

MTH443: Quiz 1

Full Marks 20

Time: 17:10-17:50

Consider the state/UT wise cyber crime data of India (*data source: <https://www.data.gov.in>; NCRB, Ministry of Home Affairs*) for the year 2020 (cyber_crime_2020.csv). The data gives the recorded cyber crimes in the state/UT, motive wise, corresponding to the following classified motives: (i) Personal Revenge, (ii) Anger, (iii) Fraud, (iv) Extortion, (v) Causing Disrepute, (vi) Prank, (vii) Sexual Exploitation, (viii) Political Motives, (ix) Terrorist Activities, (x) Inciting Hate against Country, (xi) Disrupt Public Service, (xii) Sale Purchase illegal drugs, (xiii) Developing own business, (xiv) Spreading Piracy, (xv) Steal Information, (xvi) Others Motives.

- (a) Obtain a PCA based projection of the data.
- (b) Detect outliers, if any.
- (c) Obtain scree plot and suggest the number of principal components for efficient data dimensional reduction.
- (d) Obtain complete linkage hierarchical cluster dendrogram. Partition the states/UTs into 4 clusters using the dendrogram and list the states/UTs in each of the clusters.
- (e) Obtain a K-means clustering with $k=4$.

Put your code, output/plot/interpretation of the output in a single file and upload the pdf file through the Exam (Quiz 1) link on mooKIT.

Note: A maximum of 17 mins compensatory time to 231080042, 231080053, as per GoI rules