contradiction to prove that Beam search is optimal?

1. Beam search may not find any path to the goal
2. If Beam search finds a suboptimal path to the goal, it may not have stored unexpanded node *n* on the optimal path
3. Beam search no longer guarantees f(n)=g(n)+h(n)
4. 1 and 2
5. 2 and 3

- Suppose A* finds a suboptimal path ending in goal G', where $f(G') > f^*$ = cost of optimal path
- Let's look at the first unexpanded node n on the optimal path (n exists, otherwise the optimal goal would have been found)
- $f(n) \geq f(G')$, otherwise we would have expanded n
- $f(n) = g(n)+h(n)$

  $= g^*(n)+h(n)$

  $\leq g^*(n)+h^*(n)$

  $= f^*$
- $f^* \geq f(n) \geq f(G')$, contradicting the assumption at top

Q2-1: Which of these is <u>*NOT*</u> a reason to prefer optimization instead of search for the 3-SAT optimization problem with a very large number of Boolean variables and clauses.

1. Path cost is not relevant for 3-SAT

2. The state space is very large

3. We need to guarantee we find the Boolean assignments that satisfy the most clauses possible

Q2-1: Which of these is *NOT* a reason to prefer optimization instead of search for the 3-SAT optimization problem with a very large number of Boolean variables and clauses.

1. Path cost is not relevant for 3-SAT

2. The state space is very large

3. We need to guarantee we find the Boolean assignments that satisfy the most clauses possible

Q2-2: You are building a computer that requires *k* parts. You must purchase them from *k* stores because each store has a limit of 1 part per customer. Finding the least expensive way to buy the parts is a:

1. Uninformed search problem
2. Informed search problem
3. Optimization problem

Q2-2: You are building a computer that requires *k* parts. You must purchase them from *k* stores because each store has a limit of 1 part per customer. Finding the least expensive way to buy the parts is a:

1. Uninformed search problem

2. Informed search problem

3. Optimization problem

Q2-3: You are buying *k* computer parts from *k* stores, as in question 2-2. A state is a mapping of which parts you buy from which stores. What is a good choice of neighbor?

1. Select the most expensive part; enumerate the *k*-1 ways to buy that part from the other stores by swapping parts mapped to those stores

2. For each set of 3 parts and their 3 stores, enumerate all the different ways to buy those parts from the 3 stores

3. Fix the store for 1 part; randomly reassign all other *k*-1 parts

Q2-3: You are buying *k* computer parts from *k* stores, as in question 2-2. A state is a mapping of which parts you buy from which stores. What is a good choice of neighbor?

1. Select the most expensive part; enumerate the *k*-1 ways to buy that part from the other stores by swapping parts mapped to those stores

2. For each set of 3 parts and their 3 stores, enumerate all the different ways to buy those parts from the 3 stores

3. Fix the store for 1 part; randomly reassign all other *k*-1 parts

Q3-1: In the 3-SAT example we started from the all T state. Consider what would happen if we instead started in the all F state.

1. Hill climbing would still find the global optimum with $f = 2$

2. Hill climbing would find the global optimum if it visited neighbors in the right order

3. Hill climbing would get stuck in a local optimum

$A \vee \neg B \vee C$
$\neg A \vee C \vee D$
$B \vee D \vee \neg E$
$\neg C \vee \neg D \vee \neg E$
$\neg A \vee \neg C \vee E$

(Note the recorded discussion of this quiz question incorrectly used AND operators instead of OR).

Q3-1: In the 3-SAT example we started from the all T state. Consider what would happen if we instead started in the all F state.

1. Hill climbing would still find the global optimum with $f = 2$

2. Hill climbing would find the global optimum if it visited neighbors in the right order

3. Hill climbing would get stuck in a local optimum

$A \vee \neg B \vee C$
$\neg A \vee C \vee D$
$B \vee D \vee \neg E$
$\neg C \vee \neg D \vee \neg E$
$\neg A \vee \neg C \vee E$

(Note the recorded discussion of this quiz question incorrectly used AND operators instead of OR).

Q3-2: In the 3-SAT example we reached a local optimum in the state (A=T, B=F, C=T, D=T, E=T). How can we modify hill climbing to escape this local optimum?

1. In “IF $f(t) \leq f(s)$ THEN stop” change $\leq$ to <
2. We cannot fix hill climbing to escape this local optimum
3. Pick the neighbor t with the smallest $f(t)$ instead of the largest $f(t)$

$A \vee \neg B \vee C$
$\neg A \vee C \vee D$
$B \vee D \vee \neg E$
$\neg C \vee \neg D \vee \neg E$
$\neg A \vee \neg C \vee E$

(Note the recorded discussion of this quiz question incorrectly used AND operators instead of OR).

Q3-2: In the 3-SAT example we reached a local optimum in the state (A=T, B=F, C=T, D=T, E=T). How can we modify hill climbing to escape this local optimum?

1. In "IF $f(t) \leq f(s)$ THEN stop" change $\leq$ to $<$

2. We cannot fix hill climbing to escape this local optimum

3. Pick the neighbor t with the smallest $f(t)$ instead of the largest $f(t)$

$A \lor \neg B \lor C$
$\neg A \lor C \lor D$
$B \lor D \lor \neg E$
$\neg C \lor \neg D \lor \neg E$
$\neg A \lor \neg C \lor E$

(Note the recorded discussion of this quiz question incorrectly used AND operators instead of OR).

Q3-3: What is a disadvantage of modifying hill climbing to explore a plateau?

1. Unlike the original hill climbing, it will now be sensitive to the initial state

2. It may waste time exploring and never improve the *f* score

3. It may choose a neighbor with a lower score than the current state

Q3-3: What is a disadvantage of modifying hill climbing to explore a plateau?

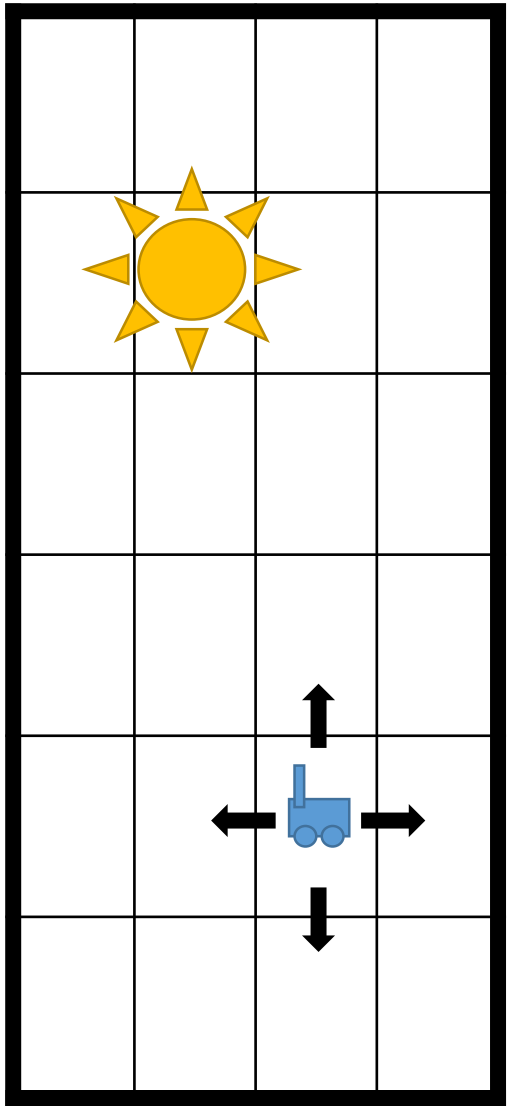
1. Unlike the original hill climbing, it will now be sensitive to the initial state

2. It may waste time exploring and never improve the *f* score

3. It may choose a neighbor with a lower score than the current state

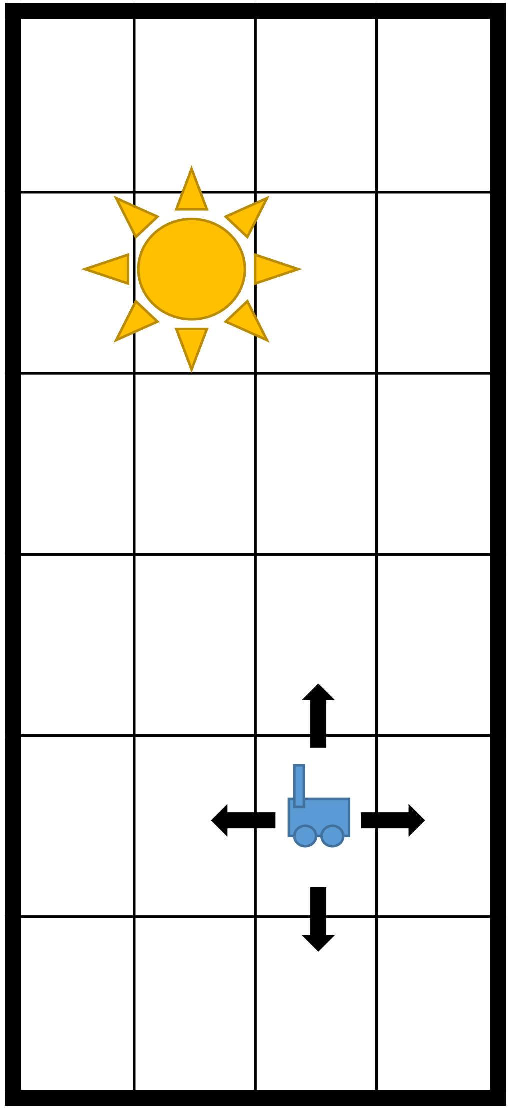
Q1-1: A robot in a large 2D room needs to find the location with the most sunlight so it can recharge. There is 1 skylight letting in light. What optimization strategy should you use?

