



# MÉTODOS DE DISCRETIZAÇÃO

*Artigo: Discretization Methods*

*Autores: Ying Yang, Geoffrey I. Webb, Xindong Wu*

# SUMÁRIO

01.

## DEFINIÇÃO

Discretização  
Qualitativo x Quantitativo  
Níveis de escalas de  
medição

02.

## TAXONOMIA

Taxonomias que classificam  
os métodos primários de  
discretização

03.

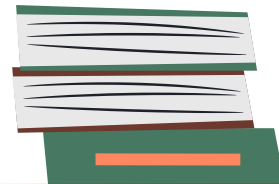
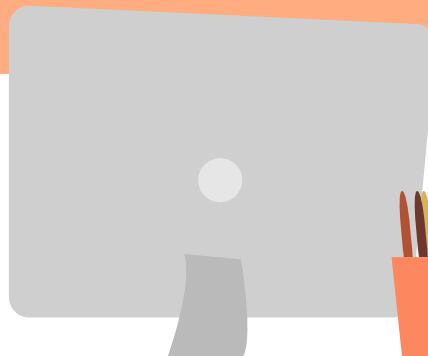
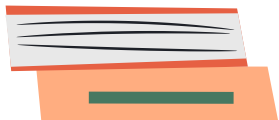
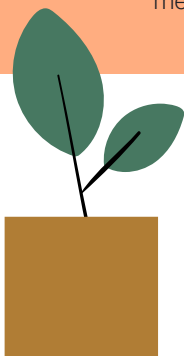
## MÉTODOS DE DISCRETIZAÇÃO

Apresentação dos métodos  
típicos  
das taxonomias discutidas

04.

## PROBLEMAS DE APRENDIZAGEM

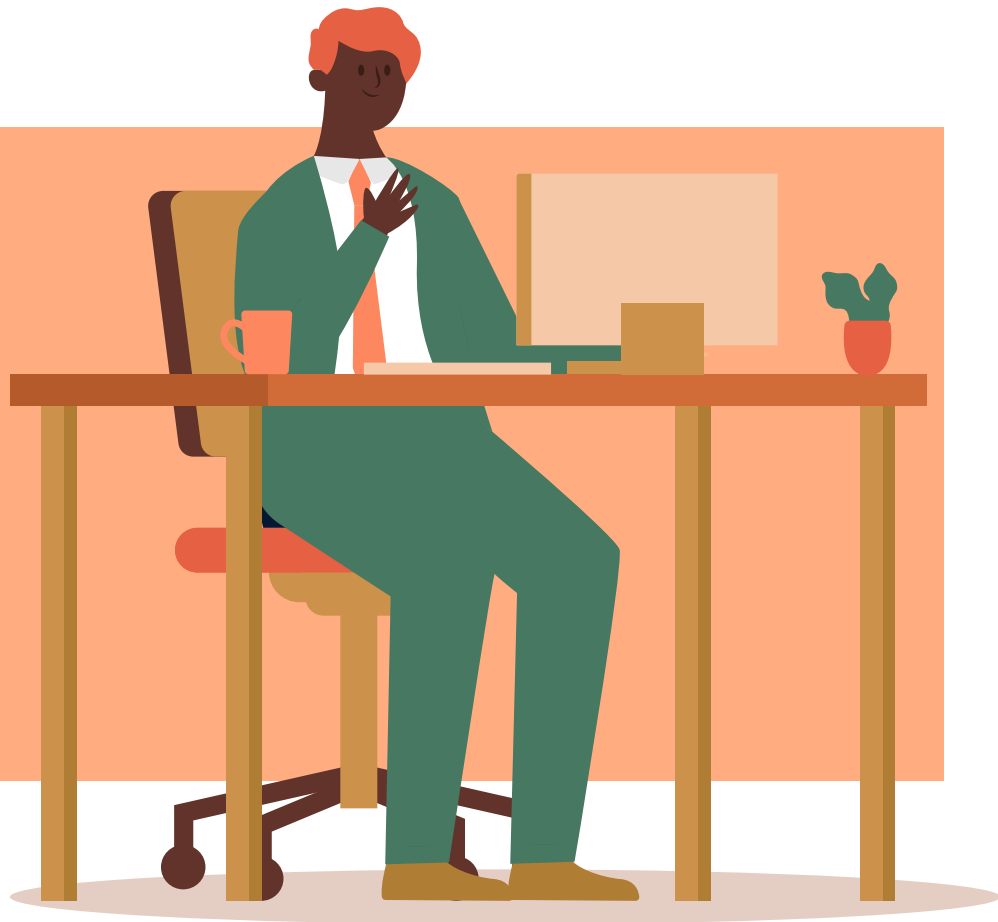
Diferentes estratégias de  
discretização são  
apropriadas para diferentes  
problemas de aprendizagem



01.

