



Machine Learning: Conceitos

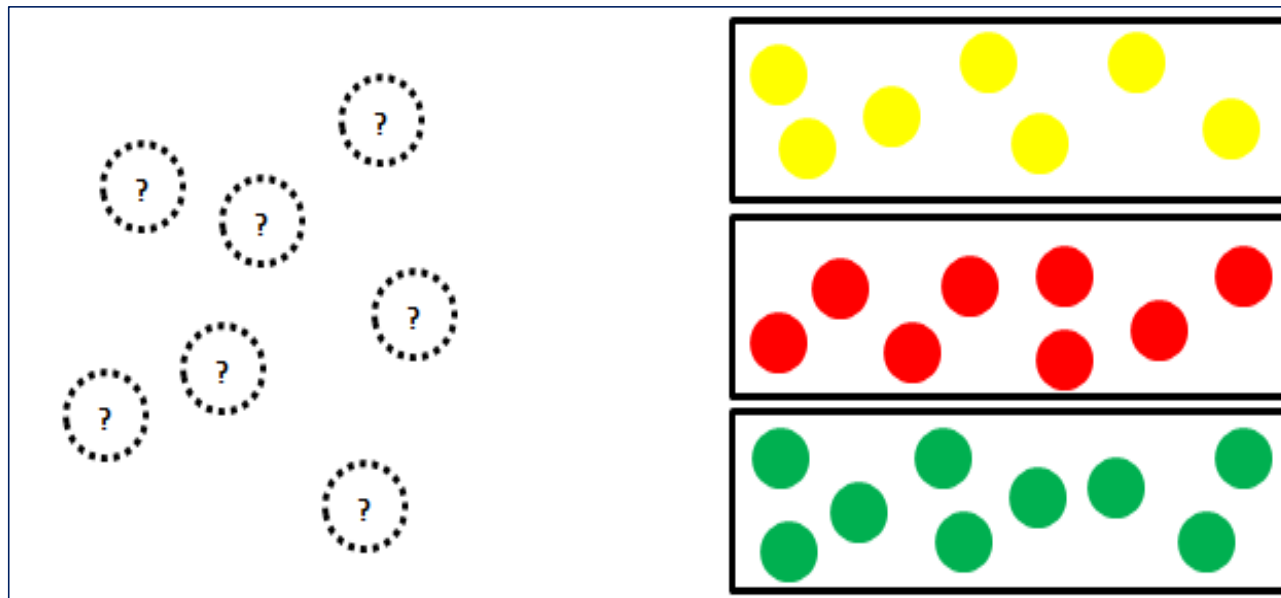
Formação Cientista de Dados

Tarefas

- Classificação
- Regressão
- Agrupamentos
- Regras de Associação

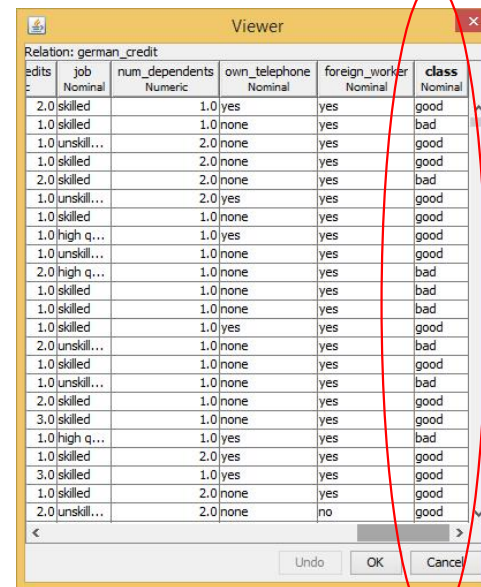


Classificação



Classificação

- Descrever ou prever um atributo especial chamado classe.
- Usamos classificação para prever uma fraude, descobrir a qual espécie um animal pertence, prever uma doença ou classificar um tipo de fungo.

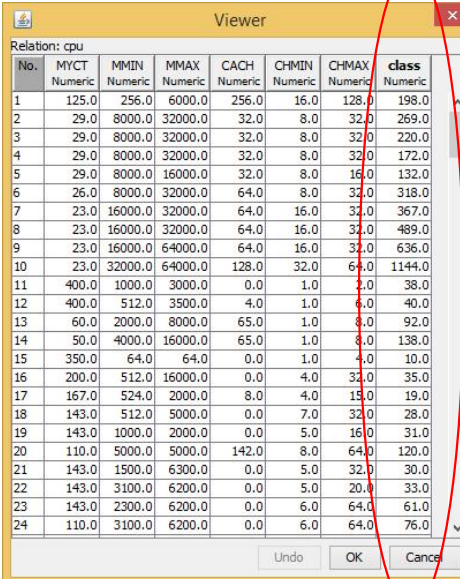


The screenshot shows a 'Viewer' window with a table titled 'Relation: german_credit'. The table has six columns: 'credits', 'job', 'num_dependents', 'own_telephone', 'foreign_worker', and 'class'. The 'class' column is circled in red. The table contains 25 rows of data.