1. Random search
2. Hill climbing without random restarts
3. Stochastic hill climbing without random restarts
4. First-choice hill climbing without random restarts



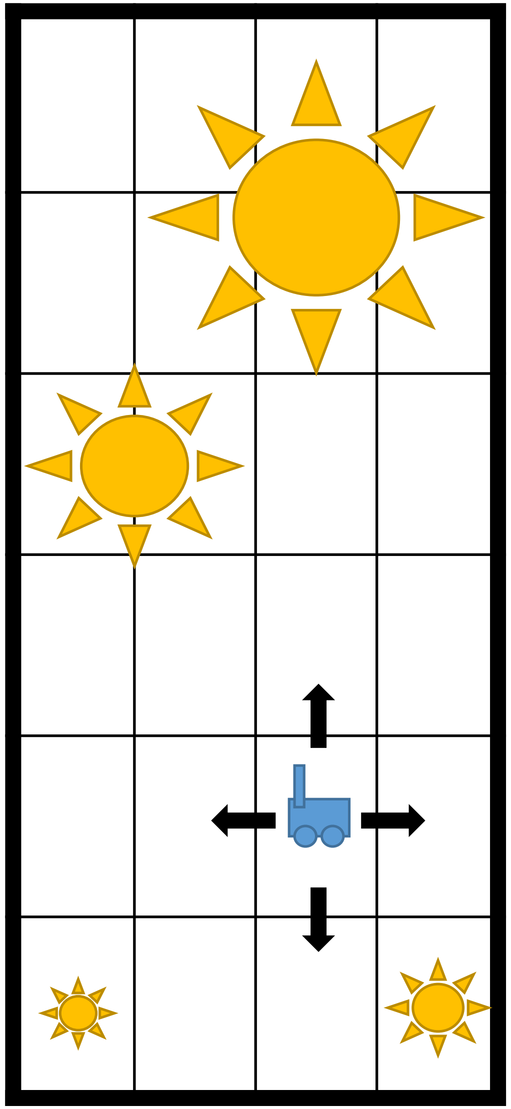
Q1-1: A robot in a large 2D room needs to find the location with the most sunlight so it can recharge. There is 1 skylight letting in light. What optimization strategy should you use?

1. Random search
2. Hill climbing without random restarts
3. Stochastic hill climbing without random restarts
4. First-choice hill climbing without random restarts



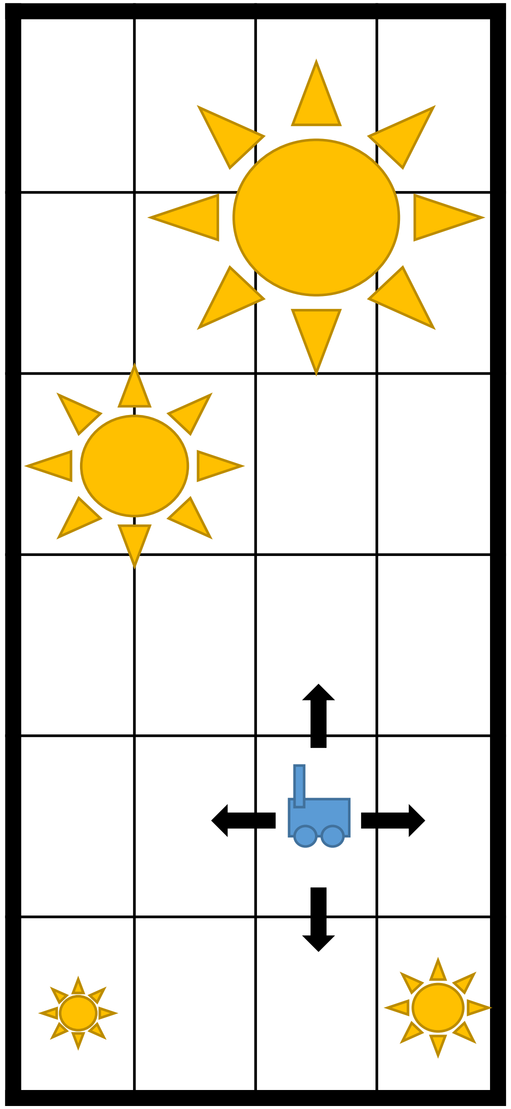
Q1-2: A robot in a large 2D room needs to find the location with the most sunlight so it can recharge. There are many skylights letting in different amounts of light. What optimization strategy should you use?

1. Random search
2. Hill climbing without random restarts
3. Stochastic hill climbing without random restarts
4. First-choice hill climbing without random restarts



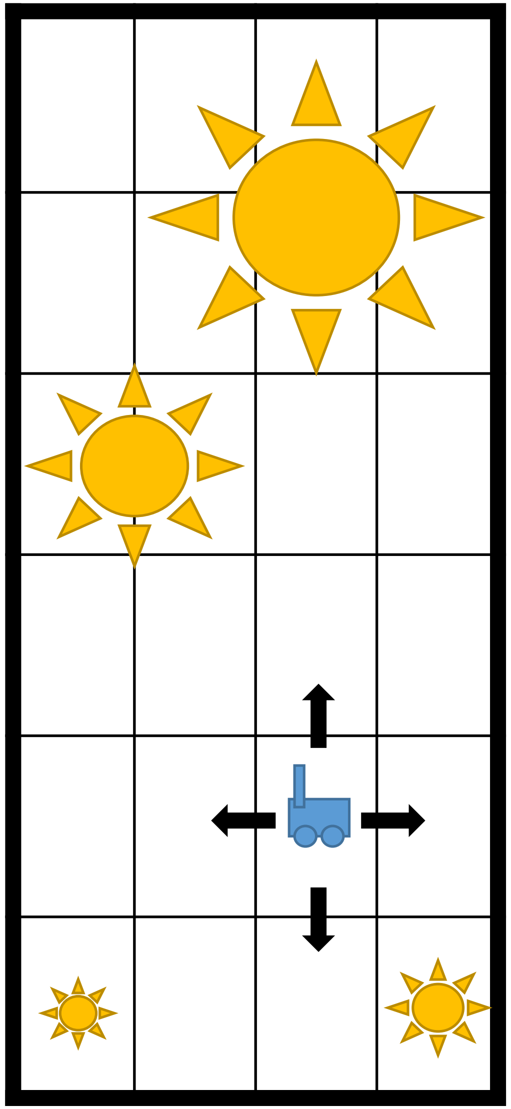
Q1-2: A robot in a large 2D room needs to find the location with the most sunlight so it can recharge. There are many skylights letting in different amounts of light. What optimization strategy should you use?

1. Random search
2. Hill climbing without random restarts
3. Stochastic hill climbing without random restarts
4. First-choice hill climbing without random restarts



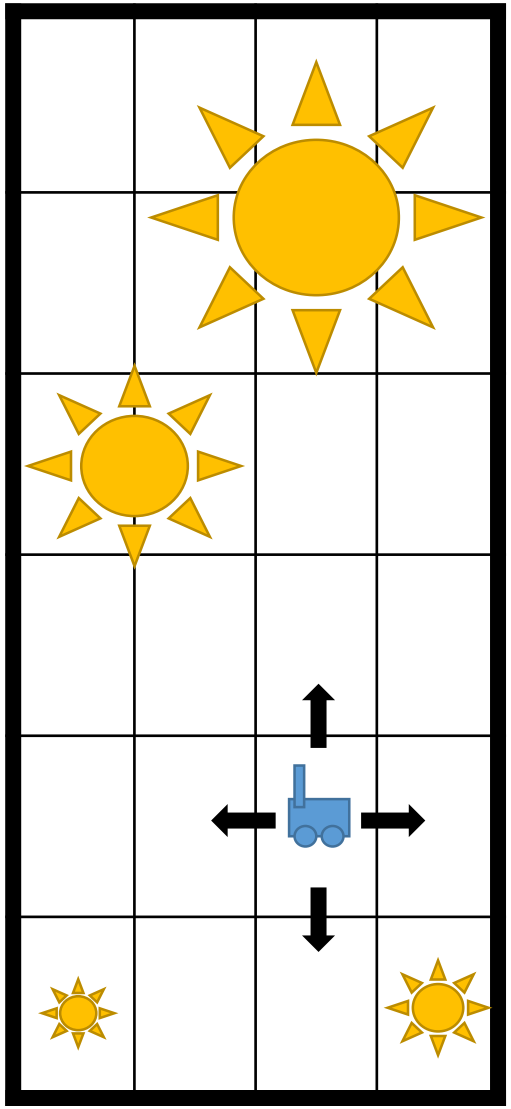
Q1-3: A robot in a large 2D room needs to find the location with the most sunlight so it can recharge. There are many skylights letting in different amounts of light. What optimization strategy should you use to guarantee you find the global maximum?

