|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Course Code:** | | | **:** | **AI244AI** | **Date** | **:** | **27-08-2024** | | | |
| **Semester** | | | **:** | **IV** | **Time** | **:** | **09:30 – 11:30 PM** | | | |
| **Max Marks** | | | **:** | **10+50 = 60** | **Duration** | **:** | **120 mins** | | | |
| **Artificial Intelligence and Machine Learning** | | | | | | | | | | |
| **Improvement Test** | | | | | | | | | | |
| **Note: Answer all the Questions** | | | | | | | | | | |
| **SL. No** | | **Part A - Quiz** | | | | | | **M** | **BT** | **CO** |
| **1** | | What does the term "search space" refer to in the context of heuristic search? | | | | | | **1** | **L1** | **CO1** |
| **2** | | In genetic algorithms, the process of introducing random changes to a solution is known as \_\_\_\_\_\_\_\_\_\_. | | | | | | **1** | **L1** | **CO1** |
| **3** | | Differentiate between Greedy Best-First Search and A\* Search. | | | | | | **1** | **L2** | **CO1** |
| **4** | | Identify the key difference between hill-climbing and local-beam search. | | | | | | **2** | **L2** | **CO1** |
| **5** | | How can K-means clustering be used for outlier detection? | | | | | | **1** | **L2** | **CO1** |
| **6** | | How is the optimal number of clusters typically determined in K-means clustering? | | | | | | **1** | **L1** | **CO1** |
| **7** | | What happens if the number of specified clusters in K-means clustering is too small? | | | | | | **1** | **L3** | **CO1** |
| **8** | | In K-means clustering, what are the various techniques that can be used to address the sensitivity to the initial placement of cluster centroids? | | | | | | **1** | **L3** | **CO1** |
| **9** | | In K-means clustering, what is the purpose of the silhouette score? | | | | | | **1** | **L1** | **CO1** |
|  | | **Part B - Test** | | | | | |  |  |  |
| **1** | | Consider a graph with nodes A, B, C, D, E, and F, where the start node is A and the goal node is F. The edges and heuristic values (estimated cost to the goal) for each node are as follows:     * Using the A\* search algorithm, find the optimal path from A to F. * Show the steps involved. | | | | | | 10 | L4 | CO2 |
| **2** | | Consider a minimax game tree with a depth of 3. The max player is at the root, and the min players are at depth 1 and 3. The leaf nodes have the following values: {3, 5, 6, 9, 1, 2, 0, -1}. Use alpha-beta pruning to determine the optimal move for the max player. Show the steps. | | | | | | 10 | L4 | CO2 |
| **3** | **a** | How do genetic algorithms operate, and in what ways do they replicate the mechanisms of natural evolution? | | | | | | 05 | L2 | CO1 |
| **b** | Give a detailed explanation of the different clustering methods, along with diagrams to illustrate each one. | | | | | | 05 | L2 | CO2 |
| **4** | **a** | Apply K (=3)-Means algorithm over the data (1,3), (2,2), (5,8), (8,5), (3,9), (10,7), (3,3), (9,4), (3,7) up to two iterations and show the clusters. Initially choose Cluster 1 = (3,3), Cluster 2 = (3,7), Cluster 3 = (9,4) as initial centroids. | | | | | | 05 | L3 | CO1 |
|  | **b** | How can cohesion and separation metrics be used to evaluate the quality of clusters in an unsupervised learning context? | | | | | | 05 | L4 | CO1 |
| **5** |  | You are given with 10 points in the Cartesian coordinate system and information is to make 2 clusters. First, we will randomly choose 2 medoids from the given data. Let us consider M1 = (3, 4), M2 = (7, 3) as the centroids of the initial clusters**. (up to 3 iterations). 2nd iteration:** M1 = (3, 4), M2 = (7, 4) as the centroids. **3rd iteration:** M1 = (3, 4), M2 = (6, 4) as the centroids.   |  |  |  |  | | --- | --- | --- | --- | | Point | Coordinates | Point | Coordinates | | A1 | (2, 6) | A6 | (7, 3) | | A2 | (3, 8) | A7 | (7,4) | | A3 | (4, 7) | A8 | (8, 5) | | A4 | (6, 2) | A9 | (7, 6) | | A5 | (6, 4) | A10 | (3, 4) |  * Calculate the distance between each data point and the medoids using the Manhattan distance. * Calculate the cost for each cluster. * Examine whether data points have changed in the clusters after changing the medoids. | | | | | | 10 | L3 | CO2 |

