

EMATM0044 Introduction to AI

Coursework Part 2

Due date: Friday 7th May 13:00

Submit your answers in the form of a pdf document. You do not need to submit any code.

Question 1 4pts

Suppose you have a dataset with classes as in table 1. Suppose you fit a k -nearest neighbours classifier with uniform weighting to this dataset with $k = 1$ (you can assume that the datapoint itself is included in the k nearest neighbours - it is its own nearest neighbour).

- (i) (1 pt) What value of accuracy will you get? Briefly explain your answer.
- (ii) (1 pt) Now suppose you choose k equal to the number of datapoints. What value of accuracy will you get? Briefly explain your answer.
- (iii) (2 pt) Why are neither of these good values of k in general?

Class	Cats	Dogs	Cows
Num. Instances	50	30	20

Table 1: Animals

Question 2 16 pts

2(a) 2 pts

In a naive Bayes classifier, the assumption is made that each feature is independent of each other feature. Give an example of a pair of features that you might encounter in the real world for which the assumption of independence doesn't hold, and a pair of features for which it does hold (more or less). Explain your answers.

2(b) 12 pts

Suppose we have the following dataset describing three plant species:

	Species 1				
Altitude (100m)	2.20	2.04	1.95	1.79	1.91
Height (cm)	1.99	2.19	2.23	2.22	2.58

	Species 2				
Altitude (100m)	2.31	2.49	2.94	2.20	1.97
Height (cm)	2.53	2.20	1.88	2.68	2.93

	Species 3				
Altitude (100m)	1.02	1.54	0.82	2.93	0.76
Height (cm)	1.15	0.95	1.05	0.97	1.10

- (i) **(10 pts)** Build a Gaussian naive Bayes classifier using this dataset and classify the point (2.1, 2.3). Show your working.
- (ii) **(2 pts)** Do any of the classes fit the independence assumption? Do any of the classes violate the independence assumption? Explain your answer. (Hint: you may wish to plot the data)

2(c) 2 pts

Would k -means do a good job at clustering this data? Explain why or why not.

Question 3 10 pts

Suppose we initialize a Gaussian mixture model with k components as follows:

- We run k -means on the dataset to get initial values for the mean of each component.
- We set the covariance matrix of each component to a small multiple of the identity matrix.
- We set the weight of each component to $\frac{1}{k}$.

Explain why this initialization is equivalent to a k -means clustering. In particular, consider the membership of a datapoint in each cluster in k -means, and in each component in the mixture of Gaussians. Why are they (approximately) equal?

Question 4 15 pts

Your friend is coming to visit you in Bristol, and you have decided to walk to the city museum. Unfortunately you don't know your way to the museum that well, but you do know how to get to various different landmarks. You start planning a route based on the distances and heuristics given in the diagram in figure 1.

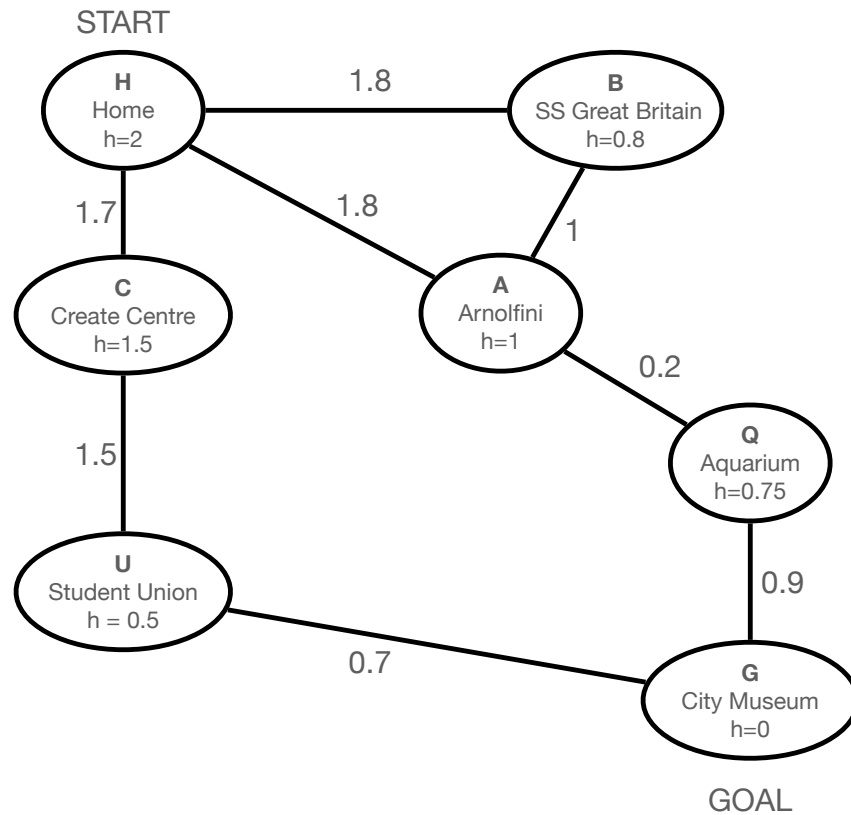


Figure 1: Bristol landmarks. Distances by road are given on the edges of the graph, and the heuristic values (h) are straight line distances between landmarks.

4(a) 6 pts

What order will you expand the nodes in for the following (break any ties alphabetically)?

- (i) (1 pt) Breadth-first search
- (ii) (1 pt) Depth-first search
- (iii) (1 pt) Best first greedy search
- (iv) (2 pt) Uniform cost search

1 pt Which, if any, of these techniques gives you the optimal route to the museum?

4(b) 2 pt

Is the heuristic:

1. Admissible?
2. Consistent?

Explain your answer.