1. Random search
2. Hill climbing without random restarts
3. Stochastic hill climbing without random restarts
4. First-choice hill climbing without random restarts



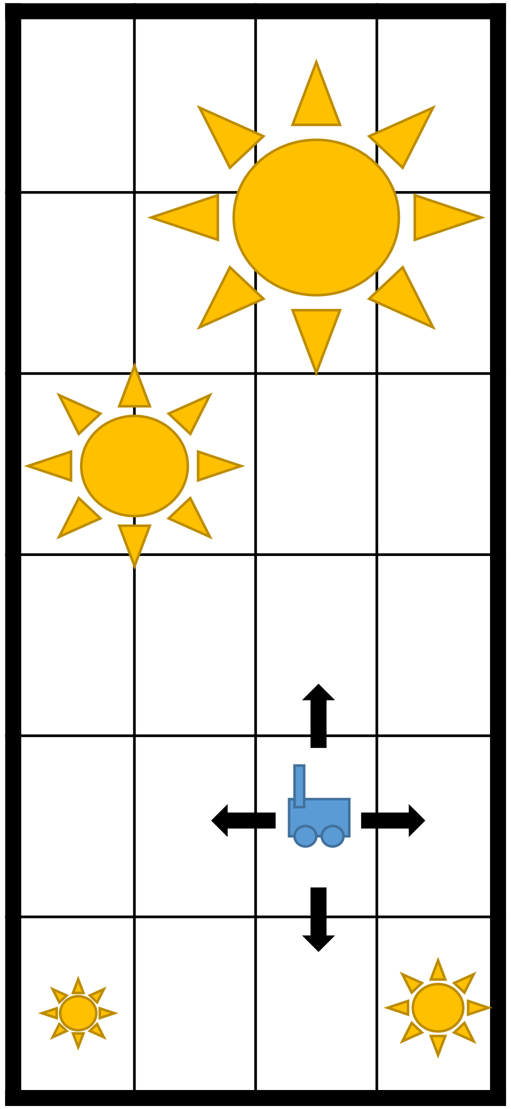
Q1-3: A robot in a large 2D room needs to find the location with the most sunlight so it can recharge. There are many skylights letting in different amounts of light. What optimization strategy should you use to guarantee you find the global maximum?

1. Random search
2. Hill climbing without random restarts
3. Stochastic hill climbing without random restarts
4. First-choice hill climbing without random restarts



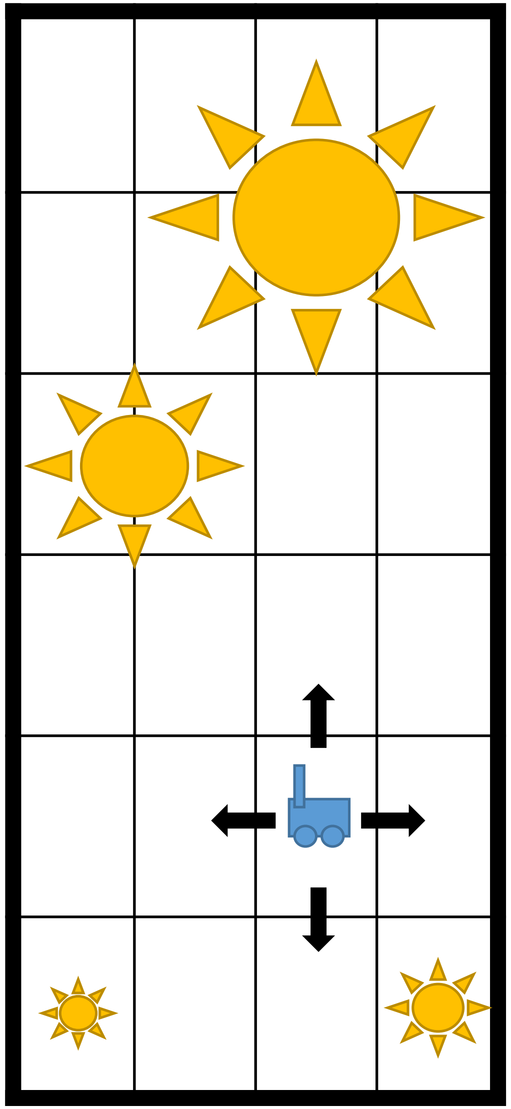
Q1-4: A robot in a large 2^100D room needs to find the location with the most sunlight so it can recharge. There are many skylights letting in different amounts of light. What optimization strategy should you use?

1. Random search
2. Hill climbing without random restarts
3. Stochastic hill climbing without random restarts
4. First-choice hill climbing without random restarts



Q1-4: A robot in a large 2^100D room needs to find the location with the most sunlight so it can recharge. There are many skylights letting in different amounts of light. What optimization strategy should you use?

1. Random search
2. Hill climbing without random restarts
3. Stochastic hill climbing without random restarts
4. First-choice hill climbing without random restarts



Q2-1: What is an advantage of simulated annealing over hill climbing, stochastic hill climbing, and first-choice hill climbing?

1. It is guaranteed to find the global optimum
2. Algorithms inspired by real world processes work better
3. It is less vulnerable to getting stuck in local optima
4. It terminates more quickly

Q2-1: What is an advantage of simulated annealing over hill climbing, stochastic hill climbing, and first-choice hill climbing?

1. It is guaranteed to find the global optimum
2. Algorithms inspired by real world processes work better
3. It is less vulnerable to getting stuck in local optima
4. It terminates more quickly

Q2-2: What is a <u>not</u> an advantage of simulated annealing over random search?

1. It is less vulnerable to getting stuck in local optima

2. It can make use of a good neighbor function

3. It can climb a local optimum

Q2-2: What is a not an advantage of simulated annealing over random search?

1. It is less vulnerable to getting stuck in local optima

2. It can make use of a good neighbor function

3. It can climb a local optimum

Q2-3: What setting will give the largest probability of moving to state *t*?

$$\exp\left(-\frac{|f(s)-f(t)|}{Temp}\right)$$

1. *f*(*t*) is close to *f*(*s*) and *Temp* is large

2. *f*(*t*) is close to *f*(*s*) and *Temp* is small

3. *f*(*t*) is much less than *f*(*s*) and *Temp* is large

4. *f*(*t*) is much less than *f*(*s*) and *Temp* is small

Q2-3: What setting will give the largest probability of moving to state *t*?

1. *f*(*t*) is close to *f*(*s*) and *Temp* is large ⟵

   $$\exp\left(-\frac{|f(s)-f(t)|}{Temp}\right)$$

2. *f*(*t*) is close to *f*(*s*) and *Temp* is small

3. *f*(*t*) is much less than *f*(*s*) and *Temp* is large

4. *f*(*t*) is much less than *f*(*s*) and *Temp* is small

Q3-1: Recall the optimization problem of buying *k* computer parts from *k* stores, one part per store, at the least cost. What would be a good encoding for a genetic algorithm?

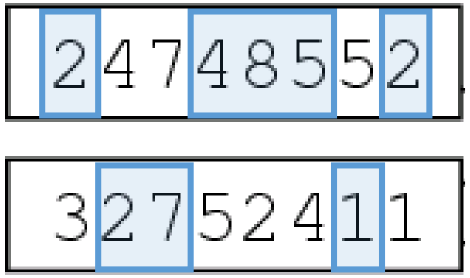
1. A graph in which edges connect a store with the part purchased there
2. A list of strings <*part*>-<*store*> sorted by the price of that part
3. A string with a fixed order for the parts and integer index for each store

Q3-1: Recall the optimization problem of buying *k* computer parts from *k* stores, one part per store, at the least cost. What would be a good encoding for a genetic algorithm?

1. A graph in which edges connect a store with the part purchased there
2. A list of strings <*part*>-<*store*> sorted by the price of that part
3. A string with a fixed order for the parts and integer index for each store

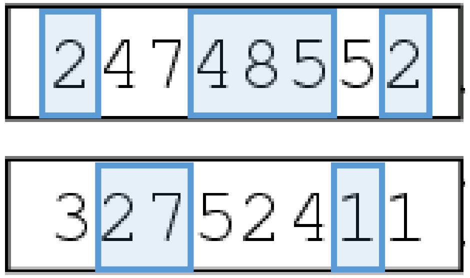
Q3-2: Consider an alternative crossover operation for a genetic algorithm run on N-queens that randomly selects several columns from one parent and the remaining columns from the other parent. Will this work as well as the standard crossover? Why?

1. No, it is impossible to select the indices in this manner
2. No, it will not preserve partial solutions on one side of the board
3. Yes, it still helps make large random moves in the state space
4. Yes, it is more closely related to real biology



Q3-2: Consider an alternative crossover operation for a genetic algorithm run on N-queens that randomly selects several columns from one parent and the remaining columns from the other parent. Will this work as well as the standard crossover? Why?