## DEFINIÇÃO

Discretização  
Qualitativo x Quantitativo  
Níveis de escalas de medição



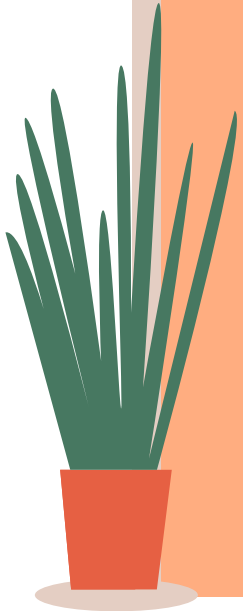
# DISCRETIZAÇÃO

Significa transformar um tipo de dados em outro tipo. Existe uma grande quantidade de termos na literatura para descrever esses dois tipos de dados. Para o presente artigo foi considerado que os dados podem ser classificados em qualitativos ou quantitativos e em diferentes níveis de escalas de medição.



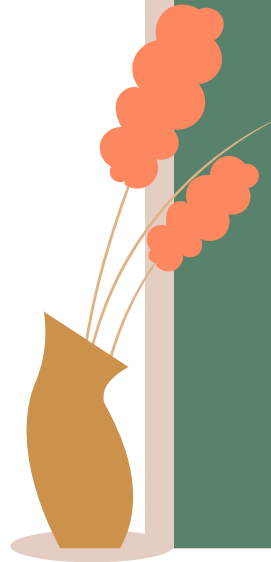
## QUALITATIVO

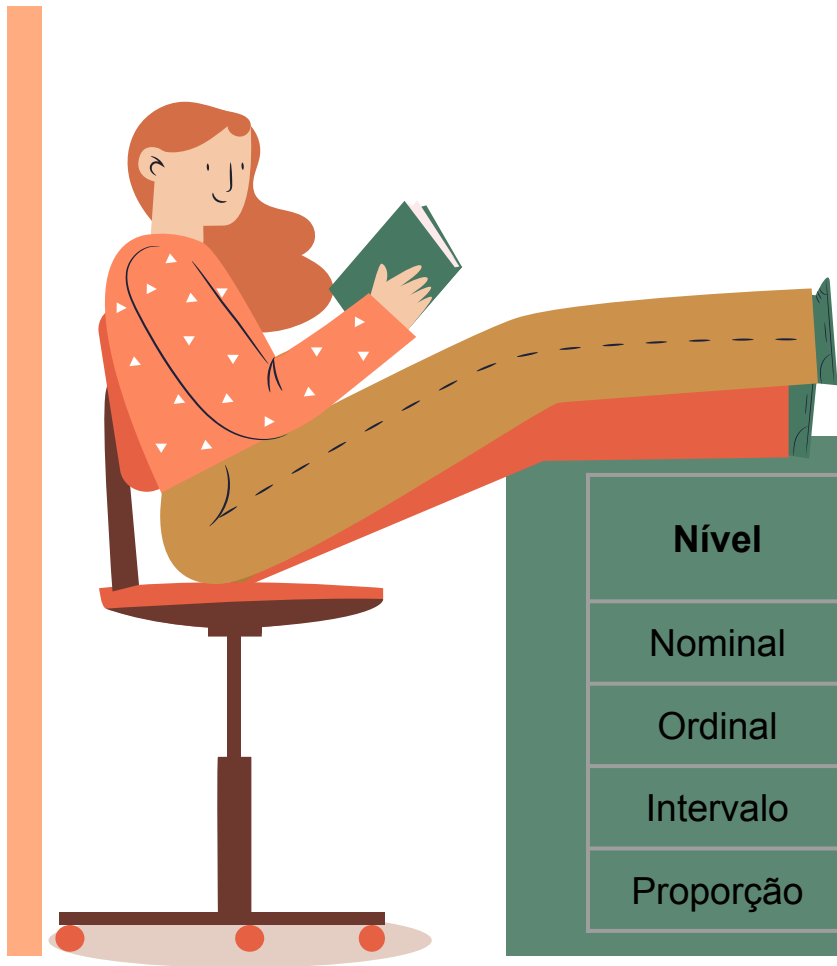
Dados que podem ser colocados em categorias distintas. Exemplo - tipo sanguíneo e avaliação de trabalho.



