Inteligência Artificial

Preparação e Pré-processamento dos dados
Parte II

Prof. Dr. Ivan Carlos Alcântara de Oliveira

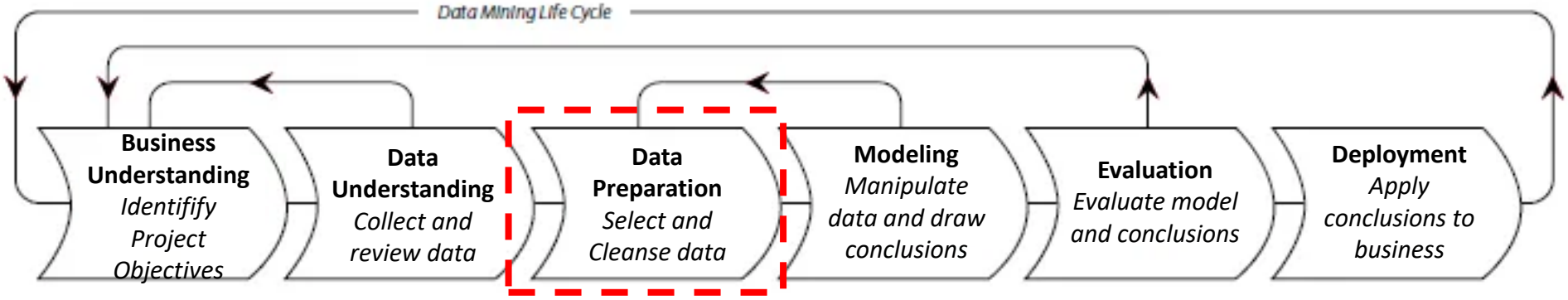
https://orcid.org/0000-0002-6020-7535

Preparação dos dados e Pré-Processamento

Ciclo de Vida de Projeto de Data Science

CRISP-DM

- **CRISP-DM (*Cross Industry Standard Process for Data Mining* -** Processo Padrão Interindústrias para Mineração de Dados) é um processo de fases bastante aceito na indústria para representar um ciclo completo de Ciência e Análise de Dados, incluído a aplicação de modelos de Aprendizado de Máquina.



Auxiliar na Identificação	Obtenção dos dados	Pré-processamento	Seleção do modelo	Validação do modelo	Aplica solução obtida ao negócio
Compreensão	Integração	Transformação de Variáveis	Cross-Validation	Otimização	Desenvolvimento de Aplicação Analítica
Pesquisas e Estudos	Análise Exploratória	Seleção de Atributos	Métricas de performance		Geração de relatório
Reunião com Especialistas		Redução de Dimensionalidade	Otimização		Construção de Dashborads etc.
Definição dos Atributos		Amostragem			
Definição medida de erro					

Ciência de Dados

Ciclo de Vida de Projeto de *Data Science*

3. **Preparação dos Dados**

Data Wrangling: pré-processamento, transformação, limpeza dos dados, etc.

Até aqui, aproximadamente 70% do tempo de um Cientista de Dados.

Preparação dos Dados

- O **desempenho** dos algoritmos de AM geralmente são **afetado**s pelo **estado dos dados**.
  - Conjunto de dados **podem apresentar diferentes características**, **dimensões** ou **formatos**.
  - Por exemplo, podem ser numéricos ou categóricos/simbólicos.

Preparação dos Dados

- **Conjunto de dados podem também**:
  - Estar limpos ou **conter ruídos** e imperfeições
  - Com valores incorretos
  - **Inconsistentes**
  - **Duplicados ou ausentes**
  - Ter atributos independentes ou relacionados
  - **Apresentar poucos ou muitos objetos**
  - Ter um **número pequeno ou elevado de atributos**.

Preparação dos Dados

- **Técnicas de Preparação dos** dados são frequentemente **utilizadas** para **melhorar** a **qualidade dos dados** por meio da eliminação ou minimização desses problemas.