1. No, it is impossible to select the indices in this manner
2. No, it will not preserve partial solutions on one side of the board
3. Yes, it still helps make large random moves in the state space
4. Yes, it is more closely related to real biology



Q1-1: Consider running a genetic algorithm with a mutation probability of 0. Is the mean population fitness at iteration i+1 <u>expected</u> to be greater than mean population fitness at iteration i?

1. No
2. Yes
3. Unsure

Q1-1: Consider running a genetic algorithm with a mutation probability of 0. Is the mean population fitness at iteration i+1 <u>expected</u> to be greater than mean population fitness at iteration i?

1. No
2. Yes
3. Unsure

Q1-2: Consider running a genetic algorithm with a mutation probability of 0. Is the mean population fitness at iteration *i*+1 always guaranteed to be greater than mean population fitness at iteration *i*?

1. No

2. Yes

3. Unsure

Q1-2: Consider running a genetic algorithm with a mutation probability of 0. Is the mean population fitness at iteration $i+1$ always guaranteed to be greater than mean population fitness at iteration i?

1. No
2. Yes
3. Unsure

Q1-3: What could go wrong in a genetic algorithm if the mutation probability is 0.95?

1. The population would converge to similar states too quickly
2. The algorithm would randomly explore the state space
3. The cross-over operation would be too slow
4. We would be unable to calculate fitness scores

Q1-3: What could go wrong in a genetic algorithm if the mutation probability is 0.95?

1. The population would converge to similar states too quickly
2. The algorithm would randomly explore the state space
3. The cross-over operation would be too slow
4. We would be unable to calculate fitness scores

Q1-4: What could go wrong in a genetic algorithm if the initial population contains the same state *N* times?

1. The population would converge to similar states too quickly
2. The algorithm would randomly explore the state space
3. The cross-over operation would be too slow
4. We would be unable to calculate fitness scores

Q1-4: What could go wrong in a genetic algorithm if the initial population contains the same state *N* times?

1. The population would converge to similar states too quickly
2. The algorithm would randomly explore the state space
3. The cross-over operation would be too slow
4. We would be unable to calculate fitness scores

Q2-1: Consider the Rubik's cube puzzle. The goal is to rotate the cube so that all 6 faces have tiles of the same color. Which is an admissible heuristic?

1. The worst case number of moves needed to reach the goal from any initial state
2. 0
3. The number of moves needed to solve the face with a blue center tile
4. 1 and 2
5. 2 and 3

Image from https://www.fiverr.com/darkdragon532/teach-you-how-to-solve-a-rubiks-cube-3x3

Q2-1: Consider the Rubik's cube puzzle. The goal is to rotate the cube so that all 6 faces have tiles of the same color. Which is an admissible heuristic?

1. The worst case number of moves needed to reach the goal from any initial state
2. 0
3. The number of moves needed to solve the face with a blue center tile
4. 1 and 2
5. 2 and 3

Image from https://www.fiverr.com/darkdragon532/teach-you-how-to-solve-a-rubiks-cube-3x3

Q2-2: Consider the Rubik’s cube puzzle. The goal is to rotate the cube so that all 6 faces have tiles of the same color. Which is an admissible heuristic?

1. The number of moves required if we can only rotate the faces up or to the right.
2. The number of misplaced edge tiles
3. The number of misplaced edge tiles on the blue face minus the number of misplaced edge tiles on the orange face

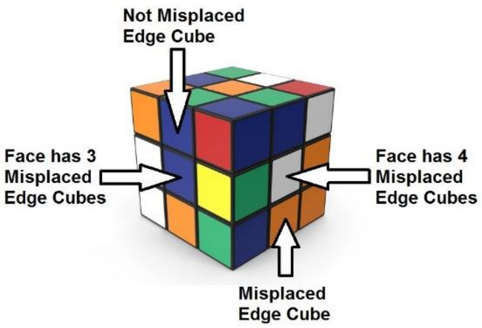


Image from https://www.cs.huji.ac.il/~ai/projects/2017/heuristic_%20search/learning_heuristics_for_the_rubiks_cube/

Q2-2: Consider the Rubik's cube puzzle. The goal is to rotate the cube so that all 6 faces have tiles of the same color. Which is an admissible heuristic?

1. The number of moves required if we can only rotate the faces up or to the right.
2. The number of misplaced edge tiles
3. The number of misplaced edge tiles on the blue face minus the number of misplaced edge tiles on the orange face

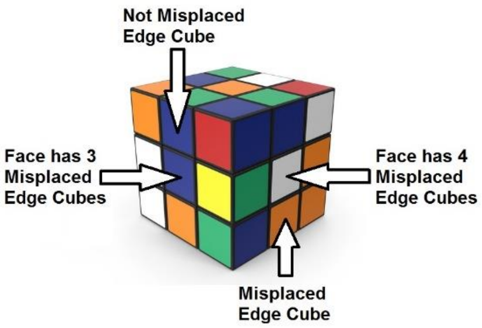


Image from https://www.cs.huji.ac.il/~ai/projects/2017/heuristic_%20search/learning_heuristics_for_the_rubiks_cube/

Q2-3: h_1 and h_2 are admissible heuristics. Which of the following are also admissible?

1. $\max(h_{1,}\ h_2)$
2. $h_1 + h_2$
3. $(1-c)*h_1 + c*h_2$, c in [0, 1]
4. 1 and 3
5. All of the above

Q2-3: h_1 and h_2 are admissible heuristics. Which of the following are also admissible?

1. $\max(h_1, h_2)$

2. $h_1 + h_2$

3. $(1-c)*h_1 + c*h_2$, c in [0, 1]

4. 1 and 3

5. All of the above

Q3-1: Which of the following is likely to give the best Temp schedule for simulated annealing?

1. $Temp_{t+1} = Temp_t * 1.25$

2. $Temp_{t+1} = Temp_t$

3. $Temp_{t+1} = Temp_t * 0.8$

4. $Temp_{t+1} = Temp_t * 0.0001$

Q3-1: Which of the following is likely to give the best Temp schedule for simulated annealing?

1. $Temp_{t+1} = Temp_t * 1.25$

2. $Temp_{t+1} = Temp_t$

3. $Temp_{t+1} = Temp_t * 0.8$

4. $Temp_{t+1} = Temp_t * 0.0001$

Q3-2: Which of the following would be better to solve with simulated annealing than A* search?

1. Finding the smallest set of vertices in a graph that involve all edges
2. Finding the fastest way to schedule jobs with varying runtimes on machines with varying processing power
3. Finding the fastest way through a maze
4. 1 and 2
5. 2 and 3

Q3-2: Which of the following would be better to solve with simulated annealing than A* search?

1. Finding the smallest set of vertices in a graph that involve all edges
2. Finding the fastest way to schedule jobs with varying runtimes on machines with varying processing power
3. Finding the fastest way through a maze
4. 1 and 2
5. 2 and 3

Q 1-1: Consider a biased coin toss. If P(heads) = 0.6, then P(tails) = ?

A. 0.4

B. 0.5

C. 0.6

D. 0.3

Q 1-1: Consider a biased coin toss. If P(heads) = 0.6, then P(tails) = ?

A. 0.4
B. 0.5
C. 0.6
D. 0.3

Q 1-2: In a presidential election, there are 3 candidates, A, B and C. Based on our polling analysis, we estimate that A has a 30 percent chance of winning the election, while B has a 40 percent chance of winning. What is the probability that either A or B win the election?

A. 50%

B. 70%

C. 40%

D. 100%

Q 1-2: In a presidential election, there are 3 candidates, A, B and C. Based on our polling analysis, we estimate that A has a 30 percent chance of winning the election, while B has a 40 percent chance of winning. What is the probability that either A or B win the election?

A. 50%

B. 70%

C. 40%

D. 100%

Q 1-3: What is the probability of selecting a black card or a number 6 from a deck of 52 cards?

A. 26 / 52

B. 4 / 52

C. 30 / 52

D. 28 / 52

Q 1-3: What is the probability of selecting a black card or a number 6 from a deck of 52 cards?

A. 26 / 52

B. 4 / 52

C. 30 / 52

D. 28 / 52

Q 2-1: Consider the joint probability distribution given below.

	weather = sunny	**weather = cloudy**	**weather = rainy**
temp= hot	150 / 365	40 / 365	5 / 365
temp = cold	50 / 365	60 / 365	60 / 365

What is the probability that the temperature is hot given the weather is cloudy?

A. 40 / 365

B. 2 / 5

C. 3 / 5

D. 195 / 365

Q 2-1: Consider the joint probability distribution given below.

	weather = sunny	**weather = cloudy**	**weather = rainy**
temp= hot	150 / 365	40 / 365	5 / 365
temp = cold	50 / 365	60 / 365	60 / 365

What is the probability that the temperature is hot given the weather is cloudy?

A. 40 / 365

B. 2 / 5

C. 3 / 5

D. 195 / 365

Q 2-2: Of a company's employees, 30% are women and 6% are married women. Suppose an employee is selected at random. If the employee selected is a woman, what is the probability that she is married?

A. 0.06

B. 0.3

C. 0.2

D. 0.24

Q 2-2: Of a company's employees, 30% are women and 6% are married women. Suppose an employee is selected at random. If the employee selected is a woman, what is the probability that she is married?