credits	job	num_dependents	own_telephone	foreign_worker	class
Nominal	Nominal	Numeric	Nominal	Nominal	Nominal
2.0	skilled	1.0	yes	yes	good
1.0	skilled	1.0	none	yes	bad
1.0	unskill...	2.0	none	yes	good
1.0	skilled	2.0	none	yes	good
2.0	skilled	2.0	none	yes	bad
1.0	unskill...	2.0	yes	yes	good
1.0	skilled	1.0	none	yes	good
1.0	high q...	1.0	yes	yes	good
1.0	unskill...	1.0	none	yes	good
2.0	high q...	1.0	none	yes	bad
1.0	skilled	1.0	none	yes	bad
1.0	skilled	1.0	none	yes	bad
1.0	skilled	1.0	yes	yes	good
2.0	unskill...	1.0	none	yes	bad
1.0	skilled	1.0	none	yes	good
1.0	unskill...	1.0	none	yes	bad
2.0	skilled	1.0	none	yes	good
3.0	skilled	1.0	none	yes	good
1.0	high q...	1.0	yes	yes	bad
1.0	skilled	2.0	yes	yes	good
3.0	skilled	1.0	yes	yes	good
1.0	skilled	2.0	none	yes	good
2.0	unskill...	2.0	none	no	good

Regressão

- Um tipo de classificação, porém, enquanto na classificação a classe é um tipo de dado nominal ou categórico, na regressão a classe é numérica.
- Prever a altura de uma pessoa a partir do peso é um exemplo de tarefa de regressão.

