

TRILHA:
Inteligência Artificial

Ministrador por:
Guilherme Feulo

Machine Learning

Learn

EaD

Aprenda



OCEAN | SAMSUNG



Nossos parceiros:

USP

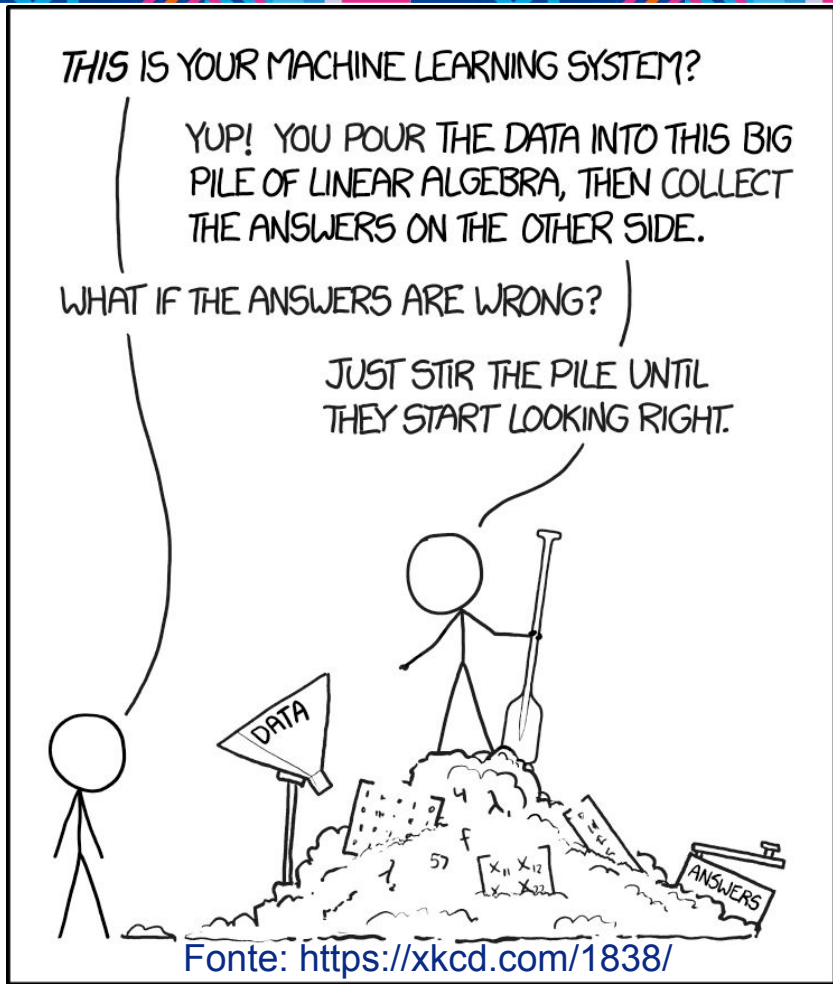
UEA

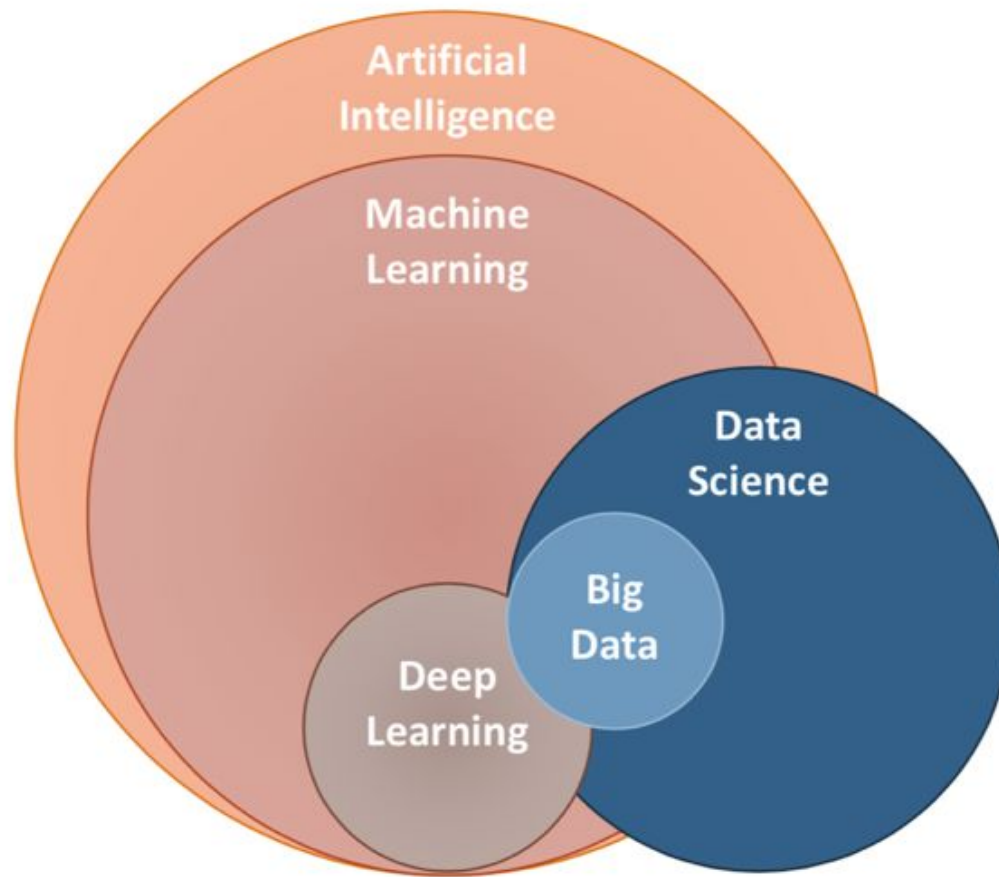
UNIVERSIDADE
DE SÃO PAULO
UNIVERSIDADE
DE SÃO PAULO

Álgebra Linear

Pra que mesmo?