**M-Marks, BT-Blooms Taxonomy Levels, CO-Course Outcomes**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Marks Distribution** | **Particulars** | **CO1** | **CO2** | **CO3** | **CO4** | **CO5** | **L1** | **L2** | **L3** | **L4** | **L5** | **L6** |
| **Max Marks CIE** | 25 | 35 | - | - | - | 04 | 14 | 17 | 25 | - | - |

|  |  |
| --- | --- |
| **Course Outcomes: After completing the course, the students will be able to:-** | |
| **CO1** | Explain and apply AI & ML algorithms to address various requirements of real-world problems. |
| **CO2** | Design and develop AI and ML solutions to benefit society, science, and industry. |
| **CO3** | Use modern tools to create AI and ML solutions. |
| **CO4** | Demonstrate effective communication through team presentations and reports to analyze the impact of AI and ML solutions on society and nature. |
| **CO5** | Conduct Performance evaluation, modeling, and validation of AI and ML solutions benefitting lifelong learning. |

Scheme and Solutions

|  |  |  |
| --- | --- | --- |
| Q.No | Scheme and Solutions **(Quiz)** | Marks |
| 1 | The search space refers to the set of all possible states or configurations that can be explored by the search algorithm to find a solution. | 1 |
| 2 | Mutation | 1 |
| 3 | Greedy Best-First Search uses only the heuristic function h(n) to estimate the cost to the goal, while A\* Search uses both the heuristic h(n) and the cost to reach the node g(n), i.e., f(n)=g(n)+h(n). | 1 |
| 4 | Hill-climbing explores a single current state to find the best neighbouring state, whereas local-beam search keeps track of multiple states simultaneously, expanding all of them in parallel. | 2 |
| 5 | By identifying data points that are far from their cluster centroids | 1 |
| 6 | By employing an elbow plot or silhouette analysis | 1 |
| 7 | The resulting clusters will be too broad and may not capture the underlying structure of the data | 1 |
| 8 | K-means++ and running the algorithm multiple times with different initializations | 1 |
| 9 | To evaluate the quality of the clustering solution | 1 |
|  | **PART B** |  |
| 1 | 1. **Initialization:**    * Start with node A. Cost to reach A: g(A) = 0. Estimated total cost from A: f(A) = g(A) + h(A) = 0 + 7 = 7. Open list = {A}. Closed list = {}. 2. **Step 1: Expand A:**    * Consider neighbors B and C. Cost to reach B: g(B) = g(A) + cost(A, B) = 0 + 4 = 4.    * Estimated total cost from B: f(B) = g(B) + h(B) = 4 + 6 = 10.    * Cost to reach C: g(C) = g(A) + cost(A, C) = 0 + 2 = 2.    * Estimated total cost from C: f(C) = g(C) + h(C) = 2 + 4 = 6.    * Open list = {B (f=10), C (f=6)}.Closed list = {A}. 3. **Step 2: Expand C (lowest f-value):**    * Consider neighbors D and E. Cost to reach D: g(D) = g(C) + cost(C, D) = 2 + 8 = 10.    * Estimated total cost from D: f(D) = g(D) + h(D) = 10 + 2 = 12.    * Cost to reach E: g(E) = g(C) + cost(C, E) = 2 + 10 = 12.    * Estimated total cost from E: f(E) = g(E) + h(E) = 12 + 1 = 13.    * Open list = {B (f=10), D (f=12), E (f=13)}.Closed list = {A, C}. 4. **Step 3: Expand B (next lowest f-value):**    * Consider neighbor D. Cost to reach D via B: g(D) = g(B) + cost(B, D) = 4 + 5 = 9 (better path).    * Update f(D) = g(D) + h(D) = 9 + 2 = 11. Open list = {D (f=11), E (f=13)}.    * Closed list = {A, C, B}. 5. **Step 4: Expand D (lowest f-value):**    * Consider neighbors E and F. Cost to reach E via D: g(E) = g(D) + cost(D, E) = 9 + 2 = 11. Update f(E) = g(E) + h(E) = 11 + 1 = 12 (better path).    * Cost to reach F: g(F) = g(D) + cost(D, F) = 9 + 6 = 15. Estimated total cost from F: f(F) = g(F) + h(F) = 15 + 0 = 15. Open list = {E (f=12), F (f=15)}.    * Closed list = {A, C, B, D}. 6. **Step 5: Expand E:**    * Consider neighbor F. Cost to reach F via E: g(F) = g(E) + cost(E, F) = 11 + 3 = 14.    * Update f(F) = g(F) + h(F) = 14 + 0 = 14 (better path).    * Open list = {F (f=14)}. Closed list = {A, C, B, D, E}. 7. **Step 6: Expand F:**    * Goal reached at node F. Optimal path = A → B → D → E → F.    * Total cost = 14.   **Final Answer:** The optimal path from A to F is A → B → D → E → F with a total cost of 14. | 10 |