A. 0.06

B. 0.3

C. 0.2

D. 0.24

Q 3-1: It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

A. 5 / 104

B. 95 / 100

C. 1 / 100

D. 1 / 2

Q 3-1: It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

A. 5 / 104

B. 95 / 100

C. 1 / 100

D. 1 / 2

Q 3-2: If a fair coin is tossed three times, find the probability of getting 2 heads and a tail

A. 1 / 8

B. 2 / 8

C. 3 / 8

D. 5 / 8

Q 3-2: If a fair coin is tossed three times, find the probability of getting 2 heads and a tail

A. 1 / 8

B. 2 / 8

C. 3 / 8

D. 5 / 8

Q1.1 On a multiple choice test, problem A has 4 choices, while problem B has 3. Assume that each problem has 1 correct answer. What is the probability of guessing the correct answer to both of the problems?

A. $\frac{1}{4} + \frac{1}{3}$

B. $\frac{1}{4} \times \frac{1}{3}$

C. $\frac{1}{4} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{3}$

D. None of the above

Q1.1 On a multiple choice test, problem A has 4 choices, while problem B has 3. Assume that each problem has 1 correct answer. What is the probability of guessing the correct answer to both of the problems?

A. $\frac{1}{4}+\frac{1}{3}$

B. $\frac{1}{4}\times\frac{1}{3}$

C. $\frac{1}{4}\times\frac{1}{3}+\frac{1}{3}\times\frac{2}{3}$

D. None of the above

Q1.2 Consider a fair die, and the following three events:

X = rolling any of {1, 2}

Y = rolling any of {2, 4, 6}

Z = rolling any of {1, 4}

In other words, P(X) = 1/3, P(Y) = 1/2, P(Z) = 1/3.

Are events X and Y independent? Are events X and Y independent given event Z?

A. Yes, Yes

B. No, No

C. Yes, No

D. No, Yes

Q1.2 Consider a fair die, and the following three events:

X = rolling any of {1, 2}

Y = rolling any of {2, 4, 6}

Z = rolling any of {1, 4}

In other words, P(X) = 1/3, P(Y) = 1/2, P(Z) = 1/3.

Are events X and Y independent? Are events X and Y independent given event Z?

A. Yes, Yes

B. No, No

C. Yes, No

D. No, Yes

Q2.1 We have a piece a text:

`It was the best of times, it was the worst of times.`

Suppose our vocabulary is `["it", "was", "best", "of", "times", "worst"]`

What is the bag of words representation of this text?

A. [2, 2, 1, 2, 2, 1]

B. [2, 2, 1, 2, 2, 1] / 6

C. [2, 2, 1, 2, 2, 1] / 10

D. [1, 1, 2, 1, 1, 2] / 10

Q2.1 We have a piece a text:

`It was the best of times, it was the worst of times.`

Suppose our vocabulary is `["it", "was", "best", "of", "times", "worst"]`

What is the bag of words representation of this text?

A. [2, 2, 1, 2, 2, 1]

B. [2, 2, 1, 2, 2, 1] / 6

C. [2, 2, 1, 2, 2, 1] / 10

D. [1, 1, 2, 1, 1, 2] / 10

Q2.2 We have a a corpus containing only the following documents.

Document ID 1: "A time to plant and a time to reap"

Document ID 2: "Time for you and time for me"

Document ID 3: "Time flies"

Given that the stemmed version of the word "flies" is the term "fly", what is the tf-idf of "fly" in document 3?

A. log(3)

B. log(3)/3

C. log(2)

D. log(2)/2

Q2.2 We have a a corpus containing only the following documents.

Document ID 1: "A time to plant and a time to reap"

Document ID 2: "Time for you and time for me"

Document ID 3: "Time flies"

Given that the stemmed version of the word "flies" is the term "fly", what is the tf-idf of "fly" in document 3?

A. log(3)

B. log(3)/3

C. log(2)

D. log(2)/2

Q2.3 Given the following two document vectors, what is their cosine similarity?

v_a = [0.5, 1, 2]

v_b = [2, 1, 0.5]

A. 0.571

B. 0.99

C. 1.909

D. -0.99

Q2.3 Given the following two document vectors, what is their cosine similarity?

$v_a = [0.5,\ 1,\ \ 2]$

$v_b = [\ 2,\ 1,\ 0.5]$

A. 0.571

B. 0.99

C. 1.909

D. -0.99

Q3.1 Suppose “the dog ran away” is our training corpus.

What is P(ran away) if we use a unigram model?

A. 0

B. 1/2

C. 1/4

D. 1/16

Q3.1 Suppose "the dog ran away" is our training corpus.

What is P(ran away) if we use a unigram model?

A. 0

B. 1/2

C. 1/4

D. 1/16

Q 3.2: Suppose "the dog ran away" is the training corpus. What is P(ran|dog) if we use a bigram model with Laplace Smoothing?

A. 1/4

B. 1

C. 2/5

D. 1/2

Q 3.2: Suppose “the dog ran away” is the training corpus. What is P(ran|dog) if we use a bigram model with Laplace Smoothing?

A. 1/4

B. 1

C. 2/5

D. 1/2

[Block 1: MLE & MAP]

Q 1.1 Consider we choose number Uniformly from a set $\{1,2,3,4, \ldots\theta\}$ (θ is an integer) with replacement. Suppose the numbers we choose are 2,5,7, then based on MLE, what value of θ do you estimate?

Options:

1) 3
2) 5
3) 7
4) 9

[Block 1: MLE & MAP]

Q 1.1 Consider we choose number Uniformly from a set {1,2,3,4, …θ} (θ is an integer) with replacement. Suppose the numbers we choose are 2,5,7, then based on MLE, what value of θ do you estimate?

Options:

5) 3
6) 5
7) **7**
8) 9

Q 1.2 For the above example, if beforehand we know that $Pr(\theta=5) = 1/3$, $Pr(\theta=8) = 1/6$, $Pr(\theta=9) = 1/2$. Then after we see the numbers we choose are 2,5,7, what value of θ do you think it is most likely to be?

Options:

1) 5
2) 7
3) 8
4) 9

Q 1.2 For the above example, if beforehand we know that Pr(θ=5) = 1/3, Pr(θ=8) = 1/6, Pr(θ=9) = 1/2. Then after we see the numbers we choose are 2,5,7, what value of θ do you think it is most likely to be?

Options:

1) 5
2) 7
3) 8
4) **9**

[Block 2: Naïve Bayes]

Q 2.1 Consider a classification problem with $n = 32$, $y \in \{1, 2, 3, \dots, n\}$, and two binary features, $x_1, x_2 \in \{0,1\}$. Suppose $P(Y=y) = 1/32$, $P(x_1 = 1 | Y = y) = y / 46$, $P(x_2 = 1 | Y = y) = y / 62$. Which class will naive Bayes classifier produce on a test item with $x_1 = 1$ and $x_2 = 0$?

Options:

1) 16
2) 26
3) 31
4) 32

[Block 2: Naïve Bayes]

Q 2.1 Consider a classification problem with $n = 32$, $y \in \{1, 2, 3, \dots, n\}$, and two binary features, $x_1, x_2 \in \{0,1\}$. Suppose $P(Y=y) = 1/32$, $P(x_1 = 1 | Y = y) = y / 46$, $P(x_2 = 1 | Y = y) = y / 62$. Which class will naive Bayes classifier produce on a test item with $x_1 = 1$ and $x_2 = 0$?

Options:

1) 16
2) 26
3) **31**
4) 32

Q 2.2 Consider the problem of detecting if an email message contains a virus. Say we use four random variables to model this problem: Boolean (binary) class variable V indicates if the message contains a virus or not, and three Boolean feature variables: A,B,C. We decide to use a Naive Bayes Classifier to solve this problem so we create a Bayesian network with arcs from V to each of A,B,C. Their associated CPTs (Conditional Probability Table) are created from the following data: $P\{V=1\}$ = 0.92, $P\{A=1|V=1\}$ = 0.65, $P\{A=1|V=0\}$ = 0.9, $P\{B=1|V=1\}$ = 0.32, $P\{B=1|V=0\}$ = 0.78, $P\{C=1|V=1\}$ = 0.12, $P\{C=1|V=0\}$ = 0.94. Compute $P\{A=a,B=b,C=c\}$ for a,b,c = 1, 0, 1.

Options:

1) 0.0637
2) 0.0149
3) 0.0488
4) 0.0766

Q 2.2 Consider the problem of detecting if an email message contains a virus. Say we use four random variables to model this problem: Boolean (binary) class variable V indicates if the message contains a virus or not, and three Boolean feature variables: A,B,C. We decide to use a Naive Bayes Classifier to solve this problem so we create a Bayesian network with arcs from V to each of A,B,C. Their associated CPTs (Conditional Probability Table) are created from the following data: $P\{V=1\}$ = 0.92, $P\{A=1|V=1\}$ = 0.65, $P\{A=1|V=0\}$ = 0.9, $P\{B=1|V=1\}$ = 0.32, $P\{B=1|V=0\}$ = 0.78, $P\{C=1|V=1\}$ = 0.12, $P\{C=1|V=0\}$ = 0.94. Compute $P\{A=a,B=b,C=c\}$ for a,b,c = 1, 0, 1.