Preparação dos Dados

- Essa melhora pode:
  - **facilitar o uso de técnicas de AM**
  - **permitir a construção de modelos mais fiéis à distribuição real dos dados**
  - **Reduzir a complexidade computacional**
  - **Tornar mais fáceis e rápidos o ajuste de parâmetros** do modelo e seu posterior uso.
  - **Facilitar a interpretação dos padrões extraídos pelo modelo**

Preparação dos Dados

- Técnicas de Preparação dos dados também **podem tornar os dados mais adequados para** sua utilização por um determinado **algoritmo de AM**.
- **Por exemplo**, **alguns** algoritmos de AM **trabalham apenas com valores numéricos**.
- **Não existe** uma **ordem** fixa **para aplicação** das diferentes **técnicas** de **pré-processamento**.

Preparação dos Dados

Transformação de Dados

- **Várias técnicas de AM estão limitadas à manipulação de valores de determinados tipos**, por exemplo, apenas valores **numéricos** ou apenas valores **simbólicos/categóricos**.

- **Algumas técnicas têm seu desempenho influenciado pelo intervalo de variação dos valores numéricos**.

Preparação dos Dados

Transformação de Dados

- As diferentes **técnicas de transformação** podem ser **divididas em**:
  - **Normalização:** modifica a escala de valores quantitativos.
  - **Conversão de Valores simbólicos para numéricos**: se os valores simbólicos são nominais ou ordinais, diferentes técnicas podem ser empregadas.
  - **Conversão de Valores numéricos para simbólicos**.
  - **Mudança de escala ou de intervalo de valores**.

Preparação dos Dados

Transformação de Dados - **Normalização**

- É um processo de **transformação dos dados** que **objetiva torná-los mais apropriados** à aplicação de algum **algoritmo de mineração**, como redes neurais artificiais ou métodos baseados em distância, como o KNN.
- Motivo: evitar a saturação dos neurônios em uma rede neural artificial de múltiplas camadas e fazer com que cada atributo dos dados de entrada tenha o mesmo domínio.

Preparação dos Dados

Transformação de Dados - **Normalização**

- O **objetivo** da normalização é **alterar os valores das colunas numéricas (variáveis quantitativas)** no conjunto de dados **para uma escala comum**, sem distorcer as diferenças nos intervalos de valores.
- Dependendo do *dataset* devem ser separadas somente as variáveis quantitativas e então aplicar a normalização.

Preparação dos Dados

Transformação de Dados - **Normalização**

- Para descobrir se algoritmo de Aprendizado de Máquina precisa que os dados estejam normalizados, basta verificar a sua documentação.

- A documentação da biblioteca Python *"scikitlearn"* fornece informações sobre a necessidade de normalização para alguns algoritmos.

Preparação dos Dados

Transformação de Dados - **Normalização**

- Não precisa aplicar normalização a todo conjunto de dados.
- É necessário apenas quando os atributos (variáveis) tiverem intervalos muito diferentes.
- A variável pode ser numérica, mas qualitativa. Então, não aplicar a normalização. Ficar atento!
- Após a normalização, os dados vão representar o mesmo tipo de informação, mas em uma escala diferente

Preparação dos Dados

Transformação de Dados – **Normalização - Exemplo**

- Considere um *dataset* contendo os atributos “idade” variando de 0 a 100 anos e “renda mensal” de 0 a 100.000.
- A “renda mensal” pode ser até 1.000 vezes maior do que “idade” e com uma variação de valores muito maior. Logo, esses dois atributos estão em intervalos muito diferentes.
- Quando análises adicionais são realizadas, como Regressão Linear Multivariada, por exemplo, a renda atribuída influenciará muito mais no resultado devido ao seu valor maior. E isso causa problemas durante o treinamento do algoritmo.

Preparação dos Dados

Transformação de Dados **- Normalização**

- As **técnicas** de **normalização são**:
  - **Max-Min:** se a distribuição não for gaussiana ou o desvio padrão for muito pequeno, esta técnica funciona melhor.
  - **Escore-z:** valores de um atributo são normalizados para que tenham distribuição normal com média = 0 e o desvio padrão = 1.
  - **Escalonamento decimal**
  - **Range interquartil**

Preparação dos Dados

Transformação de Dados – **Normalização – Max-Min**

- **Realiza uma transformação linear nos dados originais**.
- Assumir que max_a e min_a são, respectivamente, os valores máximo e mínimo de determinado atributo a.
- A normalização max-min **mapeia um valor a em um valor a' no domínio [$novo\_min_a$, $novo\_max_a$]**.
- A aplicação mais frequente dessa normalização é colocar todos os atributos de uma base de dados sob um mesmo intervalor de valores, ex. [0,1].