- Para entender melhor datasets e representação de imagens
- Na preparação de dados, exemplo: redução de dimensionalidade
- Métodos e notações de Álgebra Linear são frequentes em Data Science, Machine Learning e Deep Learning





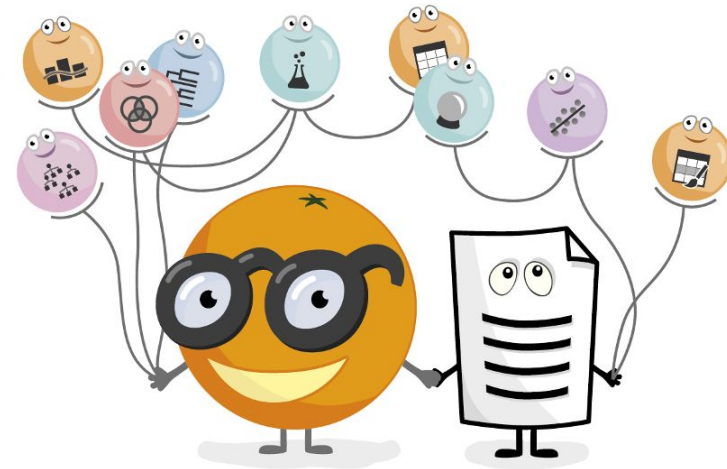
Playground

Fast.ai



Orange Canvas

- Algoritmos de aprendizagem de máquina
 - Naive Bayes
 - Árvores de decisão
 - Aprendizagem por Regras
 - KNN (vizinhos mais próximos)
- Software Orange (<https://orange.biolab.si/>)
- Prática





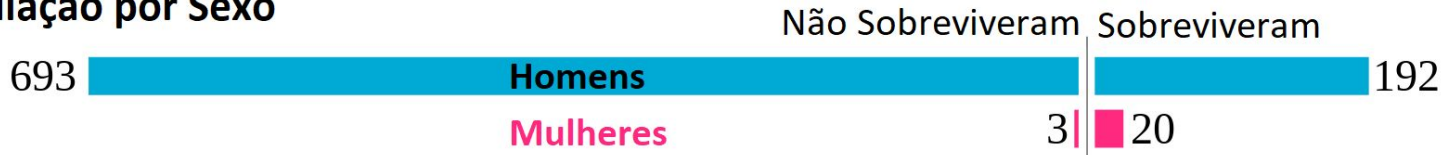
RMS Titanic - Fonte: Wikipedia

15 de abril de 1912 RMS Titanic

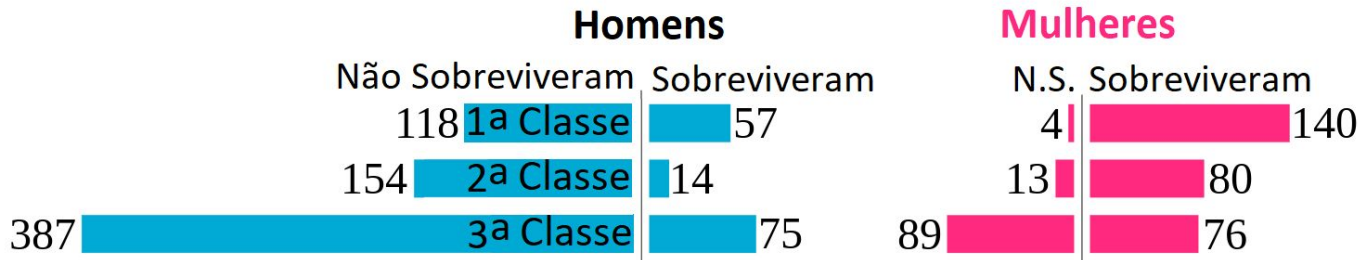
Passageiros por Sexo



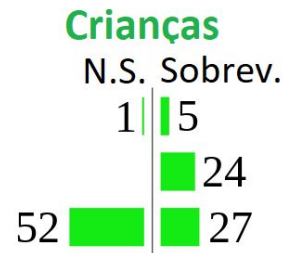
Tripulação por Sexo



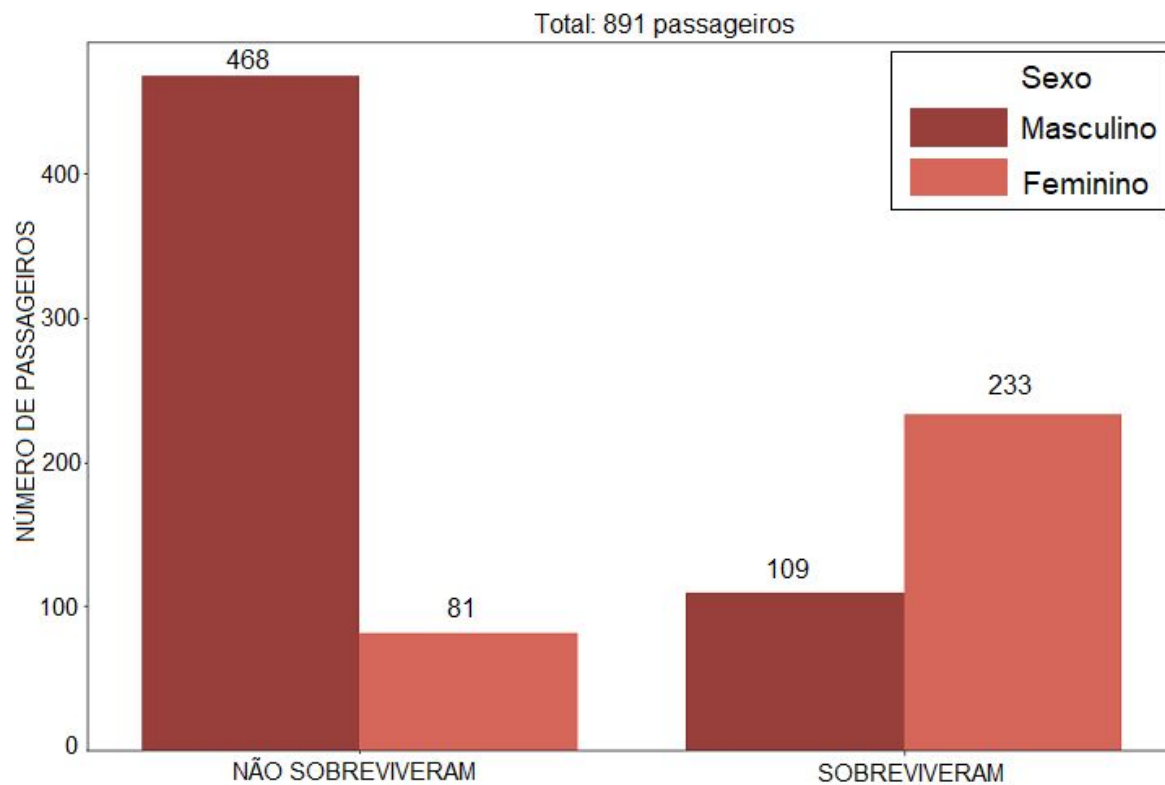
Passageiros Adultos por Classe e Sexo



Crianças por Classe



Fonte: Wikimedia Commons



Fonte: <https://paulovasconcellos.com.br>

A visualização do dataset nos mostra uma *matriz*!

Colunas; Campos; Atributos; Features...

Sobrevivente Classe Sexo Idade Tarifa

0	0	3	male	22.0	7.2500
1	1	1	female	38.0	71.2833
2	1	3	female	26.0	7.9250
3	1	1	female	35.0	53.1000
4	0	3	male	35.0	8.0500
5	0	3	male	NaN	8.4583
...
1312	0	2	male	21.0	11.5000

Linhas; Registros; Instâncias...