Options:

1) **0.0637**
2) 0.0149
3) 0.0488
4) 0.0766

[Block 3: Various Naïve Bayes Models]

Q 3.1 Consider the below dataset showing the result whether a person is pass or fail in the exam based on various factors. We want to classify an instance 'X' with Confident=Yes, Studied=Yes and Sick=No. Suppose the factors are independent to each other.

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	Yes	Pass

Options:

1. Pass
2. Fail

[Block 3: Various Naïve Bayes Models]

Q 3.1 Consider the below dataset showing the result whether a person is pass or fail in the exam based on various factors. We want to classify an instance 'X' with Confident=Yes, Studied=Yes and Sick=No. Suppose the factors are independent to each other.

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	Yes	Pass

Options:

1. **Pass**
2. Fail

Q 3.2 In a lake, there are 2 kinds of fish, salmon and tilapia. Their lengths are both in Normal distribution: P(length=x | salmon) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-3)^2}{2}\right)$, P(length=x | tilapia) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-6)^2}{2}\right)$. Their weights are also both in Normal distribution: P(w = x | salmon) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-9)^2}{2}\right)$, P(w = x | tilapia) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-5)^2}{2}\right)$.
When you catch a fish, the chance to be a salmon is 0.8. If now you catch a fish with length 5 and weight 7, which fish do you think it would be? (suppose weight and length are independent).

Options:

1. Salmon
2. Tilapia

Q 3.2 In a lake, there are 2 kinds of fish, salmon and tilapia. Their lengths are both in Normal distribution: P(length=x | salmon) = $\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-3)^2}{2}\right)$, P(length=x | tilapia) = $\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-6)^2}{2}\right)$. Their weights are also both in Normal distribution: P(w = x | salmon) = $\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-9)^2}{2}\right)$, P(w = x | tilapia) = $\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-5)^2}{2}\right)$.
When you catch a fish, the chance to be a salmon is 0.8. If now you catch a fish with length 5 and weight 7, which fish do you think it would be? (suppose weight and length are independent).

Options:

1. Salmon
2. **Tilapia**

Linear Algebra

1.1 Suppose x is a column vector. Is the equation $\| x \|_2^2 = x^{\mathrm{T}} x$ correct?

A. Yes

B. No

Answer: A

Linear Algebra

1.2 Which following statements are correct?

(1) For any square matrix X, $XI = IX = X$

(2) For any square matrix X, $XX^T v - \lambda v = (XX^T - \lambda I)v$

(3) If u_i is an eigenvector of square matrix A, then $Au_i = u_i$

A. (1)

B. (2)

C. (3)

D. (1)(2)

E. (1)(3)

F. (2)(3)

G. (1)(2)(3)

Answer: D. The eigenvalue is missing in option 3.

PCA Math

2.1 If v is a unit column vector, which one is correct :

A. $v^T v = 1$

B. $\| v \|_2 = 1$

C. both are correct

Answer: C

PCA Math

2.2 If $v_1, v_2, ..., v_d$ are principal components, which is correct :

A. $v_1^T v_2 = 0$

B. $v_2^T v_d = 0$

C. $v_1^T v_1 = 1$

D. the above options are all correct

Answer: D

PCA Dimension Reduction

3.1 Suppose we have a data matrix $X \in R^{n \times p}$ where n is the number of data points and p is the number of features. After applying PCA, we keep the first k eigenvectors with largest eigenvalues and project the data. What is the dimension of the projected data?

A. $n \times k$

B. $n \times p$

C. $k \times p$

Answer: A. After applying PCA, the data feature is reduced from p-dimension to k-dimension since we keep k principal components.

PCA Dimension Reduction

3.2 Consider the same setting as last question. We apply PCA on the data $X \in R^{n \times p}$ and keep the first k principal components with largest eigenvalues. What is the dimension of each principal component?

A. n

B. p

C. k

Answer: B. Each principal component has the same dimension as the feature of original data.

Q 1. Consider a biased coin toss. If P(heads) = 0.6, then P(tails) = ?

a) 0.4
b) 0.5
c) 0.6
d) 0.3

Q 1. Consider a biased coin toss. If P(heads) = 0.6, then P(tails) = ?

a) 0.4
b) 0.5
c) 0.6
d) 0.3

Solution:
P(tails) = 1 - P(heads) = 1 - 0.6 = 0.4

Q 2. In a presidential election, there are 3 candidates, A, B and C. Based on our polling analysis, we estimate that A has a 30 percent chance of winning the election, while B has a 40 percent chance of winning. What is the probability that either A or B win the election?

a) 50%
b) 70%
c) 40%
d) 100%

Q 2. In a presidential election, there are 3 candidates, A, B and C. Based on our polling analysis, we estimate that A has a 30 percent chance of winning the election, while B has a 40 percent chance of winning. What is the probability that either A or B win the election?

a) 50%
b) 70%
c) 40%
d) 100%

Solution:

P(A wins or B wins)
= P({A wins} U {B wins})
= P(A wins) + P(B wins) … (Note that A and B cannot win at the same time)
= 70 percent

Q 3. What is the probability of selecting a black card or a number 6 from a deck of 52 cards?

a) 26 / 52
b) 4 / 52
c) 30 / 52
d) 28 / 52

Q 3. What is the probability of selecting a black card or a number 6 from a deck of 52 cards?

a) 26 / 52

b) 4 / 52

c) 30 / 52

d) 28 / 52

Solution:

We need to find out P(card is black or card has number 6)

P(card is black) = 26/52 (either red or black)

P(card has number 6) = 4/52 (6 of clubs or 6 of diamonds or 6 of hearts or 6 of spades)

P(card is black and has number 6) = 2/52 (6 of clubs or 6 spades)

P({card is black} U {card has number 6})

= P(card is black) + P(card has number 6) - P({card is black} and {card has number 6})

= 26 / 52 + 4 / 52 - 2 / 52

= **28 / 52**

Q 4. Consider the joint probability distribution given below.
What is the probability that the temperature is hot given the weather is cloudy?

a) 40/365
b) 2/5
c) 3/5
d) 195/365

	weather = sunny	weather = cloudy	weather = rainy
temp= hot	150/365	40/365	5/365
temp = cold	50/365	60/365	60/365

Q 4. Consider the joint probability distribution given below.
What is the probability that the temperature is hot given the weather is cloudy?

a) 40/365
b) 2/5
c) 3/5
d) 195/365

	weather = sunny	weather = cloudy	weather = rainy
temp= hot	150/365	40/365	5/365
temp = cold	50/365	60/365	60/365

Solution:
P(temp = hot | weather = cloudy)
= P(temp = hot, weather = cloudy) / P(weather = cloudy)

From the table, P(temp = hot, weather = cloudy) = 40/365
P(weather = cloudy) = P(temp = hot, weather = cloudy) + P(temp = cold, weather = cloudy) = 100/365

Hence, P(temp = hot | weather = cloudy) = (40/365) / (100/365) = **2/5**

Q 5. Of a company's employees, 30% are women and 6% are married women. Suppose an employee is selected at random. If the employee selected is a woman, what is the probability that she is married?

a) 0.06
b) 0.3
c) 0.2
d) 0.24

Q 5. Of a company's employees, 30% are women and 6% are married women. Suppose an employee is selected at random. If the employee selected is a woman, what is the probability that she is married?

a) 0.06
b) 0.3
c) 0.2
d) 0.24

Solution:

P(Employee selected is Married | Employee selected is a woman)
= P(Employee selected is Married and Employee selected is a woman) / P(Employee selected is a woman)
= 0.06 / 0.30 = 0.2

Q 6. It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

a) 5 / 104
b) 95 / 100
c) 1 / 100
d) 1 / 2

Q 6. It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

a) 5 / 104
b) 95 / 100
c) 1 / 100
d) 1 / 2

Solution:

Define events

A = event that an email is detected as spam,

B = event that an email is spam,

Bc = event that an email is not spam.

We are given that, $P(B) = P(Bc) = 0.5$,

$P(A \mid B) = 0.99$,

$P(A \mid B^c) = 0.05$.

Hence by the Bayes's formula, we have

$P(B^c \mid A)$

$= P(A \mid B^c)*P(B^c) / (P(A \mid B)*P(B) + P(A \mid B^c)*P(B^c))$

$= 0.05 \times 0.5 / (0.05 \times 0.5 + 0.99 \times 0.5)$

= **5 / 104**

Q 7. If a fair coin is tossed three times, find the probability of getting 2 heads and a tail.

a) 1 / 8
b) 2 / 8
c) 3 / 8
d) 5 / 8

Q 7. If a fair coin is tossed three times, find the probability of getting 2 heads and a tail.

a) 1 / 8
b) 2 / 8
c) 3 / 8
d) 5 / 8

Solution:

P(H) = P(T) = 0.5
Each coin toss is independent of each other.
Hence, probability of getting 2 heads and a tail is given by

P(THH) + P(HTH) + P(HHT) = 3 * P(H) * P(H) * P(T) = **3 / 8**

Q 8. On a multiple choice test, problem A has 4 choices, while problem B has 3. Assume that each problem has 1 correct answer. What is the probability of guessing the correct answer to both of the problems?

a) $1/4 + 1/3$
b) $1/4 * 1/3$
c) $1/4*3/4 + 1/3*2/3$
d) None of the above

Q 8. On a multiple choice test, problem A has 4 choices, while problem B has 3. Assume that each problem has 1 correct answer. What is the probability of guessing the correct answer to both of the problems?

a) 1/4 + 1/3
b) 1/4 * 1/3
c) 1/4*3/4 + 1/3*2/3
d) None of the above

Solution: The two events are independent.

