

MBA⁺

**Artificial Intelligence &
Machine Learning**





Aprendizado não supervisionado

K-means / EM / Regras de Associação



Regras de Associação

Regras de Associação



- Regras de associação não fixam um conseqüente
- Pode-se fazer analogia com regras de classificação onde todas as variáveis podem ser conseqüentes. Entretanto existe uma diferença conceitual nesta abordagem.
- Tem como base a teoria de itens frequentes
- O Algoritmo Apriori é o mais tradicional
- Aprendizado Não-Supervisionado

O Algoritmo Apriori



Este algoritmo possui várias implementações diferentes em diferentes ferramentas de software. Tais diferentes implementações se utilizam de técnicas que visam otimizar o desempenho do algoritmo, melhorando principalmente o esforço computacional exigido para se trabalhar com grandes volumes de dados.

Apesar das diferentes implementações disponíveis, o algoritmo padrão é sempre o mesmo e possui como base 3 conceitos principais:

- Conjuntos Itens frequentes;
- Propriedade “*Apriori*”
- Operação de união (*Join*) de conjuntos

O Algoritmo Apriori



- Conjuntos de Itens frequentes: dado um conjunto de dados e suas instâncias (também chamadas de **transações** na literatura de regras de associação), estes são os conjuntos de itens (valores de atributos) que possuem um suporte maior ou igual ao mínimo desejado.
- Propriedade “*Apriori*”: qualquer subconjunto de um conjunto de itens frequentes deve ser frequente também.
- Operação de união (*Join*) de conjuntos: para se encontrar os conjuntos de itens frequentes contendo k itens, realiza-se uma operação de união dos conjuntos de itens frequentes contendo $k-1$ itens com eles próprios.

O Algoritmo Apriori



Um comentário interessante é que as instâncias do conjunto de dados são comumente chamadas de transações pelo fato da tarefa de descoberta de regras de associação estar, muitas vezes, associada à descoberta de itens frequentemente comprados em uma mesma compra (transação). Por este motivo, um dos exemplos mais famosos de extração de regras de associação utiliza como base o caso onde as transações são compras realizadas por um cliente de um supermercado. Tal exemplo ficou famoso por mencionar uma característica de consumidores que compravam cerveja e fraldas descartáveis.

O Algoritmo Apriori



- Processo de execução do *Apriori* segue o seguinte fluxo

Defina o valor do suporte mínimo;

Encontre os conjuntos unitários (chamados 1-itemsets) que contenham itens com frequência igual ou superior ao suporte mínimo;

Iterativamente, encontre os k-itemsets com base nos (k-1)-itemsets já definidos. Repita esta busca até que não haja k-itemsets que contenham itens com frequência igual ou superior ao suporte mínimo.

O Algoritmo Apriori



Com base no slide anterior, é fácil notar que a propriedade a priori está presente no algoritmo, pois utiliza-se sempre os k -itemsets para se encontrar os $(k+1)$ -itemsets.

Para facilitar a compreensão, tomemos um exemplo prático: considere o conjunto de dados definido na tabela abaixo.

Instância #	Itens (ou atributos)
1	Produto1, Produto2, Produto5
2	Produto1, Produto2, Produto4
3	Produto1, Produto2, Produto3, Produto5
4	Produto2, Produto4
5	Produto1, Produto3
6	Produto2, Produto3
7	Produto1, Produto2, Produto3
8	Produto1, Produto3
9	Produto2, Produto3

O Algoritmo Apriori



Inst. #	Itens (ou atributos)
1	Produto1, Produto2, Produto5
2	Produto1, Produto2, Produto4
3	Produto1, Produto2, Produto3, Produto5
4	Produto2, Produto4
5	Produto1, Produto3
6	Produto2, Produto3
7	Produto1, Produto2, Produto3
8	Produto1, Produto3
9	Produto2, Produto3

□ Suporte

□ **Freqüência** de ocorrência de um conjunto de itens (*itemset*)

□ Ex: $\text{suporte}(\{\text{Produto1}, \text{Produto3}\}) = 4$

□ Conjunto de itens freqüentes

□ Um *itemset* cujo suporte é maior ou igual a um dado limiar

□ Alguns conceitos importantes

□ *Itemset* (conjunto de itens)