Preparação dos Dados

Transformação de Dados – **Normalização – Max-Min - Fórmula**

- Fórmula:

$$a' = \frac{a - min_a}{max_a - min_a}(novo\_max_a - novo\_min_a) + novo\_min_a$$

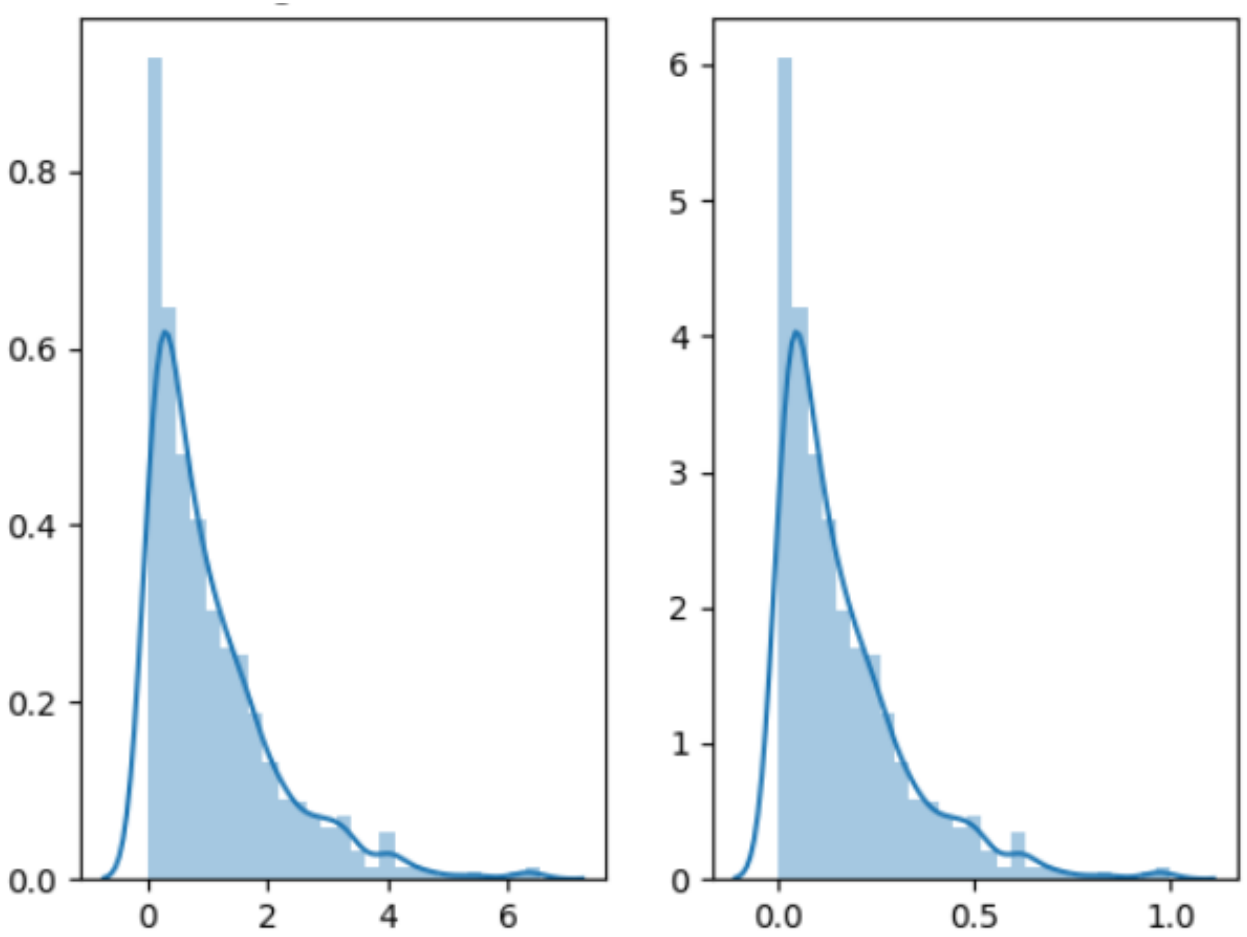
Exemplo: Normalização Max-Min para converter o atributo idade para o intervalo [0,1].
Id (a = 67)) => min_a = 28, max_a= 74,
$novo\_max_a$= 1, $novo\_min_a$ = 0. Então, Max-Min(67) = a'