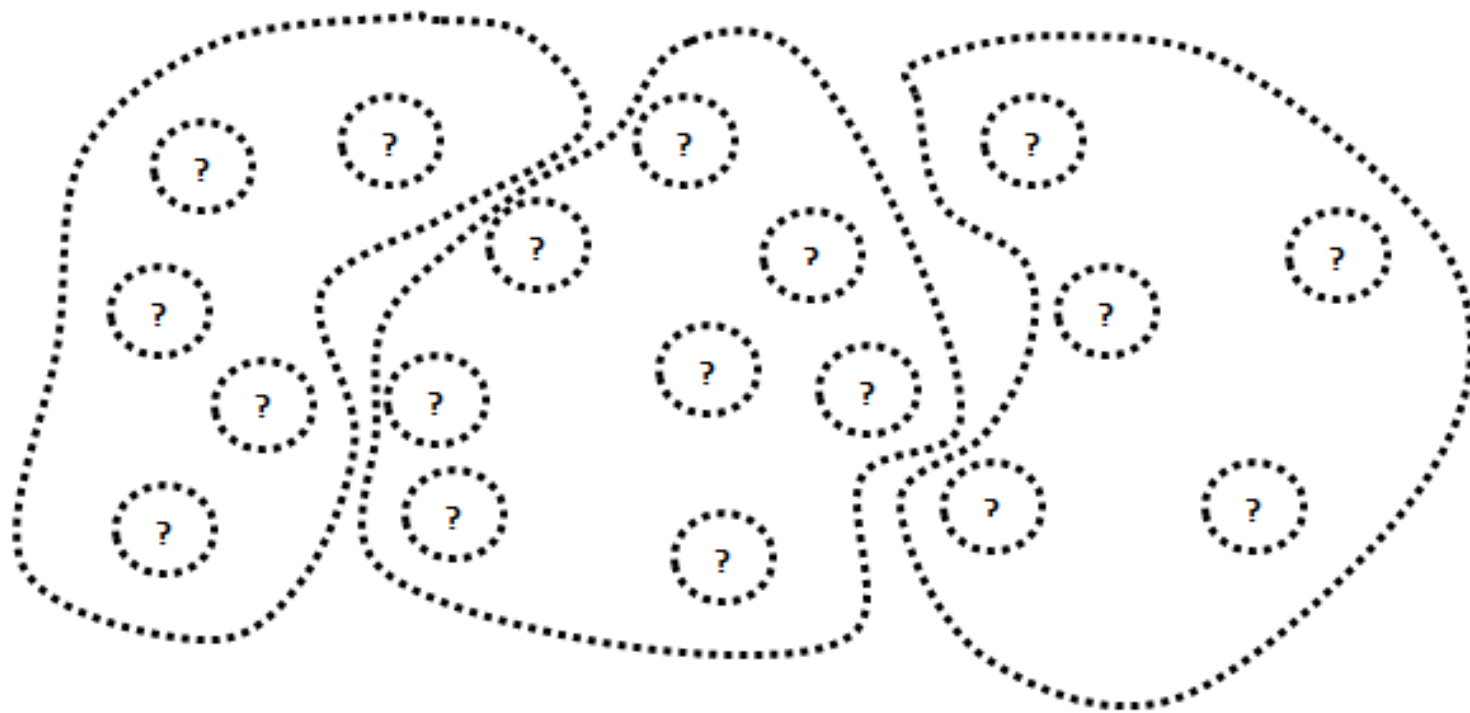


Relation: cpu

No.	MYCT Numeric	MMIN Numeric	MMAX Numeric	CACH Numeric	CHMIN Numeric	CHMAX Numeric	class Numeric
1	125.0	256.0	6000.0	256.0	16.0	128.0	198.0
2	29.0	8000.0	32000.0	32.0	8.0	32.0	269.0
3	29.0	8000.0	32000.0	32.0	8.0	32.0	220.0
4	29.0	8000.0	32000.0	32.0	8.0	32.0	172.0
5	29.0	8000.0	16000.0	32.0	8.0	16.0	132.0
6	26.0	8000.0	32000.0	64.0	8.0	32.0	318.0
7	23.0	16000.0	32000.0	64.0	16.0	32.0	367.0
8	23.0	16000.0	32000.0	64.0	16.0	32.0	489.0
9	23.0	16000.0	64000.0	64.0	16.0	32.0	636.0
10	23.0	32000.0	64000.0	128.0	32.0	64.0	1144.0
11	400.0	1000.0	3000.0	0.0	1.0	2.0	38.0
12	400.0	512.0	3500.0	4.0	1.0	6.0	40.0
13	60.0	2000.0	8000.0	65.0	1.0	8.0	92.0
14	50.0	4000.0	16000.0	65.0	1.0	8.0	138.0
15	350.0	64.0	64.0	0.0	1.0	4.0	10.0
16	200.0	512.0	16000.0	0.0	4.0	32.0	35.0
17	167.0	524.0	2000.0	8.0	4.0	15.0	19.0
18	143.0	512.0	5000.0	0.0	7.0	32.0	28.0
19	143.0	1000.0	2000.0	0.0	5.0	16.0	31.0
20	110.0	5000.0	5000.0	142.0	8.0	64.0	120.0
21	143.0	1500.0	6300.0	0.0	5.0	32.0	30.0
22	143.0	3100.0	6200.0	0.0	5.0	20.0	33.0
23	143.0	2300.0	6200.0	0.0	6.0	64.0	61.0
24	110.0	3100.0	6200.0	0.0	6.0	64.0	76.0