□ Conjunto de um ou mais itens

- Exemplo:
 $\{\text{Produto1}, \text{Produto3}\}$

□ *k-itemset*

□ Itemset com k itens

- Exemplo: 2-itemset
 $\{\text{Produto1}, \text{Produto3}\}$

O Algoritmo Apriori



Inst. #	Itens (ou atributos)
1	Produto1, Produto2, Produto5
2	Produto1, Produto2, Produto4
3	Produto1, Produto2, Produto3, Produto5
4	Produto2, Produto4
5	Produto1, Produto3
6	Produto2, Produto3
7	Produto1, Produto2, Produto3
8	Produto1, Produto3
9	Produto2, Produto3

- ***Para nosso primeiro exemplo de extração de regras de Associação, vamos considerar”:***
 - Conjunto de dados da tabela ao lado;
 - Suporte mínimo = 2
 - Confiança mínima = 70% (veremos sobre confiança ao extrairmos as regras)
 - Algoritmo Apriori

- **O primeiro passo é então encontrar o conjunto 1-itemsets frequentes, ou seja, que tenham frequência maior ou igual a dois**

O Algoritmo Apriori



Inst. #	Itens (ou atributos)
1	Produto1, Produto2, Produto5
2	Produto1, Produto2, Produto4
3	Produto1, Produto2, Produto3, Produto5
4	Produto2, Produto4
5	Produto1, Produto3
6	Produto2, Produto3
7	Produto1, Produto2, Produto3
8	Produto1, Produto3
9	Produto2, Produto3

1-itemset	frequência
{Produto1}	6
{Produto2}	7
{Produto3}	6
{Produto4}	2
{Produto5}	2

□ Assim, encontra-se o conjunto de conjuntos candidatos

□ Observamos que dos conjuntos candidatos, todos apresentam suporte mínimo, assim, nosso conjunto de 1-itemset frequentes (L1) é formado pelos 5 conjuntos candidatos.

$L1 = \{\{\text{Produto1}\}, \{\text{Produto2}\}, \{\text{Produto3}\}, \{\text{Produto4}\}, \{\text{Produto5}\}\}$

□ Iniciamos então a segunda iteração, buscando o conjunto L2

O Algoritmo Apriori



Inst. #	Itens (ou atributos)
1	Produto1, Produto2, Produto5
2	Produto1, Produto2, Produto4
3	Produto1, Produto2, Produto3, Produto5
4	Produto2, Produto4
5	Produto1, Produto3
6	Produto2, Produto3
7	Produto1, Produto2, Produto3
8	Produto1, Produto3
9	Produto2, Produto3

Candidatos 2-itemset	frequência
{Produto1, Produto2}	4
{Produto1, Produto3}	4
{Produto1, Produto4}	1
{Produto1, Produto5}	2
{Produto2, Produto3}	4
{Produto2, Produto4}	2
{Produto2, Produto5}	2
{Produto3, Produto4}	0
{Produto3, Produto5}	1
{Produto4, Produto5}	0

- ❑ Os candidatos ao conjunto L2 são formados pela união de L1 com ele próprio.
- ❑ Iniciamos então a busca pelos candidatos que possuem itemsets com frequência maior ou igual a 2. E nos quais a regra Apriori valha*.

O Algoritmo Apriori



Inst. #	Itens (ou atributos)
1	Produto1, Produto2, Produto5
2	Produto1, Produto2, Produto4
3	Produto1, Produto2, Produto3, Produto5
4	Produto2, Produto4
5	Produto1, Produto3
6	Produto2, Produto3
7	Produto1, Produto2, Produto3
8	Produto1, Produto3
9	Produto2, Produto3

2-itemset	frequência
{Produto1, Produto2}	4
{Produto1, Produto3}	4
{Produto1, Produto5}	2
{Produto2, Produto3}	4
{Produto2, Produto4}	2
{Produto2, Produto5}	2

□ **L2** = {{Produto1, Produto2}, {Produto1, Produto3}, {Produto1, Produto5}, {Produto2, Produto3}, {Produto2, Produto4}, {Produto2, Produto5}}.

