

# Inteligência Artificial

Simbólica

**Conhecimento representado por símbolos e uso da lógica para fazer inferência**

Busca e Otimização

**Espaço de soluções é percorrido preferencialmente usando uma heurística para encontrar normalmente uma solução ótima local**

Memória

**Conhecimento é armazenado no modelo e a inferência é realizada olhando esses dados memorizados**

Conexionista

**Modelo é construído compondo conexões entre estruturas menores, conhecimento é armazenado nos parâmetros internos do modelo após fase de aprendizado**

Aprendizado

Não-Supervisionado

Supervisionado

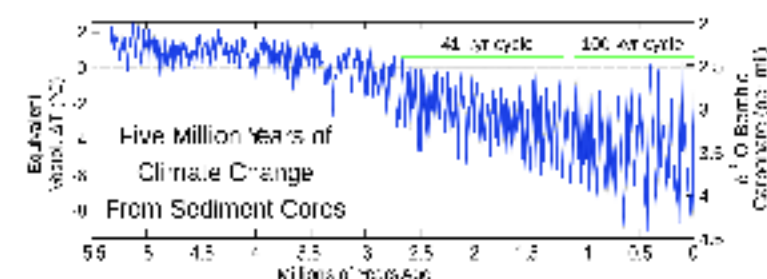
Reforço

# Dados

Estruturados

Não-estruturados

time	<u>n.risk</u>	<u>n.event</u>	survival	<u>std.err</u>	lower 95% CI	upper 95% CI
5	23	2	0.913	0.0588	0.8049	1
8	21	2	0.8261	0.079	0.6848	0.996
9	19	1	0.7826	0.086	0.631	0.971
12	18	1	0.7391	0.0916	0.5798	0.942
13	17	1	0.6957	0.0959	0.5309	0.912
18	14	1	0.646	0.1011	0.4753	0.878
23	13	2	0.5466	0.1073	0.3721	0.803
27	11	1	0.4969	0.1084	0.324	0.762
30	9	1	0.4417	0.1095	0.2717	0.718
31	8	1	0.3865	0.1089	0.2225	0.671
33	7	1	0.3313	0.1064	0.1765	0.622
34	6	1	0.2761	0.102	0.1338	0.569
43	5	1	0.2208	0.0954	0.0947	0.515
45	4	1	0.1656	0.086	0.0598	0.458
48	2	1	0.0828	0.0727	0.0148	0.462



Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt. Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt.



# Formato atributo-valor

Representação de conjunto de dados

■ Formados por **objetos**

■ Cada objeto corresponde a uma ocorrência dos dados

		Sintomas				doente
		temperatura	dor		pressão	
Objetos	paciente <sub>1</sub>	38°C	sim	...	12.7	Sim
	paciente <sub>2</sub>	36°C	não	...	12.7	Não
				:		
	paciente <sub>n</sub>	40°C	não	...	14	Sim

# Formato atributo-valor

■ Cada objeto é descrito por um conjunto de atributos de entrada (**Vetor de características**)

■ Cada atributo está associado a uma propriedade do objeto

		Sintomas				
		temperatura	dor	pressão	doente	
Dados	paciente <sub>1</sub>	38°C	sim	...	12.7	Sim
	paciente <sub>2</sub>	36°C	não	...	12.7	Não
				:		
	paciente <sub>n</sub>	40°C	não	...	14	Sim

# Tipos de atributos

## Quantitativo (numérico)

Representa quantidades

Valores podem ser ordenados e usados em operações aritméticas

Podem ser **contínuos** ou **discretos**

Possuem unidade associada

## Qualitativo (simbólico ou categórico)

Representa qualidades

Valores podem ser associados a categorias

Alguns podem ser ordenados, mas operações aritméticas não são aplicáveis

*Ex. {pequeno, médio, grande}*



# Tipos de atributos

## Atributos Quantitativos

### Contínuos

- Podem assumir um número infinito de valores
- Geralmente resultados de medidas
- Frequentemente representados por números reais
- *Ex. peso, distância*

### Discretos

- Número finito ou infinito contável de valores
- Caso especial: atributos binários (booleanos)
- *Ex. {12, 23, 45}, {0, 1}*

## Qualitativo, Quantitativo Discreto ou Contínuo?

<b>Id.</b>	<b>Nome</b>	<b>Idade</b>	<b>Sexo</b>	<b>Peso</b>	<b>Manchas</b>	<b>Temp.</b>	<b># Int.</b>	<b>Est.</b>	<b>Diagnóstico</b>
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

