AI 1 + AI&ML

Solutions

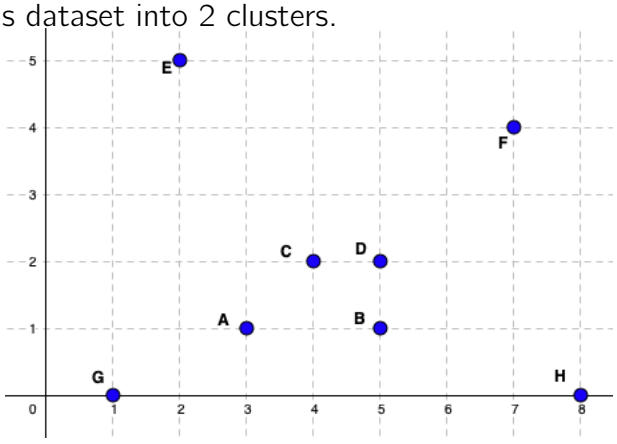
Mock Exam 2022

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]**

Model answer / LOs / Creativity:

(a)

- First, assign all samples to one of the two clusters, by measuring the Euclidean distance between each sample and c1/c2.
- A, B, C, D, G, H are closer to c1 centroid, so they form one cluster. The new centroid of this cluster c1 is: ((3+5+4+5+1+8)/6, (1+1+2+2+0+0)/6) = (13/3, 1).

- E, F are closer to c2 centroid, so they form the other cluster. The new centroid of this cluster c2 is: ((2+7)/2, (5+4)/2) = (4.5, 4.5).
- Repeat above steps: assign all samples to one of the two new clusters with centroids (13/3, 1) and (4.5, 4.5), by measuring the Euclidean distance between each sample and new c1/c2. We find the cluster assignments remain the same, i.e. A, B, C, D, G, H are closer to c1 centroid (13/3, 1) and E, F are closer to c2 centroid (4.5, 4.5).
- So, the final clusters are c1 = A, B, C, D, G, H with centroid (13/3, 1), c2 = E, F with centroid (4.5, 4.5). The clusters become stable.

(b) The calculation steps are similar to question a).

- First, assign all samples to one of the two clusters, by measuring the Euclidean distance between each sample and c1/c2.
- A, C, E, G are closer to c1 centroid, so they form one cluster. The new centroid of this cluster c1 is: ((3+4+2+1)/4, (1+2+5+0)/4) = (2.5, 2).
- B, D, F, H are closer to c2 centroid, so they form the other cluster. The new centroid of this cluster c2 is: ((5+5+7+8)/4, (1+2+4+0)/4) = (25/4, 7/4).
- Repeat above steps: assign all samples to one of the two new clusters with centroids (2.5, 2) and (25/4, 7/4), by measuring the Euclidean distance between each sample and new c1/c2. We find the cluster assignments remain the same, i.e. A, C, E, G are closer to c1 centroid (2.5, 2) and B, D, F, H are closer to c2 centroid (25/4, 7/4).
- So, the final clusters are c1 = A, C, E, G with centroid (2.5, 2), c2 = B, D, F, H with centroid (25/4, 7/4). The clusters become stable.

(c) The within-cluster variance of clustering from a) is:

$$d(A, c1)^2 + d(B, c1)^2 + d(C, c1)^2 + d(D, c1)^2 + d(G, c1)^2$$

$$+d(H, c1)^2 + d(E, c2)^2 + d(F, c2)^2 = 133/3$$

The within-cluster variance of clustering from b) is:

$$d(A, c1)^2 + d(C, c1)^2 + d(E, c1)^2 + d(G, c1)^2 + d(B, c2)^2$$

$$+d(D, c2)^2 + d(F, c2)^2 + d(H, c2)^2 = 69/2$$

The result from b) is better, because a smaller variance is preferred.

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]**

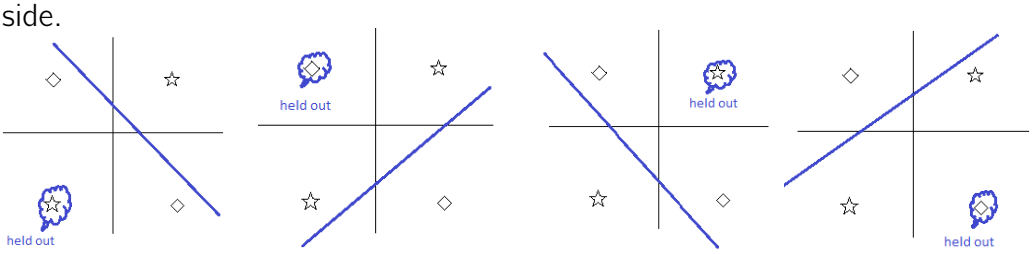
(c) Suppose you run a company and want to hire a new AI graduate. To select your your new hire, you offer an internship to 4 candidates and ask each to create a model that outperforms your existing predictor. You instruct them to use a model that has a hyperparameter called λ. You give them a labelled 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]**