| 2 | **Solution:**   1. **Initialization:**    * Start at the root node (Max player).Alpha (α) = -∞, Beta (β) = ∞. 2. **Step 1: Evaluate leftmost subtree:**    * First Min node: Evaluate left branch. Values are 3 and 5. Min value = 3.    * Update α = 3. Prune the rest of this subtree. 3. **Step 2: Evaluate second Min node:**    * Second Min node: Evaluate left branch. Values are 6 and 9.    * Min value = 6. Update α = max(3, 6) = 6. Prune the rest of this subtree. 4. **Step 3: Evaluate third Min node:**    * Third Min node: Evaluate left branch. Values are 1 and 2.    * Min value = 1. Update β = min(∞, 1) = 1.    * Prune this subtree as it's already lower than α. 5. **Step 4: Evaluate rightmost subtree:**    * Fourth Min node: Evaluate left branch.    * Values are 0 and -1. Min value = -1. Update β = min(1, -1) = -1.    * Prune this subtree. 6. **Optimal Decision:**    * Compare α values from all the subtrees. The highest α = 6.    * Max player should choose the branch leading to a value of 6.   **Final Answer:** The optimal move for the max player is to select the branch with a value of 6 after alpha-beta pruning. | 10 |
| 3a. | Genetic algorithms (GAs) are a type of optimization algorithm inspired by the process of natural evolution. They simulate evolution by iteratively improving a population of candidate solutions. The main operators in GAs are selection (choosing the fittest individuals for reproduction), crossover (combining parts of two individuals to produce offspring), and mutation (randomly altering parts of an individual to introduce diversity). These operators help maintain genetic diversity in the population, allowing the algorithm to explore a broader search space and potentially find better solutions. | 05 |
| 3b. |  | 05 |
| 4a. |  | 05 |
| 4b. | In general, we can consider expressing overall cluster validity for a set of K clusters as a weighted sum of the validity of individual clusters**,**    **Graph-Based View of Cohesion and Separation:** For graph-based clusters, the cohesion of a cluster can be defined as the sum of the weights of the links in the proximity graph that connect points within the cluster. Likewise, the separation between two clusters can be measured by the sum of the weights of the links from points in one cluster to points in the other cluster.    **Prototype-Based View of Cohesion and Separation:** For prototype-based clusters, the cohesion of a cluster can be defined as the sum of the proximities with respect to the prototype (centroid or medoid) of the cluster. Similarly, the separation between two clusters can be measured by the proximity of the two cluster prototypes. | 05 |
| 5 | After assigning clusters, we will calculate the cost for each cluster and find their sum. The cost is nothing but the sum of distances of all the data points from the medoid of the cluster they belong to.  Hence, the cost for the current cluster will be 3+4+4+2+2+0+1+3+3+0=22. Iteration 2: Now, we will select another non-medoid point (7, 4) and make it a temporary medoid for the second cluster. Hence, M1 = (3, 4), M2 = (7, 4) Now, let us calculate the distance between all the data points and the current medoids.    Now, let us again calculate the cost for each cluster and find their sum. The total cost this time will be 3+4+4+3+1+1+0+2+2+0=20.  Here, the current cost is less than the cost calculated in the previous iteration. Hence, we will make the swap permanent and make (7,4) the medoid for cluster 2. If the cost this time was greater than the previous cost i.e. 22, we would have to revert the change. New medoids after this iteration are (3, 4) and (7, 4) with no change in the clusters.  Iteration 3: | 10 |