Q 9. Consider a fair die, and the following three events:

X = rolling any of {1, 2}

Y = rolling any of {2, 4, 6}

Z = rolling any of {1, 4}

In other words,

$P(X) = 1/3$, $P(Y) = 1/2$, $P(Z) = 1/3$.

Are events X and Y independent? Are events X and Y independent given event Z ?

a) Yes, Yes
b) No, No
c) Yes, No
d) No, Yes

Q 9. Consider a fair die, and the following three events:

X = rolling any of {1, 2}
Y = rolling any of {2, 4, 6}
Z = rolling any of {1, 4}

In other words,
$P(X) = 1/3$, $P(Y) = 1/2$, $P(Z) = 1/3$.
Are events X and Y independent? Are events X and Y independent given event Z ?

a) Yes, Yes
b) No, No
c) Yes, No
d) No, Yes

$$P(X, Y) = P(X)P(Y) = \frac{1}{6}$$

So, X and Y are independent.

$$P(X|Z) = \frac{1}{2}, P(Y|Z) = \frac{1}{2}, P(X, Y|Z) = P(\{2\}|Z) = 0$$

So, X and Y are not conditionally independent given event Z.

Q 10. Bag-of-Words

We have a piece a text.
"It was the best of times, it was the worst of times.”
Suppose our vocabulary is
["it", "was", "best", "of", "times", "worst"]

What is the bag of words representation of this text?

a) [2, 2, 1, 2, 2, 1]
b) [2, 2, 1, 2, 2, 1] / 6
c) [2, 2, 1, 2, 2, 1] / 10
d) [1, 1, 2, 1, 1, 2] / 10

Q 10. Bag-of-Words

We have a piece a text.
"It was the best of times, it was the worst of times.”
Suppose our vocabulary is
["it", "was", "best", "of", "times", "worst"]

What is the bag of words representation of this text?

a) [2, 2, 1, 2, 2, 1]
b) [2, 2, 1, 2, 2, 1] / 6
c) [2, 2, 1, 2, 2, 1] / 10
d) [1, 1, 2, 1, 1, 2] / 10

$$Z = \sum_{w} c(w, d) = 2 + 2 + 1 + 2 + 2 + 1 = 10$$

Q 11. tf-idf

We have a a corpus containing only the following documents.
Document ID 1: "A time to plant and a time to reap"
Document ID 2: "Time for you and time for me"
Document ID 3: "Time flies"
Given that the stemmed version of the word "flies" is the term "fly", what is the tf-idf of "fly" in document 3?

a) log(3)
b) 1/3 log(3)
c) log(2)
d) 1/2 log(2)

Q 11. tf-idf

We have a a corpus containing only the following documents.
Document ID 1: "A time to plant and a time to reap"
Document ID 2: "Time for you and time for me"
Document ID 3: "Time flies"
Given that the stemmed version of the word "flies" is the term "fly", what is the tf-idf of "fly" in document 3?

a) log(3)
b) 1/3 log(3)
c) log(2)
d) 1/2 log(2)

Solution:
tf = 1
idf = log(3)

Q 12. Given the following two document vectors, what is their cosine similarity?

$$v_a = \begin{bmatrix} 0.5 \\ 1 \\ 2 \end{bmatrix} \quad v_b = \begin{bmatrix} 2 \\ 1 \\ 0.5 \end{bmatrix}$$

a) 0.571

b) 0.99

c) 1.909

d) -0.99

Q 12. Given the following two document vectors, what is their cosine similarity?

$$v_a = \begin{bmatrix} 0.5 \\ 1 \\ 2 \end{bmatrix} \quad v_b = \begin{bmatrix} 2 \\ 1 \\ 0.5 \end{bmatrix}$$

a) 0.571

b) 0.99

c) 1.909

d) -0.99

Solution:

$$\cos\theta = \frac{v_a^T.v_b}{||v_a||_2 * ||v_b||_2} = \frac{v_a^T.v_b}{\sqrt{v_a^T.v_a} * \sqrt{v_b^T.v_b}} = \frac{0.5*2+1*1+2*0.5}{\sqrt{0.25+1+4} * \sqrt{4+1+0.25}} = \frac{3}{5.25} = 0.571$$

Q 13. Unigram

Suppose *the dog ran away* is our training corpus. What is P(ran away) if we use a unigram model?

a) 0
b) 1/2
c) 1/4
d) 1/16

Q 13. Unigram

Suppose *the dog ran away* is our training corpus. What is P(ran away) if we use a unigram model?

a) 0
b) 1/2
c) 1/4
d) 1/16

$$P(ran\ away) = P(ran)P(away) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$$

Q 14. Smoothing

Suppose *the dog ran away* is our training corpus. What is P(ran|dog) if we use a bigram model with Laplace Smoothing?

a) 1/4
b) 1
c) 2/5
d) 1/2

Q 14. Smoothing

Suppose *the dog ran away* is our training corpus. What is P(ran|dog) if we use a bigram model with Laplace Smoothing?

a) 1/4
b) 1
c) 2/5
d) 1/2

Solution:

P(ran|dog) = $\frac{|ran\ dog|+\alpha}{|dog|+\alpha * |V|} = \frac{1+1}{1+1*4} = \frac{2}{5}$

Q 15. Consider we choose number Uniformly from a set {1,2,3,4, …θ} (θ is an integer) with replacement. Suppose the numbers we choose are 2,5,7, then based on MLE, what value of θ do you estimate?

a) 3
b) 5
c) 7
d) 9

Q 15. Consider we choose number Uniformly from a set {1,2,3,4, …θ} (θ is an integer) with replacement. Suppose the numbers we choose are 2,5,7, then based on MLE, what value of θ do you estimate?

a) 3
b) 5
c) 7
d) 9

Solution:

'Uniform' => $P(x|\theta) = 1/\theta$ if $x <= \theta$;

$P(x|\theta) = 0$ if $x > \theta$.

$P(2|\theta) * P(5|\theta) * P(7|\theta) = (1/\theta)^3$ if $\theta >= 7$

$P(2|\theta) * P(5|\theta) * P(7|\theta) = 0$ otherwise

Thus, to make it largest, we choose $\theta = 7$.

Q 16. For the above example, if beforehand we know that $\theta \sim Pr$, that $Pr(\theta=5) = 1/2$, $Pr(\theta=8) = 1/6$, $Pr(\theta=9) = 1/2$. Then after we see the numbers we choose are 2,5,7, what value of θ do you think it is most likely to be?

a) 5
b) 7
c) 8
d) 9

Q 16. For the above example, if beforehand we know that $\theta \sim Pr$, that $Pr(\theta=5) = 1/2$, $Pr(\theta=8) = 1/6$, $Pr(\theta=9) = 1/2$. Then after we see the numbers we choose are 2,5,7, what value of θ do you think it is most likely to be?

a) 5
b) 7
c) 8
d) 9 ⟵

Solution:

$\theta = 5$: $P(2|5) * P(5|5) * P(7|5) * Pr(\theta=5) = 0$ since $P(7|5) = 0$
$\theta = 7$: $P(2|7) * P(5|7) * P(7|7) * Pr(\theta=7) = 0$ since $P(\theta=7) = 0$
$\theta = 8$: $P(2|8) * P(5|8) * P(7|8) * Pr(\theta=8) = (1/8)^3 * 1/6 = 0.000326$
$\theta = 9$: $P(2|9) * P(5|9) * P(7|9) * Pr(\theta=9) = (1/9)^3 * 1/2 = 0.000686$
Thus, we choose $\theta = 9$ to maximize

Q 17. Consider a classification problem with $n = 32$, $y \in \{1, 2, 3, \dots, n\}$, and two binary features, $x_1, x_2 \in \{0,1\}$. Suppose $P(Y=y) = 1/32$, $P(x_1 = 1 \mid Y = y) = y/46$, $P(x_2 = 1 \mid Y = y) = y/62$. Which class will naive Bayes classifier produce on a test item with $x_1 = 1$ and $x_2 = 0$?

a) 16

b) 26

c) 31

d) 32

Q 17. Consider a classification problem with $n = 32$, $y \in \{1, 2, 3, \dots, n\}$, and two binary features, $x_1, x_2 \in \{0,1\}$. Suppose $P(Y=y) = 1/32$, $P(x_1 = 1 | Y = y) = y/46$, $P(x_2 = 1 | Y = y) = y/62$. Which class will naive Bayes classifier produce on a test item with $x_1 = 1$ and $x_2 = 0$?

a) 16

b) 26

c) 31

d) 32

Solution:

$$P(y|x_1 = 1, x_2 = 0) \propto P(x_1 = 1, x_2 = 0 \mid y) * P(y) = P(x_1 = 1 \mid y)\, P(x_2 = 0 \mid y) * P(y)$$

$$= y/46 * (1-y/62) * 1/32$$

Maximize above formula => $y = 31$

Q 18. Consider the problem of detecting if an email message contains a virus. Say we use four random variables to model this problem: Boolean (binary) class variable V indicates if the message contains a virus or not, and three Boolean feature variables: A,B,C. We decide to use a Naive Bayes Classifier to solve this problem so we create a Bayesian network with arcs from V to each of A,B,C. Their associated CPTs (Conditional Probability Table) are created from the following data:

$P\{V=1\} = 0.92$,
$P\{A=1|V=1\} = 0.65$,
$P\{A=1|V=0\} = 0.9$,
$P\{B=1|V=1\} = 0.32$,
$P\{B=1|V=0\} = 0.78$,
$P\{C=1|V=1\} = 0.12$,
$P\{C=1|V=0\} = 0.94$.