$$a' = \frac{67 - 28}{74 - 28}(1 - 0) + 0 = \frac{39}{46} = 0{,}85$$

Id	Idade	Max-Min
1	**67**	**0,85**
2	43	0,33
3	58	0,65
4	28	0,00
5	74	1,00
6	65	0,80
7	70	0,91
8	42	0,30
9	57	0,63
10	60	0,70

Preparação dos Dados

Transformação de Dados – **Normalização – Max-Min – Exemplo Gráfico**



Observe no eixo x do gráfico acima como a escala dos dados é diferente, embora a distribuição desses dados seja a mesma.

Preparação dos Dados

Transformação de Dados – **Normalização – Max-Min**

- É **principalmente necessária** no caso de **algoritmos que usam medidas de distância** como agrupamento (clusterização), sistemas de recomendação que usam semelhança de cosseno, etc.

- Isto é feito de forma que **uma variável que está em uma escala maior não afeta o resultado apenas porque está em uma escala maior**.

Preparação dos Dados

Transformação de Dados – **Normalização – Max-Min**

- Algoritmos de *Machine Learning* que requerem a normalização dos dados:
  - KNN com medida de distância euclidiana se quiser que todos os atributos contribuam igualmente no modelo.
  - Regressão Logística, SVM, *Perceptrons*, Redes Neurais.
  - K-Means: que trabalha com agrupamento.
  - Análise discriminante linear, análise de componentes principais, análise de componentes principais do kernel.

Preparação dos Dados

Transformação de Dados – **Normalização – Max-Min**

- **Classificadores baseados em modelo gráfico**, como Fisher *Linear discriminant analysis* (LDA) ou *Naive Bayes*, bem como Árvores de Decisão e métodos baseados em árvore, como *Random Forest*, são invariantes ao dimensionamento de recursos, **mas ainda assim pode ser uma boa ideia redimensionar os dados**.

Preparação dos Dados

Transformação de Dados – **Normalização – Escore-z**

- Chamada de normalização de média zero ou também de **Padronização**.
- Os **valores de um atributo a são normalizados para que tenham distribuição normal com média = 0 e o desvio padrão = 1**.
- Esse método de normalização é **útil quando os valores máximo** e **mínimo** reais de um atributo **são desconhecidos ou quando há *outliers* dominando a normalização Max-Min**.
- Amplamente utilizado nos algoritmos ***Support Vector Machine*** (SVM), Regressao logistica e Redes Neurais.

Preparação dos Dados

Transformação de Dados **– Normalização – Escore-z - Fórmula**

- Fórmula:

$$a' = \frac{(a - \bar{a})}{\sigma_a} \quad onde\ \bar{a} = média\ e\ \sigma_a = desvio\ \ padrão\ de\ a$$

$$\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i \qquad \sigma(a) = \sqrt[2]{\frac{1}{n-1}\sum_{i=1}^{n}(a - \bar{a})^2}$$

Preparação dos Dados

Transformação de Dados – **Normalização – Escore-z - Exemplo**

Exemplo: Normalização Escore-z para converter o atributo idade.

Id (a = 67) => Escore-z(67)