□ **Iniciamos então a terceira iteração, buscando o conjunto L3**

O Algoritmo Apriori



Inst #	Itens (ou atributos)
1	Produto1, Produto2, Produto5
2	Produto1, Produto2, Produto4
3	Produto1, Produto2, Produto3, Produto5
4	Produto2, Produto4
5	Produto1, Produto3
6	Produto2, Produto3
7	Produto1, Produto2, Produto3
8	Produto1, Produto3
9	Produto2, Produto3

Candidatos 3-itemset	frequência
{Produto1, Produto2, Produto3}	2
{Produto1, Produto2, Produto5}	2
{Produto1, Produto3, Produto5}	1
{Produto2, Produto3, Produto4}	0
{Produto2, Produto3, Produto5}	1
{Produto2, Produto4, Produto5}	0

- Os candidatos ao conjunto L3 são formados pela união de L2 com ele próprio.
- Iniciamos então a busca pelos candidatos que possuem frequência maior ou igual a 2.



O Algoritmo Apriori

Inst .#	Itens (ou atributos)
1	Produto1, Produto2, Produto5
2	Produto1, Produto2, Produto4
3	Produto1, Produto2, Produto3, Produto5
4	Produto2, Produto4
5	Produto1, Produto3
6	Produto2, Produto3
7	Produto1, Produto2, Produto3
8	Produto1, Produto3
9	Produto2, Produto3

Candidatos 3-itemset	frequência
{Produto1, Produto2, Produto3}	2
{Produto1, Produto2, Produto5}	2
{Produto1, Produto3, Produto5}	1
{Produto2, Produto3, Produto4}	0
{Produto2, Produto3, Produto5}	1
{Produto2, Produto4, Produto5}	0

Propriedade Apriori

- Os candidatos ao conjunto L3 são formados pela união de L2 com ele próprio.



O Algoritmo Apriori

Inst .#	Itens (ou atributos)
1	Produto1, Produto2, Produto5
2	Produto1, Produto2, Produto4
3	Produto1, Produto2, Produto3, Produto5
4	Produto2, Produto4
5	Produto1, Produto3
6	Produto2, Produto3
7	Produto1, Produto2, Produto3
8	Produto1, Produto3
9	Produto2, Produto3

3-itemset	frequência
{Produto1, Produto2, Produto3}	2
{Produto1, Produto2, Produto5}	2

□ **L3 = {{Produto1, Produto2, Produto3}, {Produto1, Produto2, Produto5}}.**

□ **Iniciamos então a quarta iteração, buscando o conjunto L4**

O Algoritmo Apriori



Inst .#	Itens (ou atributos)
1	Produto1, Produto2, Produto5
2	Produto1, Produto2, Produto4
3	Produto1, Produto2, Produto3, Produto5
4	Produto2, Produto4
5	Produto1, Produto3
6	Produto2, Produto3
7	Produto1, Produto2, Produto3
8	Produto1, Produto3
9	Produto2, Produto3

Candidatos 4-itemset	frequência
{Produto1, Produto2, Produto3, Produto5}	1

- Os candidatos ao conjunto L4 são formados pela união de L3 com ele próprio.



O Algoritmo Apriori

Inst .#	Itens (ou atributos)
1	Produto1, Produto2, Produto5
2	Produto1, Produto2, Produto4
3	Produto1, Produto2, Produto3, Produto5
4	Produto2, Produto4
5	Produto1, Produto3
6	Produto2, Produto3
7	Produto1, Produto2, Produto3
8	Produto1, Produto3
9	Produto2, Produto3

Candidatos 4-itemset	frequência
{Produto1, Produto2, Produto3, Produto5}	1

Propriedade Apriori

- Os candidatos ao conjunto L4 são formados pela união de L3 com ele próprio.
- Como não há 4-itemset que satisfaça a propriedade Apriori, o processo para.

O Algoritmo Apriori



□ Agora, com todos os itemsets frequentes definidos, busca-se construir as regras de associação.