4(c) 3 pt

Suppose the edge between the student union (node U) and the city museum (node G) now has a negative cost of -1 (because you particularly want to see the fountain on the way). Suppose you try to run uniform cost search on this graph (even though one edge is negative). Does uniform cost search still find the optimal path? If not, why not?

NB: do not simply answer that uniform cost search requires positive edge costs - please investigate what happens in this situation.

4(d) 4 pt

Now suppose the edge between U and G has a cost of 0.7 as before. You discover that there is a ferry between SS Great Britain (B) and the Aquarium (Q), adding an edge between B and Q at negative cost of -2 (your friend loves ferries). Suppose you try to run uniform cost search on this graph (even though one edge is negative). Does uniform cost search still find the optimal path? If not, why not?

NB: do not simply answer that uniform cost search requires positive edge costs - please investigate what happens in this situation.

Question 5 15 pts

Consider the 8-puzzle problem given the initial state shown in figure 2. Assume that the space moves up if possible, left if up is not possible, down if neither up nor left are possible, otherwise right (i.e. the order of priority of the moves is up, left, down, right).

5(a) 3 pts

Using breadth-first search, show the search tree that would be built down to level two, assuming the root of the tree, i.e. the initial state, is level zero.

1	3	4
8	2	5
	7	6

1	2	3
8		4
7	6	5

Figure 2: 8 puzzle initial state (left) and goal state (right)

5(b) 3 pts

Using depth-first search, show the search tree that would be built down to level three (stop after you have expanded one node that goes to level three).

5(c) 7 pts

Choose a heuristic function for the 8 puzzle and solve this instance using A* search.

5(d) 2 pts

On this problem, will breadth-first search find the optimal solution? Explain why, or why not.

Question 6 15 pts

You are on your way home, and you remember that you left a slice of cake out on the table. You wonder if it will be eaten (C): if your flatmate is home (F) then they may have eaten your cake. Also, the dog often gets loose (D), and if so, the dog may have eaten your cake. If the dog does get loose, it is likely to have dug a hole in the garden (H). This can be formulated as a Bayesian network as in figure 3. The probability tables are given below.

- C is a random variable specifying whether your cake is eaten, c indicates that is eaten, $\neg c$ that it is not eaten.
- D is a random variable specifying whether the dog is loose, d indicates the dog is loose, $\neg d$ indicates it is not loose.
- F is a random variable specifying whether your flatmate is home, f indicates your flatmate is home, $\neg f$ indicates they are not home.

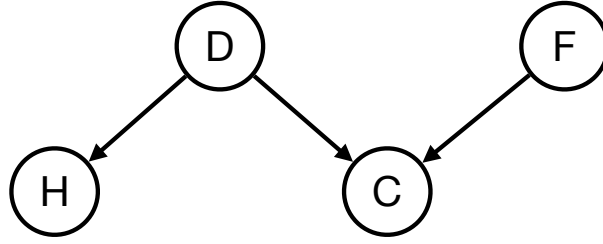


Figure 3: Bayesian network for the cake situation

- H is a random variable specifying whether there is a hole in the garden, h indicates there is a hole in the garden, $\neg h$ indicates there is not a hole in the garden.

D	$P(D)$	F	$P(F)$	D	H	$P(H D)$	D	F	C	$P(C D, F)$
d	0.3	f	0.4	d	h	0.5	d	f	c	1
$\neg d$	0.7	$\neg f$	0.6	d	$\neg h$	0.5	d	$\neg f$	c	0
				$\neg d$	h	0.1	d	$\neg f$	$\neg c$	0.8
				$\neg d$	$\neg h$	0.9	$\neg d$	f	c	0.2
							$\neg d$	f	$\neg c$	0.9
							$\neg d$	$\neg f$	c	0.1
							$\neg d$	$\neg f$	$\neg c$	0.1
										0.9

- 2 pts** What is the probability that there is a hole in the garden?
- 3 pts** What is the probability that the dog is loose, given that there is a hole in the garden?
- 4 pts** What is the probability that the dog is loose, given that your cake is eaten and your flatmate is home?
- 6 pts** What is the probability that your flatmate is home, given that your cake is eaten and there is a hole in the garden?

Question 7 25 pts

Consider an agent in a 2x3 gridworld. The **rewards** for the gridworld are given in table 2

2	-0.5	10 G
-0.5	-0.5	-0.5

Table 2: Rewards in the 2x3 gridworld

The top right cell (3, 2), marked G, is the **goal state**. Once the goal state is reached, the game is over. The agent starts in the bottom left cell (1, 1).

The **transition model** says that when the agent moves, it will go in the intended direction 80% of the time, and at right angles to the intended direction the rest of the time. So if the agent moves up, it will go up with probability 0.8, left with probability 0.1, and right with probability 0.1. Suppose the **discount factor** is 0.9.

7(a) 2 pts

We will use value iteration to calculate the utility of each state. Assume that at iteration 0, all cells have utility 0. After one iteration, the utility of each state is equal to its reward, as in table 3. Explain why this is the case.

2	-0.5	10 G
-0.5	-0.5	-0.5

Table 3: Utilities at iteration 1

7(b) 9 pts

Use value iteration to calculate the utility of each cell in the gridworld after two and three iterations, writing these utilities down each time. What is the optimal policy according to the utilities after three iterations? Does the agent reach the goal?

7(c) 3 pts

After convergence, the states have the utilities given in table 4. What is the optimal policy

17.34	14.38	10.0 G
14.38	12.25	10.13

Table 4: Utilities at iteration 90

according to the utilities after convergence? Does the agent reach the goal?

7(d) 11 pts

Suppose the discount factor is reduced to 0.01. What will the optimal policy be? Is the optimal policy different than when the discount factor is 0.9? Explain why, or why not.