Descrição do dataset

	Sobrevivente	Classe	Idade	Tarifa
count	891.000000	891.000000	714.000000	891.000000
mean	0.383838	2.308642	29.699118	32.204208
std	0.486592	0.836071	14.526497	49.693429
min	0.000000	1.000000	0.420000	0.000000
25%	0.000000	2.000000	20.125000	7.910400
50%	0.000000	3.000000	28.000000	14.454200
75%	1.000000	3.000000	38.000000	31.000000
max	1.000000	3.000000	80.000000	512.329200

Naive Bayes

Vantagens e desvantagens

● Vantagens

- Rápido
- Simplicidade de interpretação
- Trabalha com altas dimensões
- Boas previsões em bases pequenas

● Desvantagens

- Combinação de características (atributos independentes) – cada par de características são independentes – nem sempre é verdade

Base original

História do crédito	Dívida	Garantias	Renda anual	Risco
Ruim	Alta	Nenhuma	< 15.000	Alto
Desconhecida	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto
Desconhecida	Baixa	Nenhuma	>= 15.000 a <= 35.000	Moderado
Desconhecida	Baixa	Nenhuma	> 35.000	Alto
Desconhecida	Baixa	Nenhuma	> 35.000	Baixo
Desconhecida	Baixa	Adequada	> 35.000	Baixo
Ruim	Baixa	Nenhuma	< 15.000	Alto
Ruim	Baixa	Adequada	> 35.000	Moderado
Boa	Baixa	Nenhuma	> 35.000	Baixo
Boa	Alta	Adequada	> 35.000	Baixo
Boa	Alta	Nenhuma	< 15.000	Alto
Boa	Alta	Nenhuma	>= 15.000 a <= 35.000	Moderado
Boa	Alta	Nenhuma	> 35.0000	Baixo
Ruim	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

Renda anual	Risco
< 15.000	Alto
>= 15.000 a <= 35.000	Alto
>= 15.000 a <= 35.000	Moderado
> 35.000	Alto
> 35.000	Baixo
> 35.000	Baixo
< 15.000	Alto
> 35.000	Moderado
> 35.000	Baixo
> 35.000	Baixo
< 15.000	Alto
>= 15.000 a <= 35.000	Moderado
> 35.000	Baixo
>= 15.000 a <= 35.000	Alto

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa	Desconhecida	Ruim	Alta	Baixa	Nenhuma	Adequada	<15	>=15<=35	>35

História do crédito	Risco
Ruim	Alto
Desconhecida	Alto
Desconhecida	Moderado
Desconhecida	Alto
Desconhecida	Baixo
Desconhecida	Baixo
Ruim	Alto
Ruim	Moderado
Boa	Baixo
Boa	Baixo
Boa	Alto
Boa	Moderado
Boa	Baixo
Ruim	Alto

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta	Baixa	Nenhuma	Adequada	<15	>=15 <=35	>35
Alto 6/14	1/6	2/6	3/6							
Moderado 3/14	1/3	1/3	1/3							
Baixo 5/14	3/5	2/5	0							

História do crédito	Risco
Ruim	Alto
Desconhecida	Alto
Desconhecida	Moderado
Desconhecida	Alto
Desconhecida	Baixo
Desconhecida	Baixo
Ruim	Alto
Ruim	Moderado
Boa	Baixo
Boa	Baixo
Boa	Alto
Boa	Moderado
Boa	Baixo
Ruim	Alto

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma	Adequada	<15	>=15 <=35	>35
Alto 6/14	1/6	2/6	3/6	4/6	2/6					
Moderado 3/14	1/3	1/3	1/3	1/3	2/3					
Baixo 5/14	3/5	2/5	0	2/5	3/5					

Dívida	Risco
Alta	Alto
Alta	Alto
Baixa	Moderado
Baixa	Alto
Baixa	Baixo
Baixa	Baixo
Baixa	Alto
Baixa	Moderado
Baixa	Baixo
Alta	Baixo
Alta	Alto
Alta	Moderado
Alta	Baixo
Alta	Alto

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15	>=15 <=35	>35
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0			
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3			
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5			

Garantias	Risco
Nenhuma	Alto
Nenhuma	Alto
Nenhuma	Moderado
Nenhuma	Alto
Nenhuma	Baixo
Adequada	Baixo
Nenhuma	Alto
Adequada	Moderado
Nenhuma	Baixo
Adequada	Baixo
Nenhuma	Alto
Nenhuma	Moderado
Nenhuma	Baixo
Nenhuma	Alto

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

Renda anual	Risco
< 15.000	Alto
>= 15.000 a <= 35.000	Alto
>= 15.000 a <= 35.000	Moderado
> 35.000	Alto
> 35.000	Baixo
> 35.000	Baixo
< 15.000	Alto
> 35.000	Moderado
> 35.000	Baixo
> 35.000	Baixo
< 15.000	Alto
>= 15.000 a <= 35.000	Moderado
> 35.000	Baixo
>= 15.000 a <= 35.000	Alto

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

História = Boa
 Dívida = Alta
 Garantias = Nenhuma
 Renda = > 35

Risco de crédito?

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

História = Boa
Dívida = Alta
Garantias = Nenhuma
Renda = > 35

Risco de crédito?

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

História = Boa
Dívida = Alta
Garantias = Nenhuma
Renda = > 35

Risco de crédito?

$P(\text{Alto}) = 6/14 * 1/6 * 4/6 * 6/6 * 1/6$
 $P(\text{Alto}) = 0,0079$

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual			História = Boa Dívida = Alta Garantias = Nenhuma Renda = > 35 Risco de crédito?
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7	
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6	P(Alto) = 6/14 * 1/6 * 4/6 * 6/6 * 1/6 P(Alto) = 0,0079
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3	P(Moderado) = 3/14 * 1/3 * 1/3 * 2/3 * 1/3 P(Moderado) = 0,0052
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5	

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual			História = Boa Dívida = Alta Garantias = Nenhuma Renda = > 35
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7	Risco de crédito?
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6	$P(\text{Alto}) = 6/14 * 1/6 * 4/6 * 6/6 * 1/6$ $P(\text{Alto}) = 0,0079$
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3	$P(\text{Moderado}) = 3/14 * 1/3 * 1/3 * 2/3 * 1/3$ $P(\text{Moderado}) = 0,0052$
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5	$P(\text{Baixo}) = 5/14 * 3/5 * 2/5 * 3/5 * 5/5$ $P(\text{Baixo}) = 0,0514$

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual			História = Boa Dívida = Alta Garantias = Nenhuma Renda = > 35 Risco de crédito?
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7	
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6	P(Alto) = 6/14 * 1/6 * 4/6 * 6/6 * 1/6 P(Alto) = 0,0079
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3	P(Moderado) = 3/14 * 1/3 * 1/3 * 2/3 * 1/3 P(Moderado) = 0,0052
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5	P(Baixo) = 5/14 * 3/5 * 2/5 * 3/5 * 5/5 P(Baixo) = 0,0514

SOMA Risco de crédito = 0.0079 + 0.0052 + 0.0514 = 0.0645

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual			<div> História = Boa Dívida = Alta Garantias = Nenhuma Renda = > 35 </div> <div>Risco de crédito?</div>
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7	

Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

SOMA Risco de crédito = 0.0079 + 0.0052 + 0.0514 = 0.0645

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual			<div> História = Boa Dívida = Alta Garantias = Nenhuma Renda = > 35 </div> <div>Risco de crédito?</div>
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7	

Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

SOMA Risco de crédito = 0.0079 + 0.0052 + 0.0514 = 0.0645

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual			<div> História = Boa Dívida = Alta Garantias = Nenhuma Renda = > 35 </div> <div>Risco de crédito?</div>
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7	

Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

SOMA Risco de crédito = 0.0079 + 0.0052 + 0.0514 = 0.0645

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

História = Ruim
 Dívida = Alta
 Garantias = Adequada
 Renda = < 15

Risco de crédito?

