**AI&ML-Theory Exam on Unsupervised Learning**

**Total Marks: 100**

1. Perform K-Means clustering for the dataset **X** and Fill up the following table while estimating Labels **L** and Centroids **M** for each iteration. Iterate till convergence for the specified number of maximum iterations. Report the final labels and the cluster centroids. **[20]**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| # | **X** | | | | | | | | | | **M1** | **M2** | **M3** |
|  | -1 | 27 | 31 | 2 | 59 | 3 | 61 | 34 | 0 | 12 |  |  |  |
| **L0** | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 1 |  |  |  |
| **L1** |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **L2** |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **L3** |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **L4** |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **L5** |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **L6** |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **L7** |  |  |  |  |  |  |  |  |  |  |  |  |  |

1. Consider the following dataset

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **-1** | **27** | **31** | **2** | **59** | **3** | **61** | **34** | **0** | **12** |

Suggest a Threshold **T** for this dataset using Otsu’s method. Start iterations with **T0 = 3**. **[20]**

|  |  |
| --- | --- |
| (a) **T** = 27.16 | (b) **T** = 16.89 |
| (c) **T** = 12.89 | (d) **T** = 22.8 |

1. Consider the problem of incremental mean and standard deviation update for streaming data problems. For a certain problem, the respective values of mean and standard deviation from the previous instant are ***m(t-1) = 10*** and ***s(t-1) = 2***. The data at present instant is ***x(t) = 7***. Estimate the updated values of mean ***m(t)*** and standard deviation ***s(t)***. The update rate is ***r = 0.05***. **[20]**

|  |  |
| --- | --- |
| (a) ***m(t)*** = 7.16; ***s(t)*** = 3.11 | (b) ***m(t)*** = 7.16; ***s(t)*** = 2.56 |
| (c) ***m(t)*** = 9.85; ***s(t)*** = 2.06 | (d) ***m(t)*** = 8.22; ***s(t)*** = 3.11 |

1. A dataset containing **1,10,000** points is to be subjected to hierarchical K-Means clustering with **K=2**.Wewillconstruct a fully grown tree and terminate with at least **100** points at each leaf node. The root node depth is considered to be ZERO. What value of maximum depth (**dmax**) shall we set for this divisive clustering tree. **[20]**

|  |  |
| --- | --- |
| (a) ***dmax*** = 20 | (b) ***dmax*** = 10 |
| (c) ***dmax*** = 5 | (d) ***dmax*** = 8 |

1. Consider the following dataset ***D*** which is subjected to mean-shift clustering through mean-shift iterations using an Epanechnikov kernel of band-width . Let, the cluster mode value at the *k*th iteration be . Evaluate the mode value at the (*k+1*)th iteration. **[20]**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
| 4.2 | 2.5 | 4.7 | 3.4 | 2.9 | 4.3 | 6.0 | 2.6 | 3.1 | 1.9 |
| 1.7 | 0.4 | 2.0 | 4.6 | 1.5 | 0.9 | 2.8 | 2.2 | 4.8 | 3.9 |

|  |  |
| --- | --- |
| (a) | (b) |
| (c) | (d) |