UNIVERSITY OF BIRMINGHAM

School of Computer Science

AI 1 + AI&ML

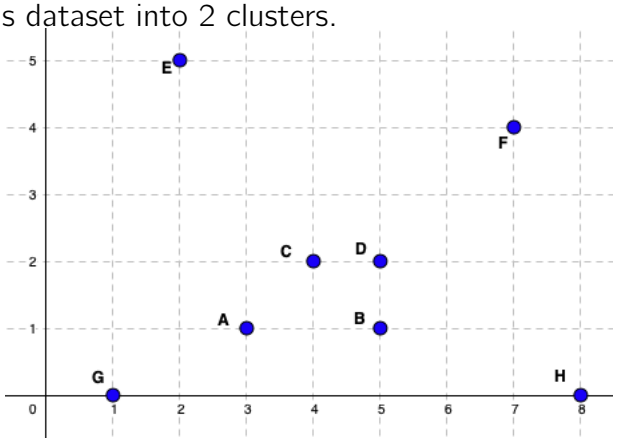
Mock Exam 2022

AI 1 + AI&ML

Exam paper

Question 1 Clustering

We have a dataset with 8 two-dimensional points: A = (3,1), B = (5,1), C = (4,2), D = (5,2), E = (2,5), F = (7,4), G = (1,0), H = (8,0). Use K-means and the Euclidean distance to cluster this dataset into 2 clusters.



(a) If the initial cluster centroids are at (1,2) and (1,4), what are the final clusters? **Show the step-by-step calculations.** **[5 marks]**

(b) If the initial cluster centroids are at (3,4) and (6,4), what are the final clusters? **Show the step-by-step calculations.** **[5 marks]**

(c) Between the two results from a) and b), which is a better grouping in terms of within-cluster variance? **Justify your answer.**

PS: Within-cluster variance is defined as the sum of squared Euclidean distance between each point and its cluster centroid: $\sum_{k=1}^{K} \sum_{x \in C_k} d\,(x, c_k)^2$, where C_k denotes the k-th cluster with centroid c_k, K is the total number of clusters, x is the data point, and d is the Euclidean distance between 2 given points. **[5 marks]**

Question 2 Supervised Learning

Consider the following labelled data set:

$$D = \{((-1,-1),0),((1,1),0),((-1,1),1),((1,-1),1)\}. \quad (1)$$

(a) Count the number of leave-one-out validation errors of Logistic Regression on this data set. **Describe briefly your steps and reasoning, and comment on what the results mean**. Hint: It is helpful to draw the data set. **[8 marks]**

(b) Would a Multi-Layer Perceptron (MLP) be likely to achieve a better leave-one-out error than Logistic Regression on this data? **Explain why or why not**. **[2 marks]**

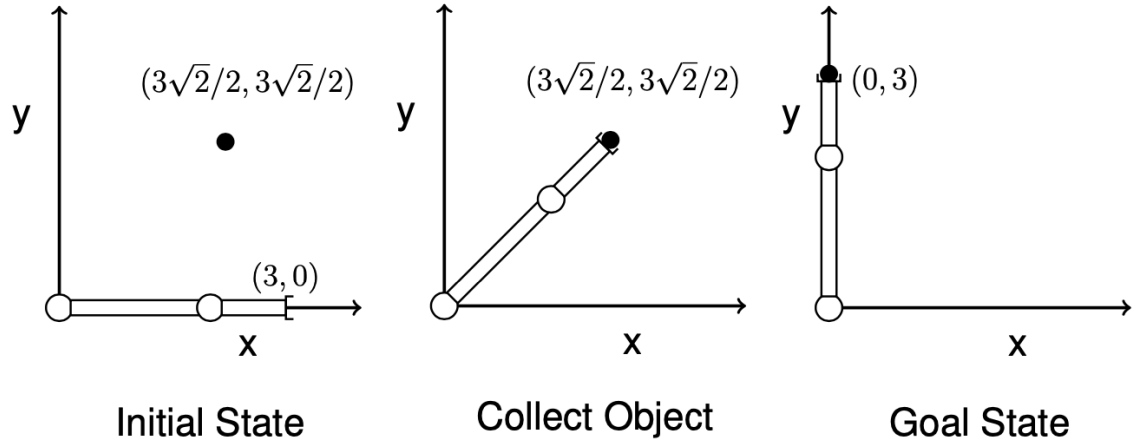
(c) Suppose you run a company and want to hire a new AI graduate. To select your your new hire, you offer an internship to 4 candidates and ask each to create a model that outperforms your existing predictor. You instruct them to use a model that has a hyperparameter called λ. You give them a labelled data set along with the code of your existing predictor. They all get to work, and after some time they come back to report to you.

- Candidate 1: "My predictor is better than yours - look at the training error!"
- Candidate 2: "My predictor is better than yours - look at the test error!"
- Candidate 3: "My predictor is better than yours - I used the hyperparameter value of $\lambda = 0.98674534286437898$, and look at the test error!"
- Candidate 4: "My predictor works better than yours - I selected λ by 10-fold cross-validation, and look at the test error!"

Which one of these candidates will you hire? **Justify your decision by commenting on each candidate**. **[5 marks]**

Question 3 Search Strategies

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 the coordinates $(0, 3)$, as shown in the Collect Object and Goal State images above, respectively. This problem can be formulated as a search problem as follows:

- Initial and goal states as shown in the images 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. When you expand the nodes, choose the next 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.**
- Nodes are identified by the coordinate of the end effector. To calculate the coordinate 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) Generate the breadth first tree until the goal node is found. Write down the steps to solve the problem, from the initial state to the goal state. When expanding the nodes, use the coordinates of the end effector to identify nodes. **[10 marks]**

(b) Based on the tree that you generated above, what is the solution for this problem? And what is the cost of this solution? **[5 marks]**

Question 4 Optimisation Problem Formulation

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 numbers.

Consider a neural network for this target regression task. The weights of this neural network are stored in a vector $\mathbf{w}$ and the output given by the neural network for an input x is given by $h(x; \mathbf{w})$. Assume that one decides to formulate the machine learning problem of learning the weights $\mathbf{w}$ for the target regression task as an optimisation problem as follows:

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

(a) Explain your understanding of what the function $f(\mathbf{w})$ is calculating. **[5 marks]**

(b) Discuss a key potential weakness of this problem formulation. **[5 marks]**

(c) Propose an adjustment of this problem formulation to overcome this weakness. **[5 marks]**