



INF0613 – Aprendizado de Máquina Não Supervisionado

Trabalho 1 - Regras de Associação

Neste primeiro trabalho vamos minerar Regras de Associação em uma base de dados que contém as vendas de uma padaria. A base de dados está disponível na página da disciplina no Moodle (arquivo bakery.csv).

Informações Importantes

- **Prazo de entrega:** 04 de setembro de 2022 (Domingo), até às 23h55.
- **Forma de entrega:** Deverá ser submetido um arquivo comprimido no formato **.zip** via [Moodle](#) contendo:
 - Arquivo `inf0613-trabalho1.Rmd` com as respostas das atividades, e
 - Arquivo no formato PDF, gerado a partir do arquivo `inf0613-trabalho1.Rmd` respondido.
- **Pontuação:** Este trabalho será pontuado de 0 a 10, e corresponderá a 30% da nota final.
- Na página da disciplina no Moodle, fornecemos um arquivo `inf0613-trabalho1.Rmd` que contém o template do trabalho, seu uso é obrigatório.
- O arquivo `bakery.csv`, disponibilizado no Moodle, não deve ser alterado diretamente, isto é, todas as alterações devem ser feitas no objetos carregados em memória.
- *Dica para esta tarefa:* explore o pacote **arules**, muitas tarefas se tornam triviais se as funções desse pacote forem usadas.
- Teste o seu código antes de submeter. Códigos com erros sintáticos serão penalizados.
- Submissões com formatos diferentes dos especificados não serão corrigidas.
- Apenas um membro de cada grupo deve enviar a solução. Os **nomes completos** dos membros devem constar no cabeçalho de cada arquivo a ser submetido no local indicado.

Atividade 1 – Análise Exploratória da Base de Dados (3,0 pts)

Dado um caminho para uma base de dados, leia as transações e faça uma análise exploratória sobre elas. Use as funções `summary`, `inspect` e `itemFrequencyPlot`. Na função `inspect` limite sua análise às 10 primeiras transações e na função `itemFrequencyPlot` gere um gráfico com a frequência relativa dos 30 itens mais frequentes.

- a) Descreva a base de dados discutindo os resultados das funções acima.
- b) Ao gerarmos o gráfico de frequências, temos uma representação visual de uma informação já presente no resultado da função `summary`. Contudo, esse gráfico nos dá uma visão mais ampla da base. Assim podemos ver a frequência de outros itens em relação aos 10 mais frequentes. Quais informações podemos obter a partir desse gráfico (e da análise anterior) para nos ajudar na extração de regras de associação com o algoritmo `apriori`? Isto é, como a frequência dos itens pode afetar os parâmetros de configuração do algoritmo `apriori`?

Atividade 2 – Minerando Regras (3,5 pts)

Use o algoritmo `apriori` para minerar regras na base de dados fornecida. Experimente com pelo menos 3 conjuntos de valores diferentes de suporte e confiança para encontrar regras de associação. Imprima as cinco regras com o maior suporte de cada conjunto escolhido. Lembre-se de usar seu conhecimento sobre a base, obtido na questão anterior, para a escolha dos valores de suporte e confiança.

- a) Quais as regras mais interessantes geradas a partir dessa base? Justifique.

Atividade 3 – Medidas de Interesse (3,5 pts)

Vimos na aula que, mesmo após as podas do algoritmo `apriori`, ainda temos algumas regras com características indesejáveis como redundâncias e dependência estatística negativa. Também vimos algumas medidas que nos ajudam a analisar melhor essas regras como o lift, a convicção e a razão de chances. Nesta questão, escolha um dos conjuntos de regras gerados na atividade anterior e o analise usando essas medidas. Compute as três medidas para o conjunto escolhido com a função `interestMeasure` e experimente ordenar as regras com cada uma das novas medidas.

- a) Quais as regras mais interessantes do conjunto? Justifique.