



TRABALHO 1 (INDIVIDUAL OU EM DUPLA)

MÓDULO 2 – APRENDIZADO DE MÁQUINA NÃO SUPERVISIONADO

O objetivo deste trabalho é exercitar o conhecimento de regras de associação e obter alguns dados estatísticos da base de dados a ser utilizada em trabalhos futuros. Este trabalho está dividido em duas atividades: *Extraindo Medidas Estatísticas Usando Regras de Associação*. Essas atividades são descritas com maiores detalhes na sequência. **Note que utilizaremos bases de dados distintas para cada atividade.** Essas bases estão disponíveis na página da disciplina no Moodle.

Atividade 1: Extraindo Medidas Estatísticas

Como foi visto na primeira aula da disciplina, é importante conhecer as características dos dados de uma base disponíveis para a mineração dos mesmos. Por isso, o objetivo desta atividade é extrair medidas estatísticas descritivas da base de dados telecom-churn.csv. A seguir são listadas as análises que devem ser feitas.

1. *Separação dos Dados Numéricos*: as análises feitas nesta atividade aplicam métodos numéricos aos dados, logo devemos primeiro selecionar os atributos que são dos tipos `integer` ou `numeric`. Para isso, escreva uma função com a seguinte especificação:

ENTRADA: caminho para a base de dados;

Selecione as colunas dos tipos especificados;

SAÍDA: `data-frame` cujos atributos (colunas) são dos tipos `integer` ou `numeric`.

2. *Informações de Características*: para cada uma das colunas de um `data frame` calcule os valores de média, variância, momentos, assimetria, curtose, mínimo, máximo, mediana e moda. Para isso, escreva uma função com a seguinte especificação:

ENTRADA: `data-frame` cujos atributos são dos tipos `integer` ou `numeric`;

Imprima os valores listados acima;

SAÍDA: – (esta função não precisa ter um retorno específico).

3. *Relações entre Características*: semelhante à questão anterior, faça uma função que imprima informações de correlação e covariância dos atributos de um dado `data frame`. Sua função deve seguir a seguinte especificação:

ENTRADA: `data-frame` cujos atributos são dos tipos `integer` ou `numeric`;

Imprima os valores listados acima;

SAÍDA: – (esta função não precisa ter um retorno específico).

Atividade 2: Usando Regras de Associação

O objetivo desta atividade é a obtenção e análise de regras de associação que representem relações entre itens nas transações contidas na base de dados bakery.

1. *Análise de Frequência*: deverá ser realizada uma análise da frequência de cada item da base de dados. Você deve apresentar em um histograma os 20 itens mais frequentes. Escreva uma função com a seguinte especificação:

ENTRADA: caminho para a base de dados;

Leia a base de dados e crie um histograma com os 20 itens mais frequentes;

SAÍDA: histograma com os 20 itens mais frequentes.

2. *Mineração de Regras*: o algoritmo *apriori* deverá ser utilizado para a mineração de regras e obtenção dos valores de suporte e confiança correspondentes. Nesta questão, você deverá experimentar com os valores de suporte e confiança para encontrar regras de associação usando a função `apriori`. Experimente com pelo menos 5 conjuntos de valores, reporte quais as regras com o maior *lift*. Escreva como um comentário no arquivo que será submetido os conjuntos de valores escolhidos e duas regras de cada conjunto que você achou interessantes.
3. *Determinação de Dependência Estatística*: deverão ser utilizadas as métricas convicção (*conviction*), *lift* e razão de chances (*odds factor*) para escolher as melhores regras. Escreva uma função com a seguinte especificação:

ENTRADA: caminho para a base de dados, valor mínimo de suporte e valor mínimo de confiança;

Leia a base de dados, crie regras de associação e as filtre com os valores de interesse de cada medida ($1.01 \leq \text{convicção} \leq 5.00$, $1.0 \leq \text{lift}$ e $0.0 \leq \text{razão de chances}$);

SAÍDA: lista de regras filtradas.

Considerações Finais

- Você não deve remover qualquer linha já existente nos arquivos das bases de dados.
- 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.
- Na página da disciplina no Moodle, fornecemos um arquivo `trabalho1.R` que contém um esqueleto da entrega da tarefa, seu uso é facultativo.
- Para os trabalhos feitos em dupla, apenas um membro da dupla deve enviar a solução. Os nomes dos membros devem constar no cabeçalho de cada arquivo “.R” a ser submetido.
- Salve os arquivos utilizando o mesmo nome, e os envie no sistema Moodle, clicando no link “Trabalho 1” da Seção “Avaliações”. Clique em “Adicionar tarefa”, anexe os arquivos e, por fim, clique em “Salvar mudanças”. Você voltará para a tela da atividade e deverá constar o status “Enviado para avaliação”. A qualquer momento, antes do prazo final de submissão, você pode alterar sua submissão clicando em “Editar envio”.

Prazo de entrega: 25 de março de 2019 (Segunda-Feira), até às 23h55.

Forma de entrega: Deverá ser submetido um arquivo com `trabalho1.R` contendo o código especificado em cada atividade via Moodle:

<https://moodle.lab.ic.unicamp.br/moodle/course/view.php?id=315>

Pontuação: Este trabalho será pontuado de 0 a 10, e corresponderá a 30% da nota final.