## QUANTITATIVO

Dados numéricos. Podem ser classificados em discretos (valores que podem ser contados) ou contínuos (valores da reta numérica).





# NÍVEIS DE ESCALAS DE MEDIÇÃO

Nível	Ranking?	Operação Aritmética?	Zero Aritmético?
Nominal	não	não	não
Ordinal	sim	não	não
Intervalo	sim	sim	não
Proporção	sim	sim	sim

02.

## TAXONOMIA

Taxonomias que classificam  
os métodos primários de  
discretização



# TAXONOMIA



## SUPERVISIONADO

Usa as informações de classe das instâncias de treinamento. Pode ser caracterizado como baseados em erro, entropia dos intervalos ou métrica estatística.



## NÃO-SUPERVISIONADO

Não usa as informações da classe.



## PARAMÉTRICO

Requer dado de entrada do usuário, como o número máximo de intervalos discretizados.



## NÃO-PARAMÉTRICO

Usa apenas informações de dados e não precisa de entrada do usuário.



# TAXONOMIA



## HIERÁRQUICO

Seleciona pontos de corte em um processo incremental, formando uma hierarquia implícita na faixa de valores. O procedimento pode ser dividido ou mesclado. E às vezes, ambos.



## NÃO-HIERÁRQUICO

Não forma nenhuma hierarquia durante a discretização.



## UNIVARIADO

Discretiza cada atributo isoladamente.



## MULTIVARIADO

Leva em consideração os relacionamentos entre os atributos durante a discretização.

# TAXONOMIA



## DISJUNTO

Discretiza a faixa de valores do atributo sob discretização em intervalos disjuntos.  
Nenhum intervalo se sobrepõe.



## NÃO-DISJUNTO

Discretiza a faixa de valores em intervalos que podem se sobrepor.



## GLOBAL

Discretiza em relação a todo o espaço de dados de treinamento.  
Realiza a discretização apenas uma vez.



## LOCAL

Permite que diferentes conjuntos de intervalos sejam formados para um único atributo, cada conjunto sendo aplicado em um contexto de classificação diferente.

# TAXONOMIA



## ANTECIPADO

Realiza discretização antes do tempo de classificação.



## LENTO

Realiza discretização durante o tempo de classificação.



## SENSÍVEL AO TEMPO

O valor qualitativo associado a um valor quantitativo pode mudar ao longo do tempo.



## NÃO SENSÍVEL AO TEMPO

Usa apenas as propriedades estacionárias dos dados quantitativos.

# TAXONOMIA



## ORDINAL

Transforma dados quantitativos em dados qualitativos ordinais.



## NOMINAL

Transforma dados quantitativos em dados qualitativos nominais.



## DIFUSO

Discretiza valores de atributos quantitativos em intervalos e em seguida, coloca algum tipo de função de associação em cada ponto de corte como bordas difusas.