$\bar{a}$ = 56,4

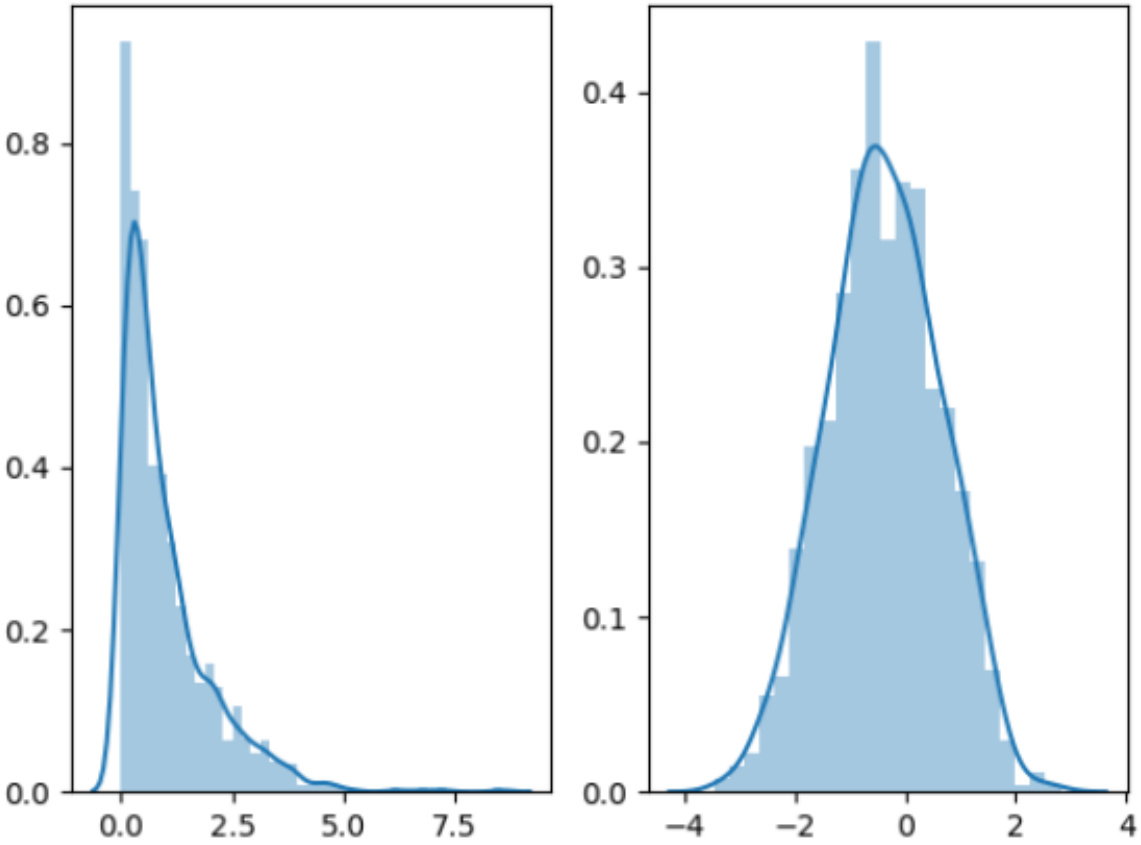
$\sigma(a)$ = 14,5

$$a' = \frac{(a - \bar{a})}{\sigma_a} = \frac{(67 - 56{,}4)}{14{,}5} = \frac{10{,}6}{14{,}5} = 0{,}73$$

Id	Idade	Escore-z
1	**67**	**0,73**
2	43	-0,92
3	58	0,11
4	28	-1,96
5	74	1,21
6	65	0,59
7	70	0,94
8	42	-0,99
9	57	0,04
10	60	0,25

Preparação dos Dados

Transformação de Dados – **Normalização – Escore-z – Exemplo Gráfico**



Observe no eixo x do gráfico acima como a distribuição dos dados segue uma distribuição normal depois de aplicado escore-z. Uma distribuição normal é caracterizada por média 0 e desvio padrão 1.

Pré-processamento de Dados

Transformação de Dados – **Normalização – Max-Min x Escore-z - Discussão**

- Se você tiver valores *outliers* em seu *dataset*, a transformação **Max-Min** aumentará os dados "normais" para um intervalo muito pequeno. E, geralmente, a maioria dos *datasets* tem *outliers*.

- Portanto, a **Max-Min** é geralmente evitada quando o *dataset* tem *outliers* (desde que inclua o valor máximo). Nesses casos, **Escore-z** é preferido.

Preparação dos Dados

Transformação de Dados – **Normalização – Max-Min x Escore-z - Discussão**

Algumas considerações importantes:

1. **Max-Min** torna o treinamento menos sensível à escala de atributos, permitindo resolver melhor os coeficientes.

