C S 488/508: Introduction to Data Mining

Assignment 6: Association Analysis

Due date: Nov 10th, 2023 @11:55pm

You need to submit **the results of each task** (e.g., numbers, plots, files, …) as well as **the source code** that generates those results. Zip everything together and submit the zip file using the link provided on Canvas. The maximum points for this assignment are 100.

We will use a dataset called Online Retail Data Set from the UCI repository (https://archive.ics.uci.edu/ml/datasets/Online+Retail). Its csv file (Online Retail.csv) can be downloaded from Canvas. This is a transactional data set that contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. Note that the InvoiceNo column contains the id of the transaction and the column StockCode contains the ids of the items.

Based on association rules generation, what are the top 10 association rules (please use confidence as the rule evaluation metric) that you can find based on this dataset?

For Association Rules Mining, you can consider using `frequent_patterns` package in MLxtend. Please check http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/association_rules/ for how to generate the association rules using MLxtend.

---END---