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

História = Ruim
 Dívida = Alta
 Garantias = Adequada
 Renda = < 15

Risco de crédito?

$$P(\text{Alto}) = 6/14 * 3/6 * 4/6 * 0 * 3/6$$

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

História = Ruim
 Dívida = Alta
 Garantias = Adequada
 Renda = < 15

Risco de crédito?

$$P(\text{Alto}) = 6/14 * 3/6 * 4/6 * 0 * 3/6$$

$$P(\text{Moderado}) = 3/14 * 1/3 * 1/3 * 1/3 * 0$$

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

História = Ruim
 Dívida = Alta
 Garantias = Adequada
 Renda = < 15

Risco de crédito?

$$P(\text{Alto}) = 6/14 * 3/6 * 4/6 * 0 * 3/6$$

$$P(\text{Moderado}) = 3/14 * 1/3 * 1/3 * 1/3 * 0$$

$$P(\text{Baixo}) = 5/14 * 0 * 2/5 * 2/5 * 0$$

Risco de Crédito	História de crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	<15 3	>=15 <=35 4	>35 7
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

História = Ruim
 Dívida = Alta
 Garantias = Adequada
 Renda = < 15

Risco de crédito?

Correção Laplaciana

$$P(\text{Alto}) = 6/14 * 3/6 * 4/6 * 0 * 3/6$$

$$P(\text{Moderado}) = 3/14 * 1/3 * 1/3 * 1/3 * 0$$

$$P(\text{Baixo}) = 5/14 * 0 * 2/5 * 2/5 * 0$$

Árvores de decisão

Vantagens e desvantagens

- Vantagens

- Fácil interpretação
- Não precisa normalização ou padronização (visto mais à frente)
- Rápido para classificar novos registros

- Desvantagens

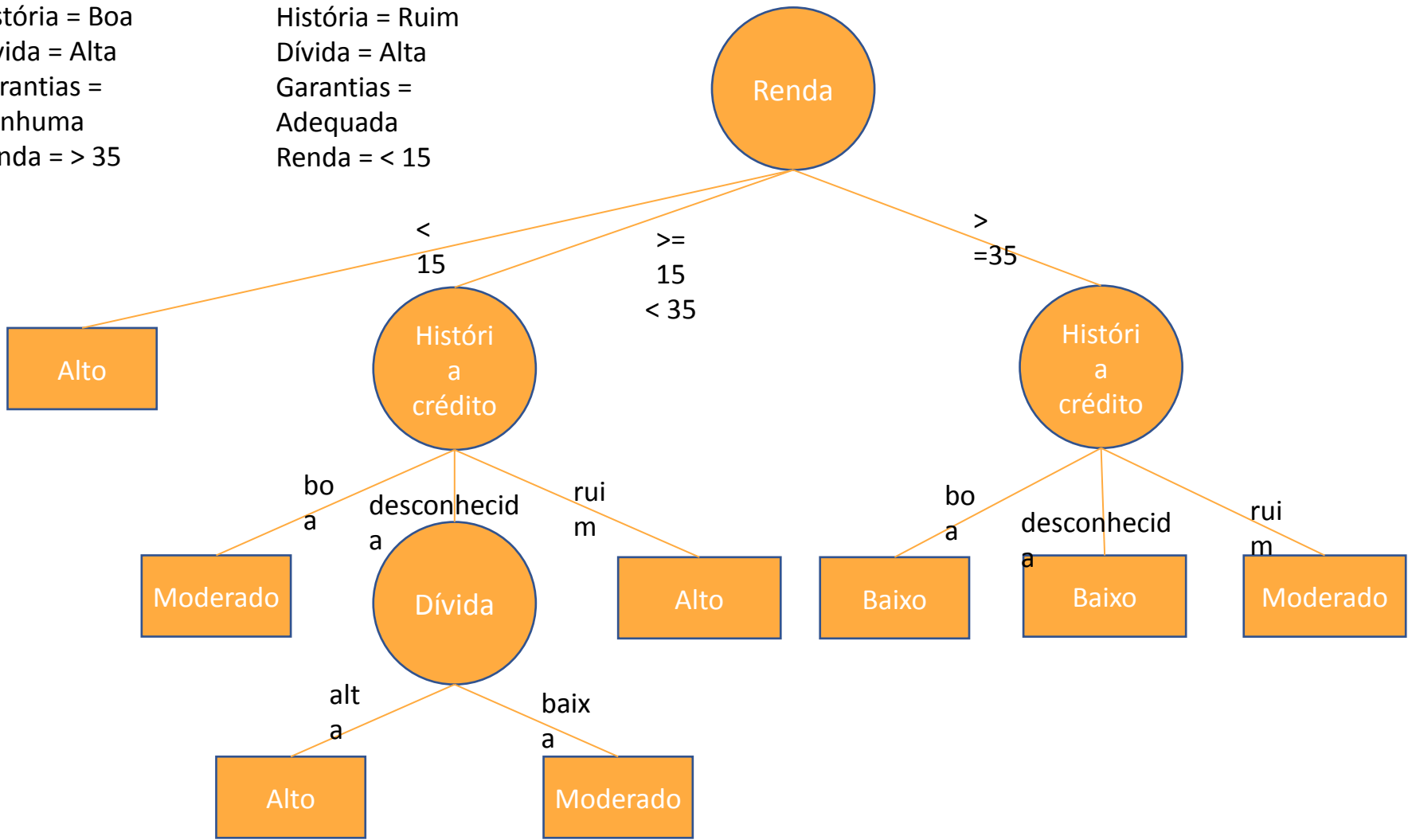
- Geração de árvores muito complexas
- Pequenas mudanças nos dados pode mudar a árvore (poda pode ajudar)

Base original

História do crédito	Dívida	Garantias	Renda anual	Risco
Ruim	Alta	Nenhuma	< 15.000	Alto
Desconhecida	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto
Desconhecida	Baixa	Nenhuma	>= 15.000 a <= 35.000	Moderado
Desconhecida	Baixa	Nenhuma	> 35.000	Alto
Desconhecida	Baixa	Nenhuma	> 35.000	Baixo
Desconhecida	Baixa	Adequada	> 35.000	Baixo
Ruim	Baixa	Nenhuma	< 15.000	Alto
Ruim	Baixa	Adequada	> 35.000	Moderado
Boa	Baixa	Nenhuma	> 35.000	Baixo
Boa	Alta	Adequada	> 35.000	Baixo
Boa	Alta	Nenhuma	< 15.000	Alto
Boa	Alta	Nenhuma	>= 15.000 a <= 35.000	Moderado
Boa	Alta	Nenhuma	> 35.0000	Baixo

História = Boa
Dívida = Alta
Garantias =
Nenhuma
Renda = > 35

História = Ruim
Dívida = Alta
Garantias =
Adequada
Renda = < 15



$$Entropy(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

***entropia:**

- medida da quantidade de incerteza que há em uma certa quantidade de dados
- medida da quantidade de informação não-óbvia presente nesses dados.