2. O uso de um método **Max-Min** melhorará a análise de múltiplos modelos.

3. **Max-Min** assegurará que um problema de convergência não tenha uma variância massiva, tornando a otimização viável.

4. **Escore-z** tende a tornar o processo de treinamento bem melhor, porque a condição numérica dos problemas de otimização é melhorada.

Preparação dos Dados

Transformação de Dados – **Normalização – Escalonamento decimal**

- Move **a casa decimal dos valores do atributo a.**

- **O número de casas decimais movidas depende** do **valor máximo absoluto do atributo a.**

Preparação dos Dados

Transformação de Dados – **Normalização – Escalonamento decimal - Fórmula**

- Fórmula:

$$a' = \frac{a}{10^j}$$

onde j é *o menor inteiro tal que* $\max(|a'|) < 1$.

Preparação dos Dados

Transformação de Dados – **Normalização – Escalonamento decimal - Exemplo**

Exemplo: Normalização "Escalonamento decimal" para converter o atributo idade usando j = 2.

Id (a = 67)) =>

Escalonamento decimal(67)

$$a' = \frac{67}{10^2} = 0{,}67$$

Id	Idade	Escalonamento decimal
1	**67**	**0,67**
2	43	0,43
3	58	0,58
4	28	0,28
5	74	0,74
6	65	0,65
7	70	0,70
8	42	0,42
9	57	0,57
10	60	0,60

Preparação dos Dados

Transformação de Dados – **Normalização – range interquartil**

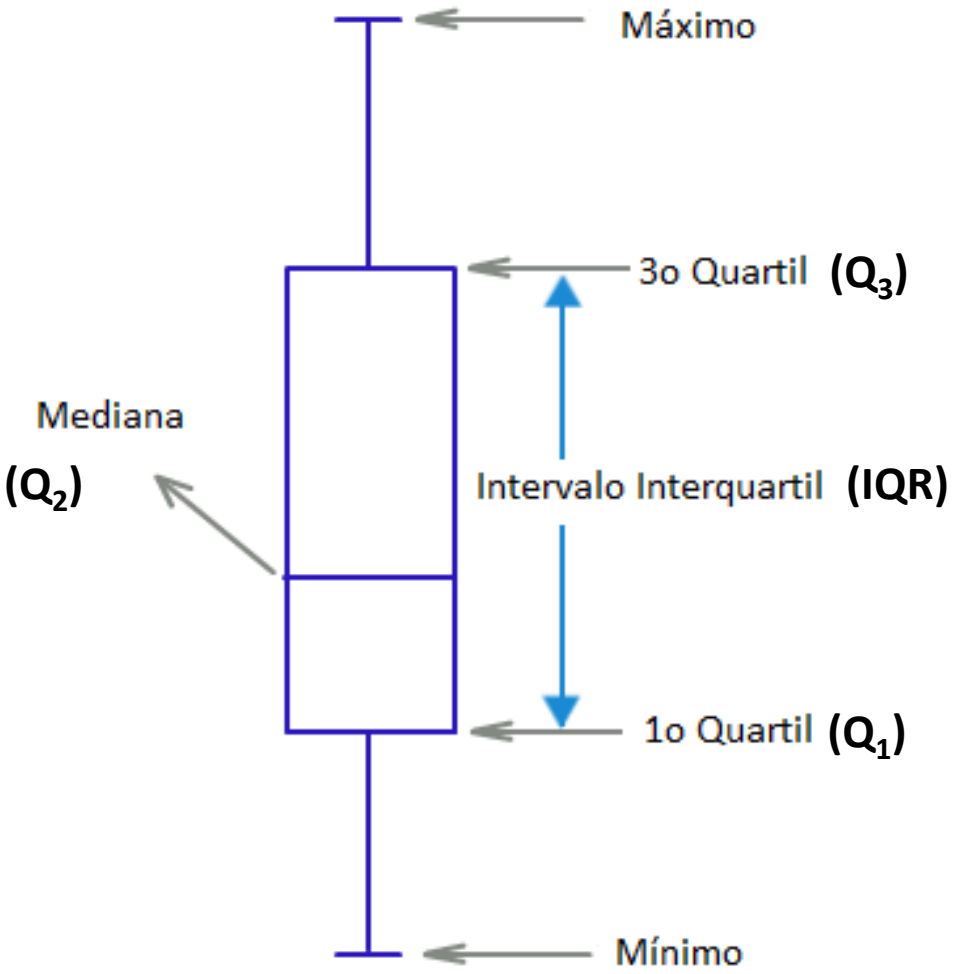
- Para **cada valor do atributo**, **subtrai a mediana (Q_2) e divide pelo** ***interquartil range*** (IQR ou FIQ, faixa interquartil).