Model answer / LOs / Creativity:

(a) The leave one out (LOO) error is obtained by cycling through the points, each time taking one labelled point out, training the LR on the remaining points, and evaluating its prediction on the held-out point. Notice that, by taking any one point out, in the remaining 3 points the 2 classes are linearly separable. Consequently, LR learns a separation boundary between them, and the held-out point ends up on the wrong side.



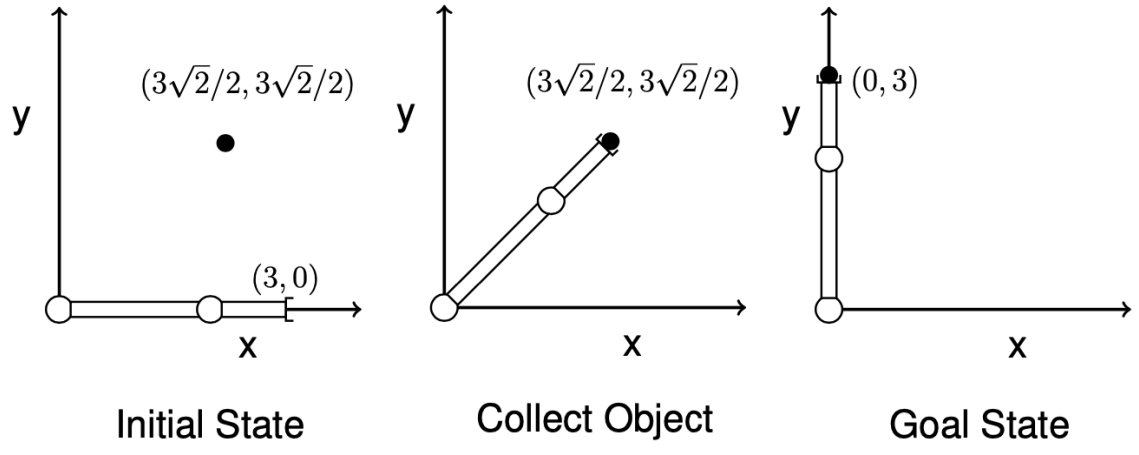
By symmetry, this happens for each of the 4 points, therefore the LOO is maximal. Comment: Notice that, upon holding out any one point, LR achieves 0 training error on the remaining points - but then it mis-classifies the held-out point. This example demonstrates the pitfall of small training sets as the classifier is misled.
Note: If you proceeded to solve the question differently, e.g. by running LR with LOO validation on your computer, and got the correct LOO count of 4 (equivalently, 100A coherent, meaningful interpretation is required for full marks. It does not have to be precisely the one above, but it must be coherent, true, and go beyond merely counting up the errors.

(b) It is very unlikely that an MLP would do better here, because any subset of the 3 points that it trains on lacks information about the non-linear nature of the problem. So, despite the fact the MLP has capacity to represent nonlinear boundaries, it will not be able to guess it from the data it sees at the training stage.

(c) C1: The training error is not indicative of future performance. C2: Didn't say how they set the hyperparameter, so you have reason to be suspicious. C3: Did not say how they set the hyperparameter. The value reported looks peculiar and might have been fine-tuned on the test set, so you have reason to be suspicious. C4: Appropriate methodology and appropriate reporting - Accept.

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 coordinates of the end effector. 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) 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]**

Model answer / LOs / Creativity:

(a) 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.

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

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]**

Model answer / LOs / Creativity:

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

(b) There may be different answers to this question. Machine learning aims not only at performing well on the given training set, but also on unseen examples. Formulating the optimisation problem in this way could lead to overfitting, which may result in poor performance on unseen examples. This is an example of key potential weakness of this problem formulation.

(c) Students may propose different solutions to their raised weakness. One possible solution would be to add a penalty term to encourage weights of small magnitude, e.g.:

$$f(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - h(x;\mathbf{w}))^2 + \lambda\sum_{j=1}^{m}|w_j|$$

where m is the size of the weights vector $\mathbf{w}$ and $\lambda > 0$ is a hyperparameter to tune how much emphasis we should place on the penalty term.