Artificial Intelligence 1

Solutions

Mock Exam 2023

Note

Answer ALL questions. Each question will be marked out of 20. The paper will be marked out of 80, which will be rescaled to a mark out of 100.

Question 1 Clustering

(a) Use hierarchical agglomerative clustering with complete linkage to cluster a 1-dimensional dataset with the following points: 3, 7, 8, 11, 17, 25, 27.

Show your step-by-step calculation of how clusters are formed. If there is a tie, select the first pair from left to right. Draw the resulting dendrogram with heights on one side. Assume that we want to cluster this dataset into 3 clusters. Which 3 clusters would the dendrogram give us? **[10 marks]**

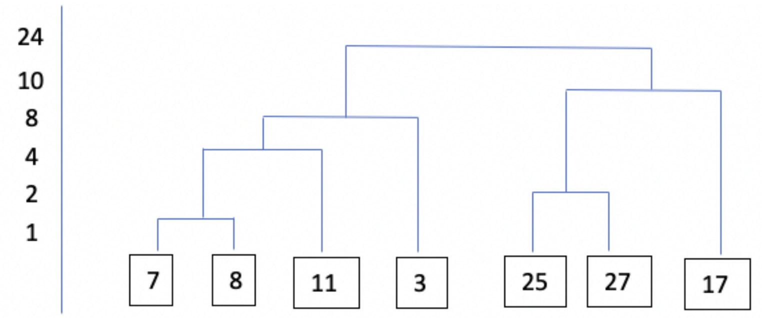
(b) Compute the per-cluster entropy and per-cluster purity of the confusion matrix given below.

Cluster	Entertainment	Financial	Foreign	Metro	National	Sports	Total
#1	1	1	0	11	4	676	693
#2	27	89	333	827	253	33	1562
#3	326	465	8	105	16	29	949
Total	354	555	341	943	273	738	3204

Note: The probability that a member of cluster i belongs to class j is $p_{i,j}$ =number of objects of class j in cluster i/ number of objects in cluster i. Then, entropy of ith cluster is $e_i = -\sum_{j=1}^{L} p_{i,j} \log_2 p_{i,j}$ with L denoting the number of classes. Purity of cluster i is given by $p_i = \max_j p_{i,j}$. **[10 marks]**

Model answer / LOs / Creativity:

(a)
- Combine 7 and 8 with distance 1.
- Combine 25 and 27 with distance 2.
- Combine (7,8) and 11 with distance 4.
- Combine (7,8,11) and 3 with distance 8.
- Combine (25, 27) and 17 with distance 10.
- Combine (3,7,8,11) and (17,25,27) with distance 24.



The dendrogram would give the following three clusters: (3, 7, 8, 11), (25, 27), 17.

(b) The probabilities can be computed as follows:

$p_{1,1} = 1/693$, $p_{1,2} = 1/693$, $p_{1,3} = 0$, $p_{1,4} = 11/693$, $p_{1,5} = 4/693$, $p_{1,6} = 676/693$
$p_{2,1} = 27/1562$, $p_{2,2} = 89/1562$, $p_{2,3} = 333/1562$, $p_{2,4} = 827/1562$, $p_{2,5} = 253/1562$, $p_{2,6} = 33/15$
$p_{3,1} = 326/949$, $p_{3,2} = 465/949$, $p_{3,3} = 8/949$, $p_{3,4} = 105/949$, $p_{3,5} = 16/949$, $p_{3,6} = 29/949$.

Then, purity of cluster 1 = 676/693, purity of cluster 2 = 827/1562, and purity of cluster 3 = 465/949. Finally, entropy of cluster 1 = 0.2, entropy of cluster 2 = 1.8407, entropy of cluster 3 = 1.6964.

Question 2 Supervised Learning

(a) The following pseudo-code represents one iteration through the training set for gradient descent applied to univariate second order polynomial regression.

```
cost = 0;
w0 = 0;
w1 = 0;
w2 = 0;
for j in size(trainingSet)
      f = w0 + w1*x(j) + w2*x(j)^2
      cost = cost + (y(j) - f)^2
      w0 = w0 - a*(f - y(j))
      w1 = w1 - a*(f - y(j))*x(j)
      w2 = w2 - a*(f - y(j))*x(j)^2
endfor
```

Assume that the value of the learning rate, a is 1.

Give the numerical values of 'w0', 'w1', 'w2' and 'cost' at the end of this pseudo-code for the following training set: $\{(1,1),(2,5)\}$. Show all your working. **[10 marks]**

(b) Consider a multivariate data set with 2 classes that are not linearly separable. Is it true that the classes will still be not linearly separable

- (i) if you remove one point from this data?
- (ii) if you remove one feature from this data?