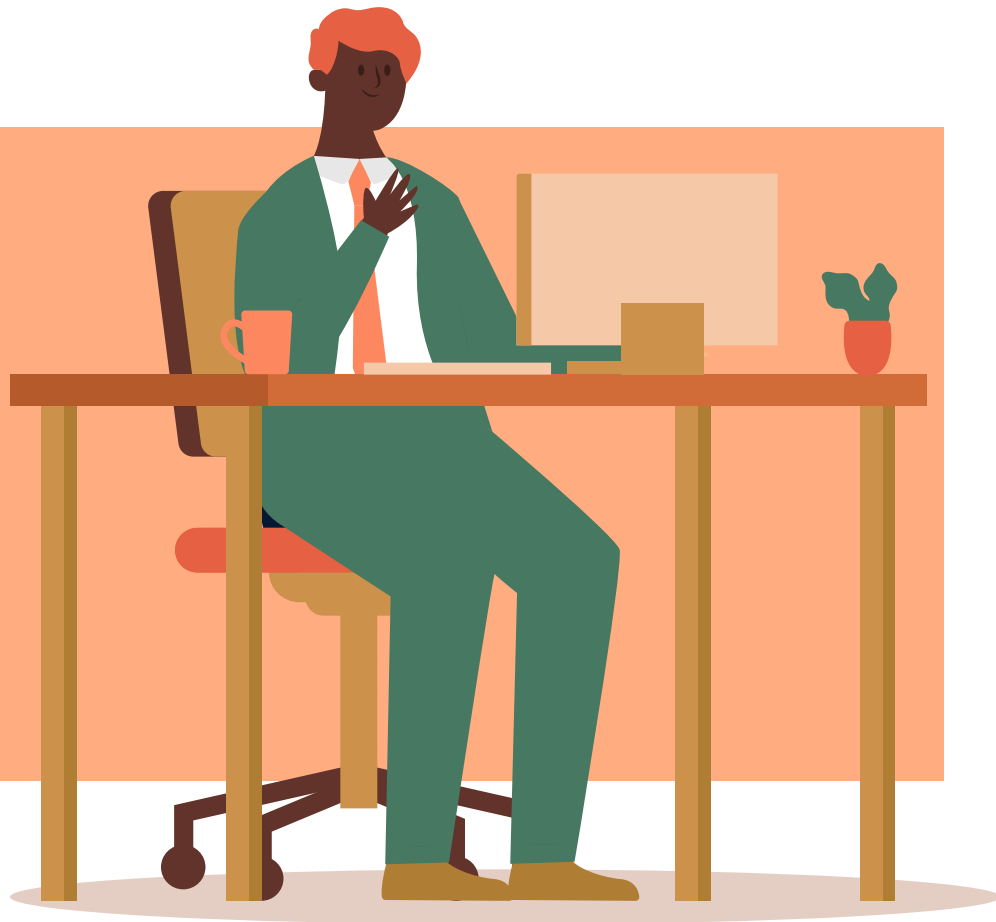
## NÃO-DIFUSO

Forma bordas nítidas sem empregar qualquer função de associação.

03.

## MÉTODOS DE DISCRETIZAÇÃO

Apresentação dos métodos típicos  
das taxonomias discutidas





## PONTO DE CORTE

Valor entre os dados quantitativos em que um limite de intervalo é localizado por um método de discretização.



## PONTO DE CORTE DE FRONTEIRA

Valores entre duas instâncias com classes diferentes na sequência de instâncias classificadas por um atributo quantitativo.

Method	Taxonomy (corresponding to Section 2)										
	0.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Equal-width	primary	unsupervised	parametric	non-hierarchical	univariate	disjoint	global	eager	time-insensitive	nominal	non-fuzzy
Equal-frequency											
Fixed-frequency											
Multi-interval-entropy-minimization	primary	supervised	non-parametric	hierarchical	univariate	disjoint	global	eager	time-insensitive	nominal	non-fuzzy
ChiMerge	primary	supervised	non-parametric	hierarchical	univariate	disjoint	global	eager	time-insensitive	nominal	non-fuzzy
StatDisc											
InfoMerge											
Cluster-based	primary	unsupervised	non-parametric	hierarchical	multivariate	disjoint	global	eager	time-insensitive	nominal	non-fuzzy
ID3	primary	supervised	parametric	hierarchical	univariate	disjoint	local	eager	time-insensitive	nominal	non-fuzzy
Non-disjoint	composite	unsupervised	*	non-hierarchical	univariate	non-disjoint	global	eager	time-insensitive	nominal	non-fuzzy
Lazy	composite	*	*	*	univariate	non-disjoint	global	lazy	time-insensitive	nominal	non-fuzzy
Dynamic-qualitative	primary	unsupervised	non-parametric	non-hierarchical	univariate	disjoint	local	lazy	time-sensitive	nominal	non-fuzzy
Ordinal	composite	*	*	*	univariate	disjoint	global	eager	time-insensitive	ordinal	non-fuzzy
Fuzzy	composite	*	*	*	univariate	non-disjoint	global	eager	time-insensitive	nominal	fuzzy
Iterative-improvement	composite	supervised	*	hierarchical	multivariate	disjoint	global	eager	time-insensitive	nominal	non-fuzzy

Note: each entry of the taxonomy is

0. primary vs. composite;
1. supervised vs. unsupervised;
2. parametric vs. non-parametric;
3. hierarchical vs. non-hierarchical;
4. univariate vs. multivariate;
5. disjoint vs. non-disjoint;
6. global vs. local;
7. eager vs. lazy;
8. time-sensitive vs. time-insensitive;
9. ordinal vs. nominal;
10. fuzzy vs. non-fuzzy.

An entry filled with '\*' indicates that the corresponding method can be conducted in either way of the corresponding taxonomy entry.