Risco
Alto
Alto
Moderado
Alto
Baixo
Baixo
Alto
Moderado
Baixo
Baixo
Alto
Moderado
Baixo
Alto

Alto = 6/14

Moderado = 3/14

Baixo = 5/14

$$Entropy(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

$$E(s) = -6/14 * \log_2(6/14) - 3/14 * \log_2(3/14) - 5/14 * \log_2(5/14) = \mathbf{1,53}$$

História do crédito	Risco
Ruim	Alto
Desconhecida	Alto
Desconhecida	Moderado
Desconhecida	Alto
Desconhecida	Baixo
Desconhecida	Baixo
Ruim	Alto
Ruim	Moderado
Boa	Baixo
Boa	Baixo
Boa	Alto
Boa	Moderado
Boa	Baixo
Ruim	Alto

$$Entropy(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

História de crédito
14

Bo
8/14

Desconhecida
5/14

Ruim
1/14

Alto
1/5

Moderado
1/5

Baixo
3/5

Alto
2/5

Moderado
1/5

Baixo
2/5

Alto
3/4

Moderado
1/4

Baixo
0

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

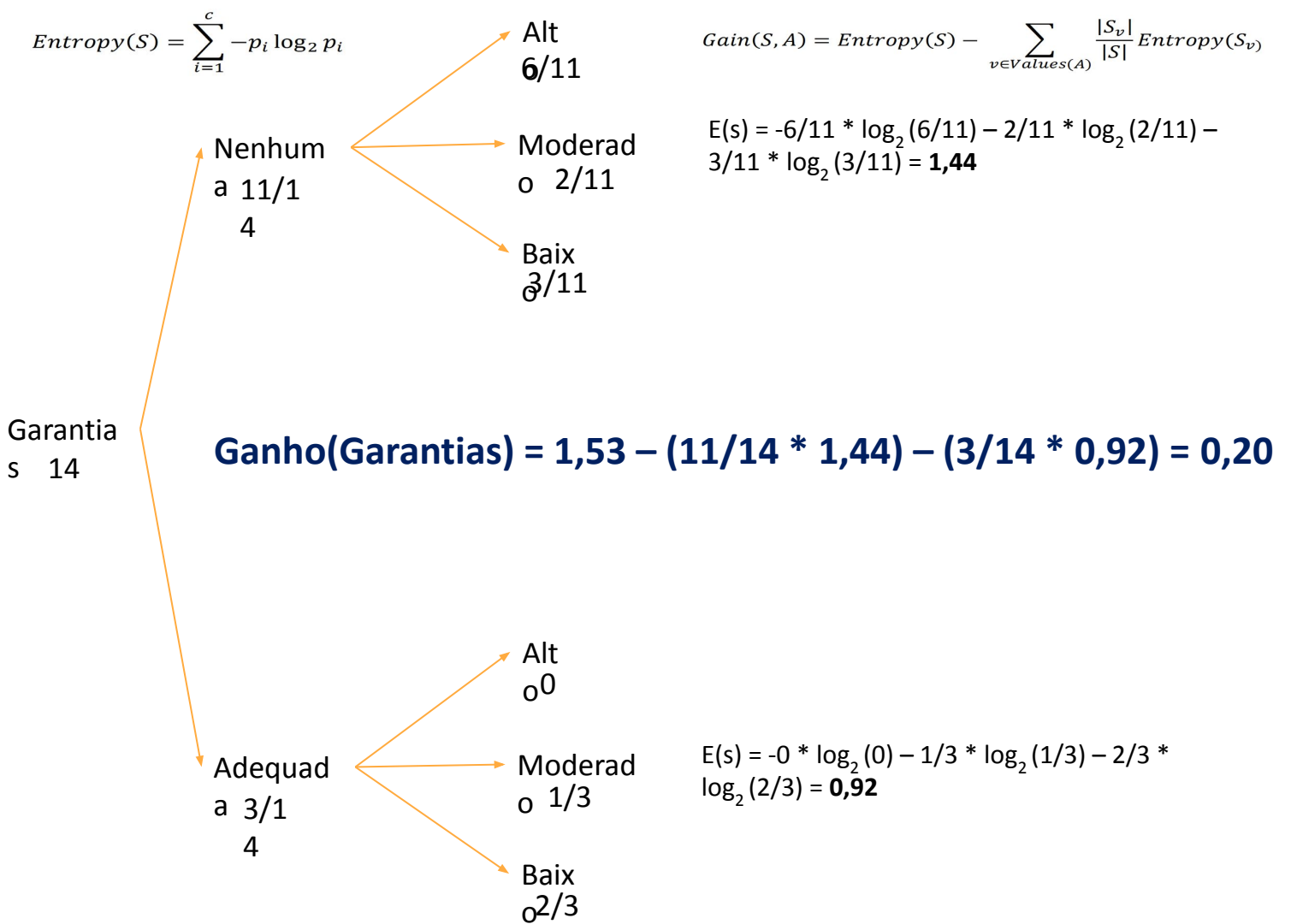
$$E(s) = -1/5 * \log_2(1/5) - 1/5 * \log_2(1/5) - 3/5 * \log_2(3/5) = 1,37$$

