

## Cluster Analysis

Find a suitable dataset and perform a cluster analysis using either Weka or R. Your data set must have more than 3000 rows.

### Rubric

1. Data set meets the requirement and is appropriate for cluster analysis. If your data set has redundant columns (e.g. "Temperature in Celcius" column is redundant to "Temperature in Fahrenheit" column) or irrelevant columns, you should remove them either in or outside of R or Weka . Describe your data. (20')
2. Use comments in the script file(s) to specify your name and explain the problem (application domain, purpose etc) (20')
3. Perform a Cluster Analysis. (30')
4. Evaluate and explain the results. (20')
5. Submit a zip file that contains your data file, script file and a WORD document with descriptions and screen shots for each of the above 4 items . File name format is lastname\_project3.zip. (10')