Compute $P\{A=a,B=b,C=c\}$ for a,b,c = 1, 0, 1.

a) 0.0637
b) 0.0149
c) 0.0488
d) 0.0766

Q 18. Consider the problem of detecting if an email message contains a virus. Say we use four random variables to model this problem: Boolean (binary) class variable V indicates if the message contains a virus or not, and three Boolean feature variables: A,B,C. We decide to use a Naive Bayes Classifier to solve this problem so we create a Bayesian network with arcs from V to each of A,B,C. Their associated CPTs (Conditional Probability Table) are created from the following data:

$P\{V=1\} = 0.92$,
$P\{A=1|V=1\} = 0.65$,
$P\{A=1|V=0\} = 0.9$,
$P\{B=1|V=1\} = 0.32$,
$P\{B=1|V=0\} = 0.78$,
$P\{C=1|V=1\} = 0.12$,
$P\{C=1|V=0\} = 0.94$.

Compute $P\{A=a,B=b,C=c\}$ for a,b,c = 1, 0, 1.

a) 0.0637
b) 0.0149
c) 0.0488
d) 0.0766

Solution:

$P(A=1,B=0, C=1) = P(A=1,B=0, C=1, V = 1) + P(A=1,B=0,C=1,V=0)$
$= P(A=1|V=1)P(B=0|V=1)P(C=1|V=1)P(V=1) + P(A=1|V=0)P(B=0|V=0)P(C=1|V=0) P(V=0)$
$= 0.65 * 0.68 * 0.12 * 0.92 + 0.9 * 0.22 * 0.94 * 0.08 = 0.0637$

Q 19. Consider the below dataset showing the result whether a person is pass or fail in the exam based on various factors. We want to classify an instance 'X' with Confident=Yes, Studied=Yes and Sick=No. Suppose the factors are independent to each other.

a) Pass
b) Fail

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	Yes	Pass

Q 19. Consider the below dataset showing the result whether a person is pass or fail in the exam based on various factors. We want to classify an instance 'X' with Confident=Yes, Studied=Yes and Sick=No. Suppose the factors are independent to each other.

a) Pass
b) Fail

Confident	**Studied**	**Sick**	**Result**
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	Yes	Pass

Solution:

First we need to calculate the class probabilities i.e. P(Pass)=3/5 and P(Fail)=2/5

Now we need to calculate individual probability with respect to each features. For example,

P(Confident=Yes| Result=Pass) = 2/3

P(Studied=Yes| Result=Pass) = 2/3

P(Sick=No| Result=Pass) = 1/3

P(Confident=Yes| Result=Fail) = 1/2

P(Studied=Yes| Result=Fail) = 1/2

P(Sick=Yes| Result=Fail) = 1/2

P(Confident, Studied, not Sick | Result=Pass) * P(Result=Pass) = (2/3) * (2/3) * (1/3) * (3/5) = **0.089**

P(Confident, Studied, not Sick | Result=Fail) * P(Result=Fail) = (1/2) * (1/2) * (1/2) * (2/5) = **0.05**

P(Result=Pass | Confident, Studied, not Sick) > P(Result=Fail | Confident, Studied, not Sick)

⇨ Pass

Q 20. In a lake, there are 2 kinds of fish, salmon and tilapia. Their lengths are both in Normal distribution: P(length=x | salmon) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-3)^2}{2}\right)$, P(length=x | tilapia) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-6)^2}{2}\right)$

Their weights are also both in Normal distribution:

P(w = x | salmon) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-9)^2}{2}\right)$,

P(w = x | tilapia) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-5)^2}{2}\right)$.

When you catch a fish, the chance to be a salmon is 0.8. If now you catch a fish with length 5 and weight 7, which fish do you think it would be? (suppose weight and length are independent).

a) Salmon
b) Tilapia

Q 20. In a lake, there are 2 kinds of fish, salmon and tilapia. Their lengths are both in Normal distribution: P(length=x | salmon) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-3)^2}{2}\right)$, P(length=x | tilapia) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-6)^2}{2}\right)$

Their weights are also both in Normal distribution:

P(w = x | salmon) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-9)^2}{2}\right)$,

P(w = x | tilapia) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-5)^2}{2}\right)$.

When you catch a fish, the chance to be a salmon is 0.8. If now you catch a fish with length 5 and weight 7, which fish do you think it would be? (suppose weight and length are independent).

a) Salmon
b) Tilapia

Solution:

P(salmon | w = 7, l = 5) $\propto$ P(w=7 | salmon) * P(l = 5 | salmon) * P(salmon)

$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(7-9)^2}{2}\right) * \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(5-3)^2}{2}\right) * 0.8$

= 0.0023

P(tilapia | w = 7, l = 5) $\propto$ P(w=7 | tilapia) * P(l = 5 | tilapia) * P(tilapia)

$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(7-5)^2}{2}\right) * \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(5-6)^2}{2}\right) * 0.2$

= 0.0026

So it's more likely a tilapia.

Q21. Suppose x is a column vector. Is the equation $\| x \|_2^2 = x^T x$ correct?

a) Yes
b) No

Q21. Suppose x is a column vector. Is the equation $\| x \|_2^2 = x^T x$ correct?

a) Yes

b) No

Q22. Which following statements are correct? (I is the identity matrix)
(1) For any square matrix X, $XI = IX = X$
(2) For any square matrix X, $XX^T v - \lambda v = (XX^T - \lambda I)v$
(3) If u_i is an eigenvector of square matrix A, then $Au_i = u_i$

a) (1)
b) (2)
c) (3)
d) (1)(2)
e) (1)(3)
f) (2)(3)
g) (1)(2)(3)

Q22. Which following statements are correct? (I is the identity matrix)
(1) For any square matrix X, $XI = IX = X$
(2) For any square matrix X, $XX^T v - \lambda v = (XX^T - \lambda I)v$
(3) If u_i is an eigenvector of square matrix A, then $Au_i = u_i$

a) (1)
b) (2)
c) (3)
d) (1)(2)
e) (1)(3)
f) (2)(3)
g) (1)(2)(3)

Solution:
The eigenvalue is missing in option (3).

Q23. If v is a unit column vector, which one is correct?

a) $v^T v = 1$
b) $\| v \|_2 = 1$
c) Both are correct

Q23. If v is a unit column vector, which one is correct?

$a)$ $v^T v = 1$

$b)$ $\| v \|_2 = 1$

c) Both are correct

Q24. If v_1, v_2, ..., v_d are principal components, which one is correct?

a) $v_1^T v_2 = 0$
b) $v_2^T v_d = 0$
c) $v_1^T v_1 = 1$
d) All are correct

Q24. If $v_1, v_2, \ldots, v_d$ are principal components, which one is correct?

a) $v_1^T v_2 = 0$
b) $v_2^T v_d = 0$
c) $v_1^T v_1 = 1$
d) All are correct

Q25. Suppose we have a data matrix $X \in \mathbb{R}^{n \times p}$ where n is the number of data points and p is the number of features. After applying PCA, we keep the first k eigenvectors with largest eigenvalues and project the data. What is the dimension of the projected data?

a) $n \times k$
b) $n \times p$
c) $k \times p$

Q25. Suppose we have a data matrix $X \in \mathbb{R}^{n \times p}$ where n is the number of data points and p is the number of features. After applying PCA, we keep the first k eigenvectors with largest eigenvalues and project the data. What is the dimension of the projected data?

a) $n \times k$
b) $n \times p$
c) $k \times p$

Solution:
After applying PCA, the data feature is reduced from p-dimension to k-dimension since we keep k principal components.

Q26. Consider the same setting as the previous question. We apply PCA on the data $X \in \mathbb{R}^{n \times p}$ and keep the first k principal components with largest eigenvalues. What is the dimension of each principal component?

a) n

b) p

c) k

Q26. Consider the same setting as the previous question. We apply PCA on the data $X \in \mathbb{R}^{n \times p}$ and keep the first k principal components with largest eigenvalues. What is the dimension of each principal component?

$a)\ n$

$b)\ p$

$c)\ k$

Solution:
Each principal component has the same dimension as the feature of original data.