- Os *quartis* de um atributo ordenado são os três pontos que dividem o domínio do atributo em quatro grupos de cardinalidade iguais.

- Cada qual composto por 1/4 da quantidade total de dados.

Preparação dos Dados

Transformação de Dados – **Normalização – range interquartil**



Preparação dos Dados

Transformação de Dados – **Normalização – range interquartil - Fórmula**

- Fórmula:

$$a' = \frac{(a - Q_2)}{IQR}$$

$$onde\ Q_i = i * \frac{n+1}{4}\ e\ IQR = Q_3 - Q_1.$$

Preparação dos Dados

Transformação de Dados – **Normalização – range interquartil - Exemplo**

Exemplo: Normalização Range interquartil para converter o atributo idade.

Id (a = 67)) =>

Range Interquartil(67)

$Q_1 = 46{,}5$, $Q_2 = 59$, $Q_3 = 66{,}5$

$$a' = \frac{(a - Q_2)}{IQR} = \frac{(67-59)}{66{,}5-46{,}5} = \frac{8}{20} = 0{,}4$$

Id	Idade	range interquartil
1	**67**	**0,40**
2	43	-0,80
3	58	-0,05
4	28	-1,55
5	74	0,75
6	65	0,30
7	70	0,55
8	42	-0,85
9	57	-0,10
10	60	0,05

Preparação dos Dados

Transformação de Dados – **Conversão Simbólico/Numérico**

- **Algumas Técnicas de AM** como **redes neurais** artificiais e ***support vector machines*** e vários algoritmos de agrupamento **lidam apenas com dados numéricos**.

Preparação dos Dados

Transformação de Dados – **Conversão Simbólico/Numérico**

- Quando o **atributo** é do tipo **nominal** e assume apenas **dois valores**:
  - se **os valores denotam** a **presença ou ausência** de uma **característica**, o valor **0 indica a ausência** e **1 a presença**.
  - **se apresentam numa relação de ordem**, o **menor valor ordinal assume** o valor **0** e **o outro** assume o valor **1**.

Preparação dos Dados

Transformação de Dados – **Conversão Simbólico/Numérico**

Codificação 1 – de - c

- No caso de um **atributo** simbólico **nominal com mais de dois valores sem uma relação de ordem**, pode-se codificar **cada valor nominal** por **uma sequência de c bits**, em que **c** é igual ao número de possíveis valores ou categorias.
- Na codificação 1-de-c, também chamada de canônica ou topológica, **cada sequência possui apenas um bit com o valor 1 e os demais com o valor zero**.
- A diferença entre as sequências é definida pela posição que o valor 1 ocupa nelas.

Preparação dos Dados

Transformação de Dados – **Conversão Simbólico/Numérico**

Codificação 1 – de - c

Atributo nominal	Código $1-de-c$
Azul	100000
Amarelo	010000
Verde	001000
Preto	000100
Marrom	000010
Branco	000001

Preparação dos Dados

Transformação de Dados – **Conversão Simbólico/Numérico**
Pseudoatributos

- **Dependendo da quantidade de valores nominais, a sequência binária** para representar um valor **pode ficar muito longa**.
- Por exemplo, codificar os nomes dos 193 países com a codificação 1-de-c necessita de um vetor com 193 elementos.
- Uma **alternativa** é a **representação dos possíveis valores nominais por um conjunto de pseudoatributos**, onde **(b) denota valor do tipo binário ou (i) inteiro** (por questões de simplificação).