This often happens for composite methods, whose taxonomy depends on their primary methods.

# MÉTODOS X TAXONOMIAS

- Total de 15 métodos discutidos;
- Aqueles rotulados com “\*” são métodos compostos (aceitam qualquer uma das taxonomias), ficando a cargo do método primário.

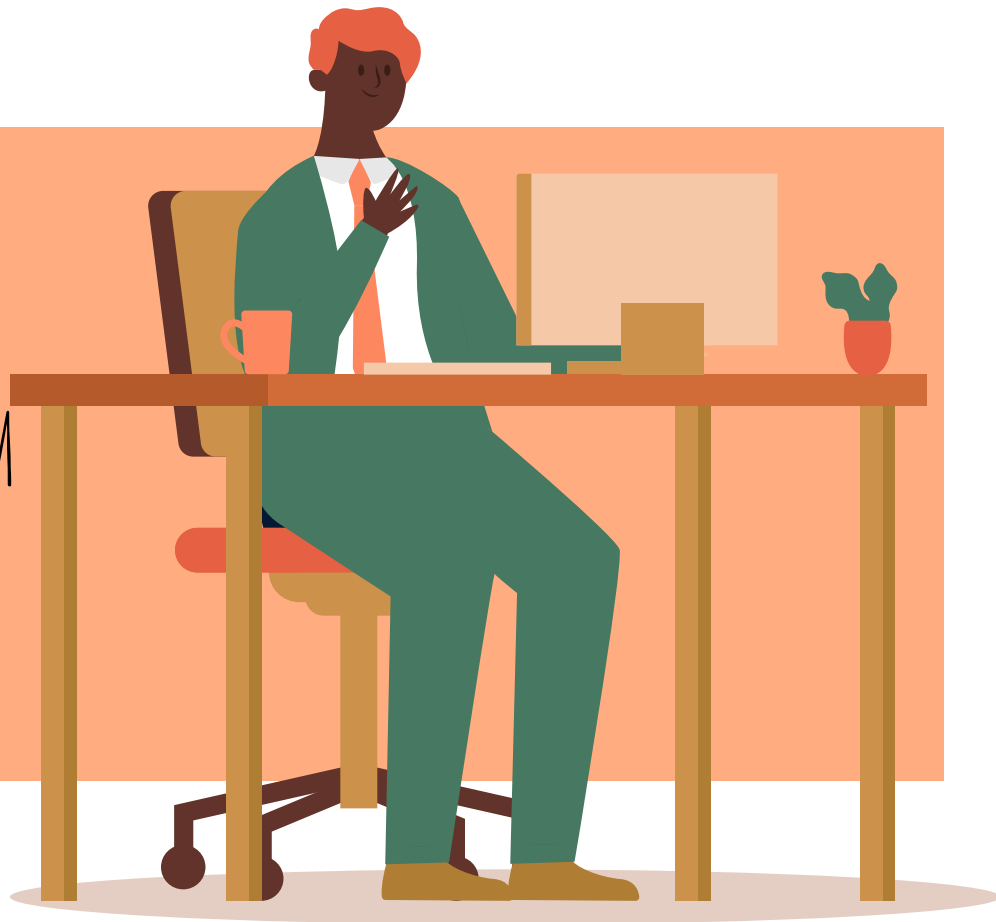




04.

## PROBLEMAS DE APRENDIZAGEM

Diferentes estratégias de discretização são apropriadas para diferentes problemas de aprendizagem





# PROBLEMAS DE APRENDIZAGEM

Cada método de discretização disponível é apropriado para um tipo de aprendizagem diferente.

Cada tipo de aprendizado requer estratégias diferentes pois possuem características distintas.

**EXEMPLO:** Para o método de *árvores de decisão* a fragmentação é um problema, então um método de discretização que gera poucos intervalos é mais vantajoso.



An illustration of a man and a woman in a classroom or meeting setting. The man, on the left, has dark curly hair and is wearing an orange long-sleeved shirt and brown pants. He is sitting on a brown stool at a dark brown table, gesturing with his right hand while looking towards the center. A laptop is open on the table in front of him. The woman, on the right, has brown hair tied in a bun and is wearing a brown patterned sweater and orange pants. She is sitting on a brown stool at a dark brown table, looking at a computer monitor that displays several lines of text. The background is a large, solid green rectangle.

# OBRIGADO!

GIAN GIOVANNI RODRIGUES

JESSYCA MORAES

TAMMY GUSMÃO

THALITA ALVES