# Tipos de atributos

## ■ Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

Alguns atributos qualitativos são representados por números, mas não faz sentido a utilização de operadores aritméticos sobre seus valores

Qualitativo

Quantitativo discreto

Quantitativo contínuo



# Escala de atributos

■ Define operações que podem ser realizadas sobre os valores dos atributos

■ Nominais

■ Ordinais

**Qualitativos**

■ Intervalares

■ Racionais

**Quantitativos**

# Escalas de atributos Qualitativos

## Escala nominal

- Valores são nomes diferentes e carregam a menor quantidade de informação possível
- Não existe relação de ordem entre os valores
- **Operações aplicáveis:** =,  $\neq$
- *Ex.: número de conta em banco, cores, sexo*

## Escala ordinal

- Valores refletem ordem das categorias representadas
- **Operações aplicáveis:** =,  $\neq$ , <, >,  $\leq$ ,  $\geq$
- *Ex.: hierarquia militar, avaliações qualitativas de temperatura*

# Escalas de atributos Quantitativos

## Escala intervalar

- Números que variam em um intervalo
- É possível definir ordem e diferença em magnitude entre dois valores
- Origem da escala definida de maneira arbitrária
- Operações aplicáveis:  $=$ ,  $\neq$ ,  $<$ ,  $>$ ,  $\leq$ ,  $\geq$ ,  $+$ ,  $-$
- Ex.: temperatura em  $^{\circ}\text{C}$  ou  $^{\circ}\text{F}$ , datas

## Escala racional

- Carregam mais informações
- Têm significado absoluto (existe 0 absoluto)
- Razão tem significado
- Operações aplicáveis:  $=$ ,  $\neq$ ,  $<$ ,  $>$ ,  $\leq$ ,  $\geq$ ,  $+$ ,  $-$ ,  $*$ ,  $/$
- Ex.: tamanho, distância, salário, saldo em conta

<b>Id.</b>	<b>Nome</b>	<b>Idade</b>	<b>Sexo</b>	<b>Peso</b>	<b>Manchas</b>	<b>Temp.</b>	<b># Int.</b>	<b>Est.</b>	<b>Diagnóstico</b>
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

# Exercício

## ■ Definir o tipo e escala dos seguintes atributos:

tipo: (quantitativo [contínuo ou discreto], qualitativo [categórico])

escala: (Nominal, ordinal, intervalar e racional)

- Renda mensal:
- Número de palavras de um texto:
- Número de matrícula:
- Data de nascimento:
- Código postal:
- Posição em uma corrida:



# Exercício

## ■ Definir o tipo e escala dos seguintes atributos:

tipo: (quantitativo [contínuo ou discreto], qualitativo [categórico])

escala: (Nominal, ordinal, intervalar e racional)

- Renda mensal: quantitativo contínuo | racional
- Número de palavras de um texto: quantitativo discreto | racional
- Número de matrícula: qualitativo | nominal
- Data de nascimento: quantitativo discreto | intervalar
- Código postal: qualitativo | nominal
- Posição em uma corrida: qualitativo | ordinal

# Exploração de dados

- **Estatística descritiva:** resumo quantitativo das principais características de um conjunto de dados
  - Muitas medidas podem ser calculadas rapidamente
  - Captura de informações como:
    - Frequência
    - Localização ou tendência central
    - Dispersão ou espalhamento
    - Distribuição ou formato

Informações obtidas podem ajudar na seleção de técnicas apropriadas de pré-processamento e aprendizado

# Exploração de dados

## Frequência

- Proporção de vezes que um atributo assume um dado valor
- Aplicável a valores numéricos e simbólicos
- *Ex.: 40% dos pacientes têm febre*

## Localização, dispersão e distribuição

- Diferem para dados **univariados** e **multivariados**
  - *Maioria dos dados em AM é multivariado, mas análises em cada atributo podem fornecer informações valiosas*
- Geralmente aplicados a valores numéricos

# Frequência

■ Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

Frequência: 25% das manchas são médias

# Dados univariado

- Objetos com apenas um atributo

- Conjunto com  $n$  objetos  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$

**Observação:** termo conjunto não tem o mesmo significado do usado em teoria dos conjuntos

*Em um conjunto de dados, o mesmo valor pode aparecer mais de uma vez em um atributo*



# Dados univariados: medidas de localidade

- Definem pontos de **referência** nos dados
  - Valor “típico”, resume os dados

## Valores numéricos

- **Média**
- **Mediana**
- **Percentil**

## Valores simbólicos

- **Moda**: valor mais frequente

# Moda

## ■ Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