$$E(s) = -2/5 * \log_2(2/5) - 1/5 * \log_2(1/5) - 2/5 * \log_2(2/5) = 1,52$$

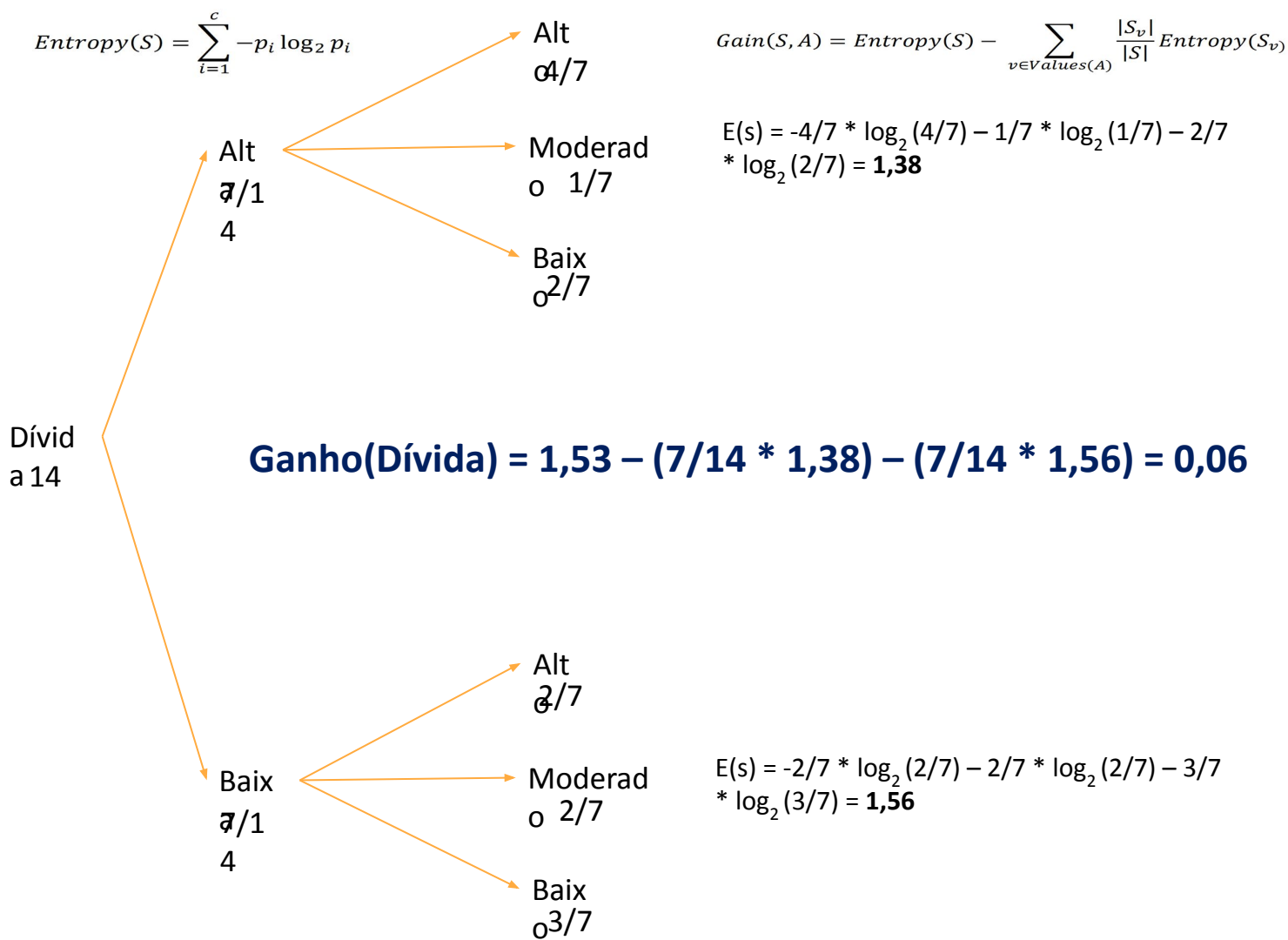
$$Ganho(História) = 1,53 - (5/14 * 1,37) - (5/14 * 1,52) - (4/14 * 0,81) = 0,26$$

$$E(s) = -3/4 * \log_2(3/4) - 1/4 * \log_2(1/4) - 0 * \log_2(0) = 0,81$$

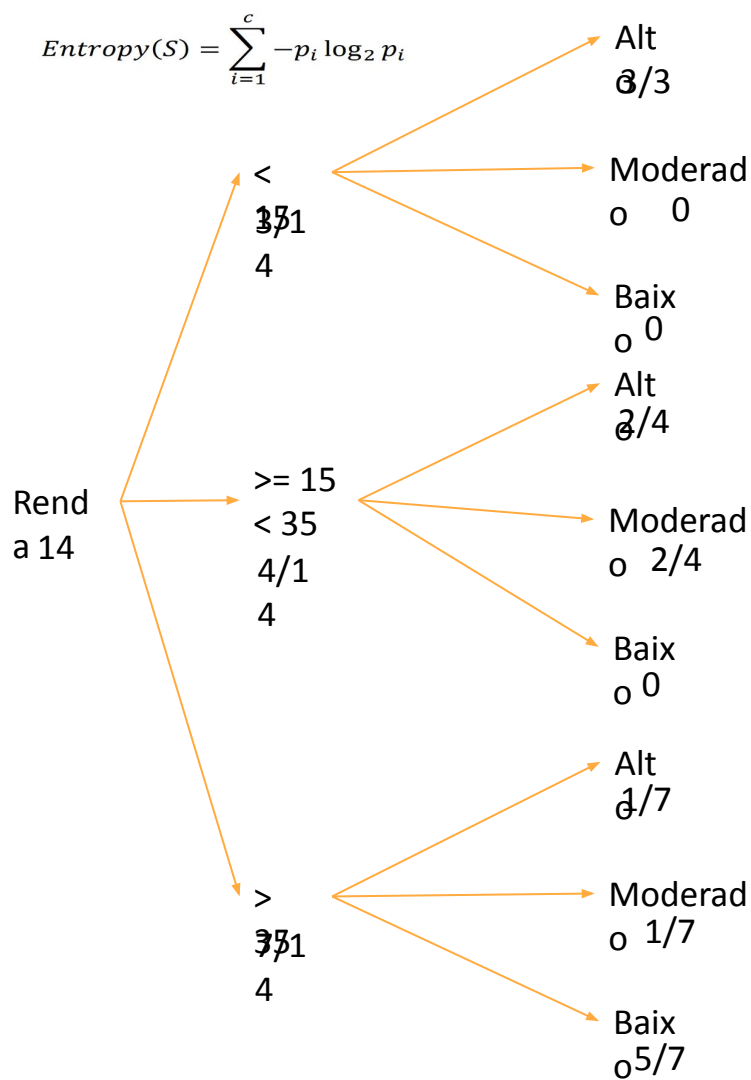
Garantias	Risco
Nenhuma	Alto
Nenhuma	Alto
Nenhuma	Moderado
Nenhuma	Alto
Nenhuma	Baixo
Adequada	Baixo
Nenhuma	Alto
Adequada	Moderado
Nenhuma	Baixo
Adequada	Baixo
Nenhuma	Alto
Nenhuma	Moderado
Nenhuma	Baixo



Dívida	Risco
Alta	Alto
Alta	Alto
Baixa	Moderado
Baixa	Alto
Baixa	Baixo
Baixa	Baixo
Baixa	Alto
Baixa	Moderado
Baixa	Baixo
Alta	Baixo
Alta	Alto
Alta	Moderado
Alta	Baixo
Alta	Alto