□ Considere que o conjunto de itemsets frequentes é:

$L = \{\{\text{Produto1}\}, \{\text{Produto2}\}, \{\text{Produto3}\}, \{\text{Produto4}\}, \{\text{Produto5}\},$
 $\{\text{Produto1}, \text{Produto2}\}, \{\text{Produto1}, \text{Produto3}\}, \{\text{Produto1}, \text{Produto5}\},$
 $\{\text{Produto2}, \text{Produto3}\}, \{\text{Produto2}, \text{Produto4}\}, \{\text{Produto2}, \text{Produto5}\},$
 $\{\text{Produto1}, \text{Produto2}, \text{Produto3}\}, \{\text{Produto1}, \text{Produto2}, \text{Produto5}\}\}$

□ Para cada itemset frequente escolhido extrai-se as regras com base no seguinte processo:

□ Para cada subconjunto S (não vazio) do itemset I escolhido, gere regras da forma “Se S , então $I-S$ ” se a confiança for maior do que o limiar definido.

□ Para o cálculo da confiança de uma regra “Se S , então $I-S$ ” faça $(\text{suporte}(I)/\text{suporte}(S))$.

□ Tomemos como exemplo o itemset $\{\text{Produto1}, \text{Produto2}, \text{Produto5}\}$

O Algoritmo Apriori



□ Os subconjuntos (não vazios) do itemset {Produto1, Produto2, Produto5} são:

{Produto1, Produto2}, {Produto2, Produto5}, {Produto1, Produto5},
{Produto1}, {Produto2}, {Produto5}

□ Assim, temos as seguintes regras:

1) Se (Produto1 e Produto2), Então Produto5

Confiança = $\text{Suporte}(\{\text{Produto1}, \text{Produto2}, \text{Produto5}\}) / \text{Suporte}\{\text{Produto1}, \text{Produto2}\} = 2/4 = 0,5 = 50\%$

2) Se (Produto2 e Produto5), Então Produto1

Confiança = $\text{Suporte}(\{\text{Produto1}, \text{Produto2}, \text{Produto5}\}) / \text{Suporte}\{\text{Produto2}, \text{Produto5}\} = 2/2 = 1 = 100\%$

3) Se (Produto1 e Produto5), Então Produto2

Confiança = $\text{Suporte}(\{\text{Produto1}, \text{Produto2}, \text{Produto5}\}) / \text{Suporte}\{\text{Produto1}, \text{Produto5}\} = 2/2 = 1 = 100\%$

O Algoritmo Apriori



4) Se Produto1, Então (Produto2 e Produto5)

Confiança = $\text{Suporte}(\{\text{Produto1}, \text{Produto2}, \text{Produto5}\}) / \text{Suporte}\{\text{Produto1}\} = 2/6 = 0,33 = 33\%$

5) Se Produto2, Então (Produto1 e Produto5)

Confiança = $\text{Suporte}(\{\text{Produto1}, \text{Produto2}, \text{Produto5}\}) / \text{Suporte}\{\text{Produto2}\} = 2/7 = 0,28 = 28\%$

6) Se Produto5, Então (Produto1 e Produto2)

Confiança = $\text{Suporte}(\{\text{Produto1}, \text{Produto2}, \text{Produto5}\}) / \text{Suporte}\{\text{Produto5}\} = 2/2 = 1 = 100\%$

O Algoritmo Apriori



Se considerarmos a confiança mínima sendo 70% (como definido no slide 8), temos 3 regras extraídas do itemset {Produto1, Produto2, Produto5}

- 2) Se (Produto2 e Produto5), Então Produto1 Confiança = 100%
- 3) Se (Produto1 e Produto5), Então Produto2 Confiança = 100%
- 6) Se Produto5, Então (Produto1 e Produto2) Confiança = 100%

Questions and Feedback



[Thank you!](#)

Obrigado !



Vinicius Fernandes Caridá
vfcarida@gmail.com



@Vinicius Caridá



@Vinicius Caridá
@machine learning Brasil



@vfcarida

MBA⁺

Copyright © 2018 **Prof. Vinicius Fernandes Caridá**
 Todos direitos reservados. Reprodução ou divulgação
 total ou parcial deste documento é expressamente
 proibido sem o consentimento formal, por escrito, do
 Professor (autor).