In both cases, justify your answers in the following way: if your answer is yes, then explain why; if your answer is no then give a counter-example. **[10 marks]**

Model answer / LOs / Creativity:

(a) Iteration j=1:
w0=1, w1=1, w2=1, cost=1. Iteration j=2:
w0=-1, w1=-3, w2=-7, cost=5. These are the values at the end of the pseudo-code that the question was asking.

(b) (i) No: If you remove a point it can become linearly separable. For example, the following 2D labelled data set ((-1,-1), 0), ((1,1), 0), ((-1,1),1), ((1,-1),1) is not linearly separable, but after removing (any) one point from it it becomes linearly separable. (Draw the data for yourself to see.)

(c) (ii) Yes, it remains not linearly separable. That is, a weight vector with which a linear classifier would perfectly label this data does not exist. The reason is a follows: Removing a feature is equivalent to setting the component of the weight vector that multiplies that feature to 0. If this would make the classes linearly separable then this modified weight vector would perfectly label the original data set, which is a contradiction.

(a). Apply AI approaches to solve real-world problems. Develop, or work with, implementations of algorithms for AI problems (b). Question (a) is a step-by-step manual run of the algorithm. Question (b) is creative.

Question 3 Optimisation

(a) Explain what are constraints in optimisation problems, and how they are usually mathematically depicted in a problem formulation. **[10 marks]**

(b) Consider a regression task represented by a (potentially noisy) training set as follows:

$$D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(n)}, y^{(n)})\},$$

where, for any $i \in \{1, 2, \ldots, n\}$, $x^{(i)}$ and $y^{(i)}$ are real values.

Consider a linear regression model for this regression task. The weights of this model are stored in a 2-dimensional vector $\mathbf{w}$ of real values, and the output of the model for an input x is given by $h(x; \mathbf{w})$. Assume that one proposes to formulate the machine learning problem of learning the weights $\mathbf{w}$ to be adopted for the regression task as an optimisation problem as follows:

$$\text{minimize} \quad f(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - h(x^{(i)}; \mathbf{w}))^2$$

(i) Explain your understanding of what the function $f(\mathbf{w})$ is calculating.

Question 3 continued over the page

(ii) Consider that one wishes to solve the problem given by the proposed formulation using simulated annealing with the design below:

Representation: direct representation of the design variable. In other words, a 2-dimensional vector $\mathbf{w}$ of real values.

Algorithm 1: Initialisation Procedure.

Output: Candidate solution $\mathbf{w}$.

```
w = new vector of size 2;
w_1 = real value picked uniformly at random between 0 (inclusive) and 1 (inclusive);
w_2 = real value picked uniformly at random between 0 (inclusive) and 1 (inclusive);
return w
```

Algorithm 2: Neighbourhood Operator.

Input: Current solution $\mathbf{w}$.

Output: Neighbour $\mathbf{w}'$.

```
w' = copy of w;
i = value picked uniformly at random from the set {1, 2};
j = value picked uniformly at random from the set {-1, +1};
w'_i = w_i + j;
return w'
```

Discuss a key potential weakness of this simulated annealing design in the context of solving the problem given by the proposed formulation.

Note: assume that the proposed formulation truly represents the problem to be solved.

[10 marks]

Model answer / LOs / Creativity:

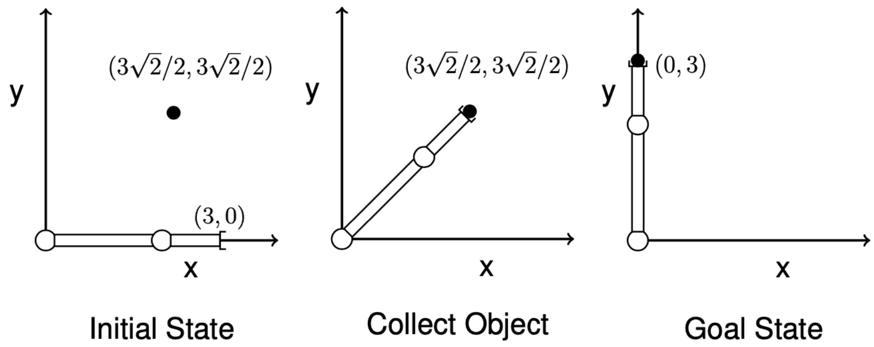
(a) Constraints are conditions that the design variable(s) must satisfy for the solution to be feasible. They are usually captured by functions that take the design variable(s) as input and output a numeric value. The constraints then specify the values that these functions are allowed to take for the solution to be feasible.

(b) (i) This function is computing the error (mean square error) of the neural network on the training set D, given weights $\mathbf{w}$.