Renda anual	Risco
< 15.000	Alto
>= 15.000 a <= 35.000	Alto
>= 15.000 a <= 35.000	Moderado
> 35.000	Alto
> 35.000	Baixo
> 35.000	Baixo
< 15.000	Alto
> 35.000	Moderado
> 35.000	Baixo
> 35.000	Baixo
< 15.000	Alto
>= 15.000 a <= 35.000	Moderado
> 35.0000	Baixo
>= 15.000 a <= 35.000	Alto



$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$E(s) = -3/3 * \log_2(3/3) - 0 * \log_2(0) - 0 * \log_2(0) = \mathbf{0,00}$$

$$E(s) = -2/4 * \log_2(2/4) - 2/4 * \log_2(2/4) - 0 * \log_2(0) = \mathbf{1,00}$$

$$\mathbf{Ganho(Renda) = 1,53 - (3/14 * 0,00) - (4/14 * 1,00) - (7/14 * 1,15) = 0,66}$$

$$E(s) = -1/7 * \log_2(1/7) - 1/7 * \log_2(1/7) - 5/7 * \log_2(5/7) = \mathbf{1,15}$$

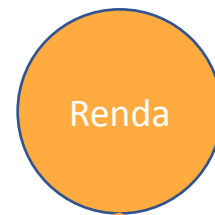
Risco
Alto
Alto
Moderado
Alto
Baixo
Baixo
Alto
Moderado
Baixo
Baixo
Alto
Moderado
Baixo
Alto

$$\text{Ganho(História)} = 1,53 - (5/14 * 1,37) - (5/14 * 1,52) - (4/14 * 0,81) = 0,26$$

$$\text{Ganho(Dívida)} = 1,53 - (7/14 * 1,38) - (7/14 * 1,56) = 0,06$$

$$\text{Ganho(Garantias)} = 1,53 - (11/14 * 1,44) - (3/14 * 0,92) = 0,20$$

$$\text{Ganho(Renda)} = 1,53 - (3/14 * 0,00) - (4/14 * 1,00) - (7/14 * 1,15) = 0,66$$



<
15

>=
15
< 35

>
35

História do crédito	Dívida	Garantias	Renda anual	Risco
Ruim	Alta	Nenhuma	< 15.000	Alto
Ruim	Baixa	Nenhuma	< 15.000	Alto
Boa	Alta	Nenhuma	< 15.000	Alto

História do crédito	Dívida	Garantias	Renda anual	Risco
Desconhecida	Baixa	Nenhuma	> 35.000	Alto
Desconhecida	Baixa	Nenhuma	> 35.000	Baixo
Desconhecida	Baixa	Adequada	> 35.000	Baixo
Ruim	Baixa	Adequada	> 35.000	Moderado
Boa	Baixa	Nenhuma	> 35.000	Baixo
Boa	Alta	Adequada	> 35.000	Baixo
Boa	Alta	Nenhuma	> 35.000	Baixo

História do crédito	Dívida	Garantias	Renda anual	Risco
Desconhecida	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto
Desconhecida	Baixa	Nenhuma	>= 15.000 a <= 35.000	Moderado
Boa	Alta	Nenhuma	>= 15.000 a <= 35.000	Moderado
Ruim	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto

Random Forest (floresta randômica)

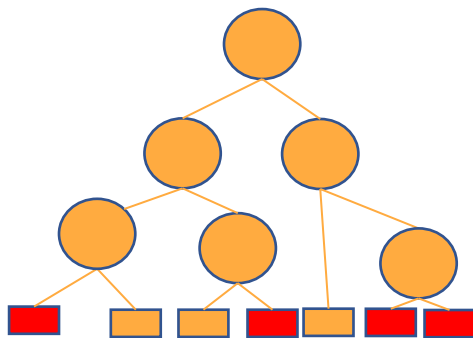


Random Forest

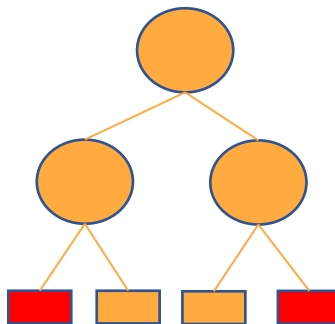
- Ensemble learning (aprendizagem em conjunto)

- “Consultar diversos profissionais para tomar uma decisão”
- Vários algoritmos juntos para construir um algoritmo mais “forte”
- Usa a média (regressão) ou votos da maioria (classificação) para dar a resposta final

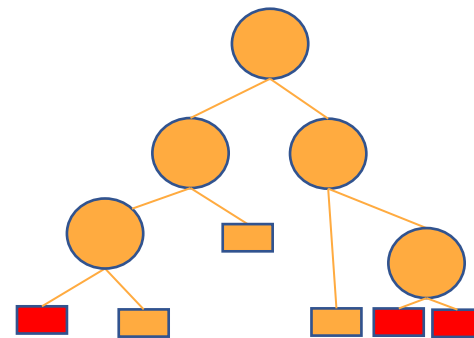
História do crédito	Dívida	Garantias	Renda anual	Risco
Ruim	Alta	Nenhuma	< 15.000	Alto
Desconhecida	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto
Desconhecida	Baixa	Nenhuma	>= 15.000 a <= 35.000	Moderado
Desconhecida	Baixa	Nenhuma	> 35.000	Alto
Desconhecida	Baixa	Nenhuma	> 35.000	Baixo
Desconhecida	Baixa	Adequada	> 35.000	Baixo
Ruim	Baixa	Nenhuma	< 15.000	Alto
Ruim	Baixa	Adequada	> 35.000	Moderado
Boa	Baixa	Nenhuma	> 35.000	Baixo
Boa	Alta	Adequada	> 35.000	Baixo
Boa	Alta	Nenhuma	< 15.000	Alto
Boa	Alta	Nenhuma	>= 15.000 a <= 35.000	Moderado
Boa	Alta	Nenhuma	> 35.000	Baixo
Ruim	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto



Risco =
Alto



Risco =
Baixo



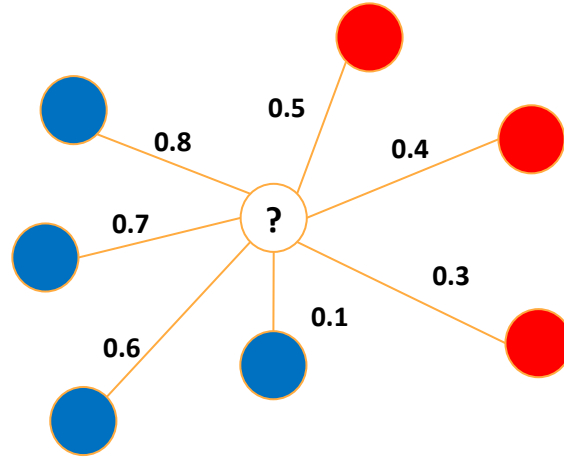
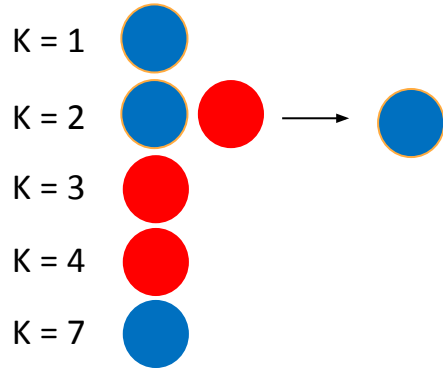
Risco =
Baixo

