M4L2 Homework Assignment

This homework assignment focuses on Clustering and is due on the 8th October 2017. You will provide a written analysis based on the following information:

- First, go to the UC Irvine Machine Learning Repository at https://archive.ics.uci.edu/ml/ and find a dataset for clustering. Note that every student MUST use a different dataset so you MUST get approved for which data you are going to use. You will use this dataset for this module on unsupervised learning and the next on supervised learning. For approval please pick the dataset you propose to use and let me know at s.arunagiri@northeastern.edu.
- Next, cluster some of your data using k-means, PAM and hierarchical clustering.
- Finally, answer the following questions:
    1. How did you choose a k for k-means?
    2. Evaluate the model performance. How do the clustering approaches compare on the same data?
    3. Generate and plot confusion matrices for the k-means and PAM. What do they tell you?
    4. Generate centroid plots against the 1st two discriminant functions for k-means and PAM. What do they tell you?
    5. Generate silhouette plots for PAM. What do they tell you?
    6. For the hierarchical clustering use all linkage methods (Single Link, Complete Link, Average Link, Centroid and Minimum energy clustering) and generate dendograms. How do they compare on the same data?
    7. For the hierarchical clustering use both agglomerative and divisive clustering with a linkage method of your choice and generate dendograms. How do they compare on the same data?
    8. For the hierarchical clustering use centroid clustering and squared Euclidean distance and generate dendograms. How do they compare on the same data?

As a reminder, please provide a written analysis/report as an .Rmd file and please also submit the pdf generated by 'knit'ing your Rmd. Note: This is a graded assignment due by Sunday at 11:59 pm.

M4L3 Homework Assignment

This homework assignment is due on the 15th October 2017 and focuses on Expectation-maximization (EM). Provide a written analysis based on the following information:

- First, you will use the same dataset you chose for previous assignment (M04 Lesson 02 for the partition (k-means, PAM) and hierarchical clustering) from the he the UC Irvine Machine Learning Repository at https://archive.ics.uci.edu/ml/
- Next, cluster some of your data using EM based clustering and answer the following questions:
    1. How did you choose a model for EM? Evaluate the model performance.
    2. Cluster some of your data using EM based clustering that you also used for k-means, PAM, and hierarchical clustering. How do the clustering approaches compare on the same data?

As a reminder, please provide a written analysis/report as an .Rmd file and please also submit the pdf generated by 'knit'ing your Rmd . Note: This is a graded assignment due by Sunday at 11:59 pm.

M4L4 Homework Assignment

This homework assignment is due on the 15th October 2017 and focuses on Association Rules. You will provide a written analysis based on the following information:

- First, select a transaction dataset from the Frequent Itemset Mining Dataset Repository at http://fimi.ua.ac.be/data/ or another transaction dataset of your choice from the Web.
- Next, generate a set of 50 or so (non-redundant) rules.
- Finally answer the following questions:
    1. Which rules make sense to you? Highlight the five best and five worst of your rule set.
    2. How did you choose the level of support and confidence?
    3. What is the lift and conviction of your best and worst rules?
    4. Visualize your 50 association rules. Where do the best and worst end up in your plot?
    5. Does the model make sense?

As a reminder, please provide a written analysis/report as an .Rmd file and please also submit the pdf generated by 'knit'ing your Rmd . Note: This is a graded assignment due by Sunday at 11:59 pm.