(ii) The neighbourhood operator only allows to sum or subtract one from a given position of the weight vector. This value may be too large, potentially causing the algorithm to jump over the optimum in a given optimisation run.

Question 4 Search

A planar robot with two degrees of freedom consists of two links that can rotate around the two rotational joints. The planar robot is placed at the origin as shown in the image below (initial state). The first link has length 2, while the second link has length 1 so that the end effector (i.e., the end of the robotic arm that is used to manipulate objects) is placed at coordinates $(3, 0)$ in the initial state.



The goal of the robot is to collect an object placed at coordinates $(3\sqrt{2}/2, 3\sqrt{2}/2)$ and move this object to the position identified by coordinates $(0, 3)$, as shown in the Collect Object and Goal State figure above, respectively. This problem can be formulated as a search problem as follows:

Question 1 continued over the page

- Initial and goal states as shown in the figure above.
- Actions: you can rotate one of the links by 45° or −45°, and you can collect the object only if the end effector is placed above it.
- Nodes are identified by the coordinates of the end effector and by the information if the robot is holding the object. To calculate the coordinates, use the following equations (forward kinematics):

$$x = 2\cos(\theta_1) + \cos(\theta_1 + \theta_2), \quad y = 2\sin(\theta_1) + \sin(\theta_1 + \theta_2),$$

  where θ_1 and θ_2 are the angles of rotation of the first and second joint, respectively, and cos and sin are the cosine and sine functions.
- The cost of each action is equal to 1. Always avoid loopy paths.

To calculate the cosine and sine of a given angle, please refer to the table below.

angle	cosine	sine
0	1	0
45°	$\sqrt{2}/2$	$\sqrt{2}/2$
90°	0	1

(a) Consider the state space of the above problem. In a real-life context, the robotic arm can rotate its links by any specified value, even non-integer values, e.g., rotate by 22.5° or 1.6666°. What impact would this have on the execution of a breadth first search on this problem and, more generally, how would this relate to the formulation of a search problem? **Justify** your answer. **[10 marks]**

(b) Generate the breadth first tree until the goal node is found. **[10 marks]**

When choosing which node to expand in the frontier and all nodes are at the same depth, always expand the node corresponding to the action in the following order: collect object (only if above the object), rotate link 1 by 45°, rotate link 1 by −45°, rotate link 2 by 45° and rotate link 2 by −45°. **Important: we only consider rotations if both coordinates of the position of the end effector are positive.**

Write down the following:

- Search tree produced by breadth first search, indicating which nodes are in the frontier when the algorithm terminates.
- The solution retrieved by breadth first search and its cost.
- The order in which the nodes are visited by breadth first search.

Model answer / LOs / Creativity:

(a) The first consideration is that a breadth first search on this problem would not be complete as the branching factor is infinite. The second consideration is that this approach would violate the formalism of the formulation of a search problem, since the actions would not be finite.

(b) The solution to this part is the following:

- We expand the root node and obtain the following 2 children (the other two are not possible because the coordinates of the end effector would not be positive): $(3\sqrt{2}/2, 3\sqrt{2}/2)$ and $(2+\sqrt{2}/2, \sqrt{2}/2)$.

  We expand node $(3\sqrt{2}/2, 3\sqrt{2}/2)$ and add the following nodes to the frontier: $(3\sqrt{2}/2, 3\sqrt{2}/2)$ (collect the object), $(0, 3)$, $(\sqrt{2}, \sqrt{2}+1)$ and $(\sqrt{2}+1, \sqrt{2})$.

  We expand node $(2+\sqrt{2}/2, \sqrt{2}/2)$ and add the following node to the frontier (since all other nodes would be considered loopy paths):
  $(2, 1)$.

  Finally, we expand node $(3\sqrt{2}/2, 3\sqrt{2}/2)$ (holding the object) and add the following nodes to the frontier:
  $(0, 3)$ (holding the object), $(3, 0)$ (holding the object), $(\sqrt{2}, \sqrt{2}+1)$ (holding the object) and $(\sqrt{2}+1, \sqrt{2})$ (holding the object).

  Since we added the goal node to the frontier, we stop.

- The solution is: rotate link 1 by 45°, collect the object, rotate link 1 by 45°. The cost of this solution is 3.

- The order in which the nodes are visited is the following: $(3, 0)$, $(3\sqrt{2}/2, 3\sqrt{2}/2)$, $(2+\sqrt{2}/2, \sqrt{2}/2)$, $(3\sqrt{2}/2, 3\sqrt{2}/2)$ (holding the object). We do not add the goal node as this node is not visited, but only added to the frontier (stop condition for BFS).