KNN
(vizinhos mais próximos)

kNN

- Algoritmo simples e poderoso
- Indicado quando o relacionamento entre as características é complexo
- Valor de k pequeno: dados com ruídos ou outliers podem prejudicar
- Valor de k grande: tendência a classificar a classe com mais elementos (overfitting) – valor default 3 ou 5
- Lento para fazer as previsões
- Outras distâncias
 - Coeficiente de Pearson
- Índice de Tanimoto
 - City Block

K-Nearest Neighbour (kNN)



KNN

- A maioria dos métodos de aprendizagem constroem um modelo após o treinamento (os dados são descartados após a criação do modelo)
- Métodos baseados em instâncias simplesmente armazenam os exemplos de treinamento
- A generalização/previsão é feita somente quando uma nova instância precisa ser classificada (lazy)
- Paradigmas de aprendizagem de máquina

$$DE(x, y) = \sqrt{\sum_i^p (x_i - y_i)^2}$$

- $X = 5, 7, 9$
- $Y = 5, 5, 5$
- Subtração de cada posição do vetor
 - $5 - 5 = 0$
 - $7 - 5 = 2$
 - $9 - 5 = 4$
- Elevação ao quadrado
 - 0^2
 - 2^2
 - 4^2
- Somatório
 - $0 + 4 + 16 = 20$
- Raiz quadrada (20)
 - 4,47
- **Distância Euclidiana = 4,47**

História do crédito	Dívida	Garantias	Renda anual	Risco
Ruim	Alta	Nenhuma	< 15.000	Alto
Desconhecida	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto
Desconhecida	Baixa	Nenhuma	>= 15.000 a <= 35.000	Moderado
Desconhecida	Baixa	Nenhuma	> 35.000	Alto
Desconhecida	Baixa	Nenhuma	> 35.000	Baixo
Desconhecida	Baixa	Adequada	> 35.000	Baixo
Ruim	Baixa	Nenhuma	< 15.000	Alto
Ruim	Baixa	Adequada	> 35.000	Moderado
Boa	Baixa	Nenhuma	> 35.000	Baixo
Boa	Alta	Adequada	> 35.000	Baixo
Boa	Alta	Nenhuma	< 15.000	Alto
Boa	Alta	Nenhuma	>= 15.000 a <= 35.000	Moderado
Boa	Alta	Nenhuma	> 35.0000	Baixo
Ruim	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto

História do crédito	Dívida	Garantias	Renda anual	Risco
3	1	1	1	Alto
2	1	1	2	Alto
2	2	1	2	Moderado
2	2	1	3	Alto
2	2	1	3	Baixo
2	2	2	3	Baixo
3	2	1	1	Alto
3	2	2	3	Moderado
1	2	1	3	Baixo
1	1	2	3	Baixo
1	1	1	1	Alto
1	1	1	2	Moderado
1	1	1	3	Baixo
3	1	1	2	Alto

História = Boa (1)

Dívida = Alta (1)

Garantias = Nenhuma (1)

Renda = > 35 (3)

Novo x 9º

1 1 1 3

1 2 1 3

$0 + 1^2 + 0 + 0$

$0 + 1 + 0 + 0 = 1$

Raiz(1) = **1**

Novo x 3º

1 1 1 3

2 2 1 2

$1^2 + 1^2 + 0 + 1^2$

$1 + 1 + 0 + 1 = 3$

Raiz(3) = **1,7**

$$DE(x, y) = \sqrt{\sum_i^p (x_i - y_i)^2}$$

kNN – variáveis na mesma escala

Idade	Renda anual
60	30.000
65	75.000
20	29.500

$$DE(x, y) = \sqrt{\sum_i^p (x_i - y_i)^2}$$

1º x 2º

60 30.000

65 75.000

$$5^2 + 45.000^2$$

$$25 + 2.025.000.000 = 2.025.000.025$$

$$\text{Raiz}(2.025.000.000) = \mathbf{45.000}$$

1º x 3º

60 30.000

20 29.500

$$40^2 + 500^2$$

$$1.600 + 250.000 = 251.600$$

$$\text{Raiz}(251.600) = \mathbf{501,59}$$

Normalização (Normalization)

$$x = \frac{x - \text{mínimo}(x)}{\text{máximo}(x) - \text{mínimo}(x)}$$

Idade	Renda anual
60	30.000
35	45.000
20	29.500

Idade	Renda anual
0,80	0,03
0,30	1,00
0,00	0,00

$$x = \frac{60 - 20}{65 - 20} = 0,80 \quad x = \frac{30.000 - 29.500}{45.000 - 29.500} = 0,03$$

$$x = \frac{35 - 20}{65 - 20} = 0,30 \quad x = \frac{45.000 - 29.500}{45.000 - 29.500} = 1,00$$

$$x = \frac{20 - 20}{65 - 20} = 0,00 \quad x = \frac{29.500 - 29.500}{45.000 - 29.500} = 0,00$$

1º x 2º

0,80 0,03

0,30 1,00

$0,50^2 + 0,97^2$

$0,25 + 0,940 = 1,19$

Raiz(1,19) = **1,09**

1º x 3º

0,80 0,03

0,00 0,00

$0,80^2 + 0,03^2$

$0,64 + 0,0009 = 0,6409$

Raiz(0,6409) = **0,80**

Padronização (Standardization)

$$x = \frac{x - \text{média}(x)}{\text{desvio padrão}(x)}$$

$$x = \frac{60 - 38,33}{20,20} = 1,07 \quad x = \frac{30.000 - 34.833,33}{8.808,14} = -0,54$$

$$x = \frac{35 - 38,33}{20,20} = -0,16 \quad x = \frac{45.000 - 34.833,33}{8.808,14} = 1,15$$

$$x = \frac{20 - 38,33}{20,20} = -0,90 \quad x = \frac{29.500 - 34.833,33}{8.808,14} = -0,60$$

Idade	Renda anual	Idade	Renda anual
60	30.000	1,07	-0,54
35	45.000	-0,16	1,15
20	29.500	-0,90	-0,60

Idade

Média = 38,33

Desvio padrão = 20,20

1º x 2º

0,80 0,03

0,30 1,00

$0,50^2 + 0,97^2$

$0,25 + 0,940 = 1,19$

Raiz(1,19) = **1,09**

Renda

Média = 34.833,33

Desvio padrão = 8.808,14

1º x 3º

1,07 -0,54

-0,90 -0,60

$1,97^2 + 0,006^2$

$3,88 + 0,000036 = 3,880036$

Raiz(3,880036) = **1,96**



OBRIGADO!

Guilherme Feulo

PROFESSOR OCEAN SP

guilherme.feulo@oceanbrasil.com

