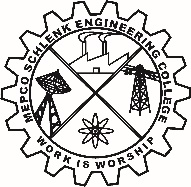
**MEPCO SCHLENK ENGINEERING COLLEGE (Autonomous), SIVAKASI**

**Department of Computer Science and Engineering**

**ASSIGNMENT – 3**

**Class: 4th Year B.E. CSE**

|  |  |
| --- | --- |
| **Sub. Code & Name: 19CSC17 –Machine Learning Using R Total Marks: 100** | |
| **Date of Announcement: 06-10-2022** | **Date of Submission: 26-10-2022** |

|  |  |  |  |
| --- | --- | --- | --- |
| **CL, CO** | **#** | **Question(s)** | **Mark** |
| 1. A, CO4 | 1. | Create and Read the file (cars.csv).  ## $ origin : chr "usa" "usa" "usa" "usa" ...  ## $ price : num 4099 4749 3799 4816 7827 ...  ## $ mileage : num 8.8 6.8 8.8 8 6 7.2 10.4 8 6.4 7.6 ...  ## $ repair : num 3 3 3 3 4 3 3 3 3 3 ...  ## $ headspace : num 6.25 7.5 7.5 11.25 10 ...  ## $ trunkspace : num 308 308 336 448 560 588 280 448 476 364 ...  ## $ weight : num 1318 1508 1188 1462 1836 ...  ## $ length : num 465 432 420 490 555 ...  ## $ turningcircle: num 12.2 12.2 10.7 12.2 13.1 ...  ## $ gear\_ratio : num 3.58 2.53 3.08 2.93 2.41 2.73 2.87 2.93 2.93 3.08 ...  The variable for origin (US, versus non-US) is a factor variable. We cannot calculate distances from a factor variable. Because we want to include it anyway, we have to make it a dummy (0/1) variable.  Normalize the data. Determine the number of clusters using the (graphical) method described above. Determine the clustering, and add the cluster to the data set. Describe the clusters in terms of all variables used in the clustering. Characterize (label) the clusters. Repeat the exercise with more or fewer clusters, and decide if the new solutions are better than the original solution. | (15) |
| 1. A, CO4 | 2. | Perform K medoids clustering approach by using the USArrests dataset built into R, which contains the number of arrests per 100,000 residents in each U.S. state in 1973 for Murder, Assault, and Rape along with the percentage of the population in each state living in urban areas, UrbanPop. Perform the following  Remove any rows with missing values  Scale each variable in the dataset to have a mean of 0 and a standard deviation of 1  Murder Assault UrbanPop Rape  Alabama 1.24256408 0.7828393 -0.5209066 -0.003416473  Alaska 0.50786248 1.1068225 -1.2117642 2.484202941  Arizona 0.07163341 1.4788032 0.9989801 1.042878388  Arkansas 0.23234938 0.2308680 -1.0735927 -0.184916602  California 0.27826823 1.2628144 1.7589234 2.067820292  Colorado 0.02571456 0.3988593 0.8608085 1.864967207 | (15) |
| 1. A, CO4 | 3. | Apply hierarchical clustering on the seeds dataset. This dataset consists of measurements of geometrical properties of kernels belonging to three different varieties of wheat: Kama, Rosa and Canadian. It has variables which describe the properties of seeds like area, perimeter, asymmetry coefficient etc. There are 70 observations for each variety of wheat.  Dataset Description: [https://archive.ics.uci.edu/ml/datasets/seeds#](https://archive.ics.uci.edu/ml/datasets/seeds)  Dataset download link: https://archive.ics.uci.edu/ml/machine-learning-databases/00236/ | (15) |
| 1. A, CO4 | 4. | Perform stock market analysis using Density based clustering approach.  Dataset description: It is a basic data about the customers going to the supermarket mall. This can be used for customer segmentation. There are 200 observations (customers) and no missing data. It consists of four columns ie. measured attributes: ​   * CustomerID is the customer identification number. * Gender is Female and Male. * Age is the age of customers. * Annual Income (k) is the annual income of clients in thousands of dollars. * Spending Score (1-100) is the spending score assigned by the shopping center according to the customer's purchasing behavior   $ CustomerID 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 1...  $ Gender Male, Male, Female, Female, Female, Female, ...  $ Age 19, 21, 20, 23, 31, 22, 35, 23, 64, 30, 67, ...  $ Annual.Income..k.. 15, 15, 16, 16, 17, 17, 18, 18, 19, 19, 19, ...  $ Spending.Score..1.100. 39, 81, 6, 77, 40, 76, 6, 94, 3, 72, 14, 99,... | (15) |
| 1. A, CO5 | 5. | Consider an input matrix of size 5\*5 and filter of size 3\*3. Apply the convolution with zero padding as 1 and find the resultant matrix for the following:  *Input Matrix:*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | 1 | 2 | 4 | 7 | 8 | | 8 | 5 | 3 | 34 | 9 | | 13 | 12 | 78 | 33 | 67 | | 66 | 31 | 11 | 88 | 53 | | 99 | 10 | 15 | 18 | 19 |   *Filter or Kernel:*   |  |  |  | | --- | --- | --- | | 1 | 0 | -1 | | 1 | 1 | 0 | | -1 | 0 | 0 | | (10) |
| 1. A, CO5 | 6. | Consider an input matrix of size 7\*7 and filter of size 3\*3. Apply the convolution with stride as 2 and find the resultant matrix for the following:  *Input Matrix:*   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 1 | 2 | 4 | 7 | 4 | 5 | 8 | | 8 | 5 | 3 | 34 | 12 | 67 | 9 | | 13 | 12 | 78 | 33 | 62 | 87 | 67 | | 66 | 31 | 11 | 88 | 18 | 90 | 53 | | 99 | 10 | 15 | 18 | 27 | 26 | 19 | | 1 | 3 | 5 | 7 | 9 | 11 | 13 | | 2 | 56 | 78 | 35 | 34 | 12 | 3 |   *Filter or Kernel:*   |  |  |  | | --- | --- | --- | | 1 | 0 | -1 | | 1 | 1 | 0 | | -1 | 0 | 1 | | (10) |
| A, CO5 | 7. | Consider the feature map produced from convolution layer is of size 5\*5 with matrix as follows. Apply the min pooling operation with filter of size 2\*2 with default stride:   |  |  |  |  |  | | --- | --- | --- | --- | --- | | 1 | 2 | 4 | 7 | 8 | | 8 | 5 | 3 | 34 | 9 | | 13 | 12 | 78 | 33 | 67 | | 66 | 31 | 11 | 88 | 53 | | 99 | 10 | 15 | 18 | 19 | | (10) |
| A,CO5 | 8. | Consider the feature map produced from convolution layer is of size 5\*5 with matrix as follows. Apply the average pooling operation with filter of size 3\*3 with stride=2:   |  |  |  |  |  | | --- | --- | --- | --- | --- | | -34 | 12 | 5 | -3 | 32 | | 7 | 15 | 3 | 12 | 11 | | -5 | 9 | 7 | 0 | -5 | | 11 | 9 | 8 | 72 | 3 | | 56 | 34 | 12 | 89 | -34 | | (10) |

***Faculty Signature HOD/CSE***