Indian Institute of Space Science and Technology – Thiruvananthapuram

MA613 Data Mining

Assignment-V

Date: 11-11-2014

1. Cluster the following data: $A_1(2,10), A_2(2,5), A_3(8,4), A_4(5,8), A_5(7,5), A_6(6,4), A_7(1,2)$, $A_8(4,9)$. Use (a) Manhattan metric and (b) Euclidean metric. Use k-means and k mediods algorithm to find the clusters.

   (a) Find the best k using Davies Bouldin index.

   (b) Plot $J(c,\mu)$ and the clusters.

2. Apply agglomerative and divisive clustering on the above data (manually) and draw the dendogram.

3. Apply either SOM algorithm or DBSCAN (any one) on IRIS data.

   (a) Report the hyperparameter values and the performance of the algorithm.

4. Write short notes on visualization techniques used in SOM.

5. Write short notes on cluster evaluation techniques.

6. Write short notes on Mahalanobis distance.

7. Apply Gooogle Page ranking on:

page id	outgoing links
1	4, 3, 2
2	3
3	4,5

   (a) Report all the important steps involved.

Notes

- All the files related with the assignment should be saved in a single folder and send to sumitra@iist.ac.in.
- Last date of submission: 19-11-2014.
- **Late submission will be penalised.**
- **As far as assignments are concerned, students are expected to observe academic honesty and integrity. Though the students can collaborate and discuss, copying directly other students' assignment or allowing your own assignment to be copied constitute academic dishonesty and is highly discouraged.**