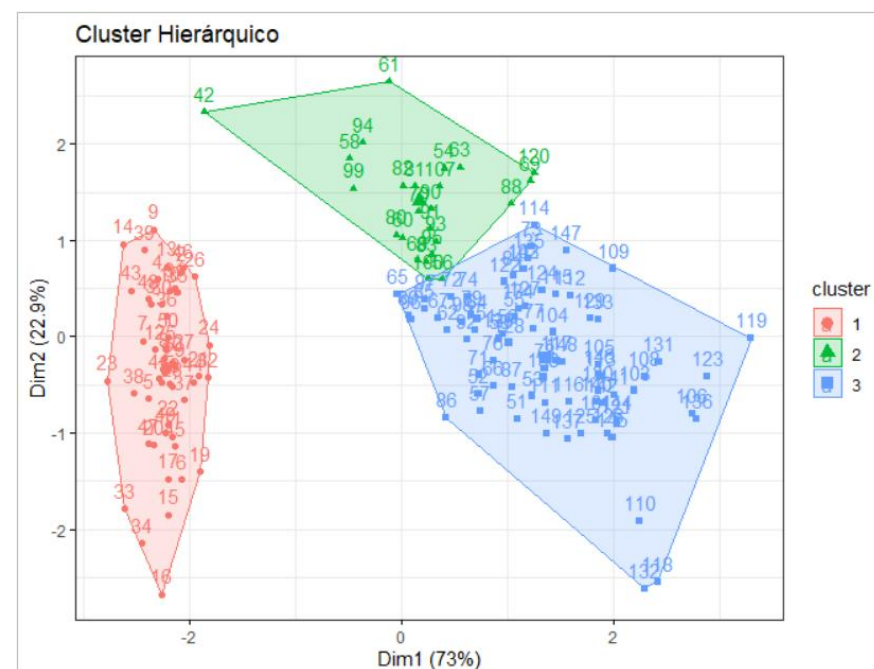
Undo OK Cancel

Agrupamentos

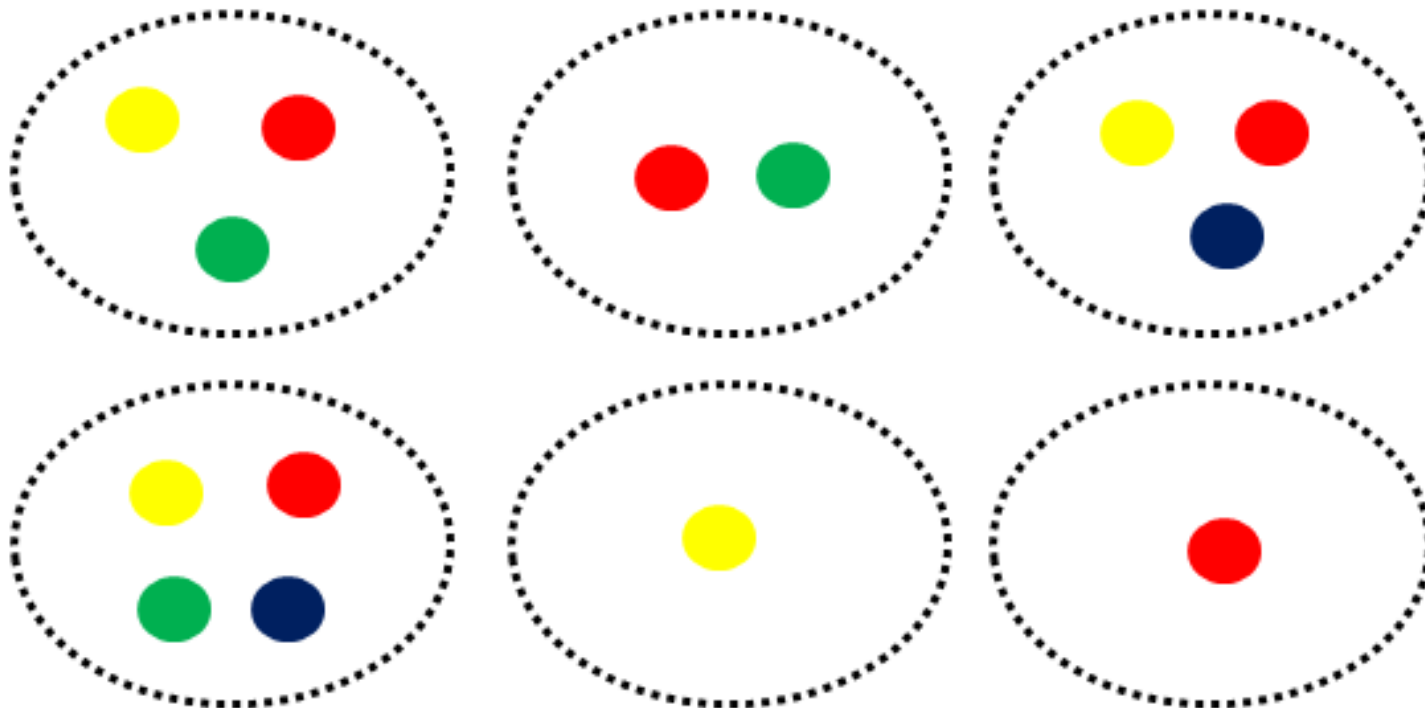


Agrupamentos

- Não existe classe.
- O objetivo é criar grupos e atribuir às instancias dos dados nestes grupos, a partir das características, ou atributos destas instancia.
- Exemplos de uso: identificar grupos de clientes e direcionar campanhas de marketing específicas; identificar tentativas de acesso a rede; categorizar uma nova espécie entre outros.



Regras de Associações



Regras de Associações

Relation: supermarket

mens toiletries Nominal	medicines Nominal	haircare Nominal	dental needs Nominal	lotions-creams Nominal	sanitary pads Nominal
t					
		t			
t					
		t	t		
					t
					t
			t		
		t	t		
				t	
		t		t	
		t	t		
			t		

- Buscam a relação entre itens. A aplicação clássica é em cestas de compras: quem comprou produto A também comprou produto B. Porém, têm ampla aplicação em diagnósticos em medicina, sensores etc.

Supervisionado Versus Não Supervisionado

- As tarefas de mineração de dados são ditas supervisionadas quando existe uma classe, ou um atributo especial com o qual se pode comparar e validar o resultado.

Tarefa não é Algoritmo

Tarefas	Tipos de Algoritmos	Algoritmos
Classificação	Bayes	NaiveBayes BayesNet
	Rules	OneR Part
Agrupamentos	Por Densidade	DBSCAN
	Centróide	K-means
Regras de Associação	Frequência de Sub-conjuntos	Apriori
	Árvore de Dados	FP Growth



Resposta



Identificar quais sintomas estão relacionados com quais doenças

- ☐ Classificação
- ☐ Regressão
- ☐ Agrupamentos
- ☐ Regras de Associação