Preparação dos Dados

Transformação de Dados – **Conversão Simbólico/Numérico**

Pseudoatributos para os 193 países. TMA = Temperatura Média Anual

Pseudoatributo	#Valores
Continente	7 (b)
PIB	1 (i)
População	1 (i)
TMA	1 (i)
Área	1 (i)

Uma **combinação de valores para os cinco pseudoatributos representa um único país**.

Preparação dos Dados

Transformação de Dados – **Conversão Simbólico/Numérico**
***One-Hot Encoding* (Similar a 1 de c)**

- *One-Hot Encoding* é a representação numérica de uma variável categórica de muitos níveis em que as colunas possuem 0s em todas as linhas, exceto onde o valor corresponde à nova coluna, que seria 1.
- Uma técnica que é amplamente utilizada em Inteligência Artificial para tarefas de processamento de linguagem natural (PLN) e Visão Computacional.
- Na realidade é uma técnica de manipulação de variáveis categóricas, na qual converte-se o valor da variável categórica para o correspondente numérico na forma de um vetor codificado de 0s e 1s.

Preparação dos Dados

Transformação de Dados – **Conversão Simbólico/Numérico**

***One-Hot Encoding* (Similar a 1 de c) – Exemplo**

- Considere a variável cor que tem três valores possíveis: vermelho, verde, azul. Como converter esta variável no seu correspondente numérico com One-Hot?

Cor
Vermelho
Verde
Azul
Verde
Vermelho
Vermelho
Azul

Cor.vermelho	Cor.verde	Cor.azul
1	0	0
0	1	0
0	0	1
0	1	0
1	0	0
1	0	0
0	0	1

- Com isso, a partir de uma coluna (atributo) foram criadas três novas colunas (atributos)

Preparação dos Dados

Transformação de Dados – **Conversão Simbólico/Numérico**
Valor ordinal para inteiro

- **Quando existe uma relação de ordem**, o atributo é do tipo ordinal, e a **codificação deve preservar essa relação**.
- Quando o valor numérico é um número inteiro ou real, essa transformação é simples e direta, basta ordenar os valores categóricos e codificar cada valor de acordo com sua posição na ordem.
- **A distância entre os valores varia de acordo com a proximidade deles**.

Preparação dos Dados

Transformação de Dados – **Conversão Simbólico/Numérico**

Valor ordinal para inteiro

Valor ordinal	Valor inteiro
Primeiro	0
Segundo	1
Terceiro	2
Quarto	3
Quinto	4
Sexto	5

A distância entre os valores varia de acordo com a proximidade deles.

Preparação dos Dados

Transformação de Dados – **Conversão Simbólico/Numérico**
Valor ordinal para Binário

- Se for necessário **converter valores ordinais em** valores **binários**, pode-se **utilizar** o **código cinza ou** o **código termômetro**.

- Em **ambos os casos os próximos valores diferem por apenas um bit** (chamada de **distância de Hamming igual a 1**).

- O **código termômetro utiliza sequências binárias maiores** (mais bits) **que** o código **cinza**.

Preparação dos Dados

Transformação de Dados – **Conversão Simbólico/Numérico**

Valor ordinal para Binário

Valor ordinal	Código cinza	Código termômetro
Primeiro	000	00000
Segundo	001	00001
Terceiro	011	00011
Quarto	010	00111
Quinto	110	01111
Sexto	100	11111

Preparação dos Dados

Transformação de Dados – **Conversão Numérico/Simbólico**

- **Algumas técnicas de AM** foram desenvolvidas para **trabalhar com valores qualitativos**, como uma parcela dos algoritmos de classificação e associação.

- **Alguns** dos algoritmos **podem lidar com dados quantitativos**, **mas tem seu desempenho reduzido**.

- Se o **atributo quantitativo** for do tipo **discreto e binário**, com apenas dois valores, a conversão é trivial. **Basta associar um nome a cada valor**.

Preparação dos Dados

Transformação de Dados – **Conversão Numérico/Simbólico**

- Se o **atributo original** for formado por **sequências binárias sem uma relação de ordem** entre si, **cada sequência** pode ser **substituída por** um **nome ou categoria**.

- Nos **demais casos**, **métodos de discretização permitem transformar atributos quantitativos em qualitativos, transformando valores numéricos em intervalos ou categorias**.

- **Cada intervalo é convertido em um valor qualitativo**.

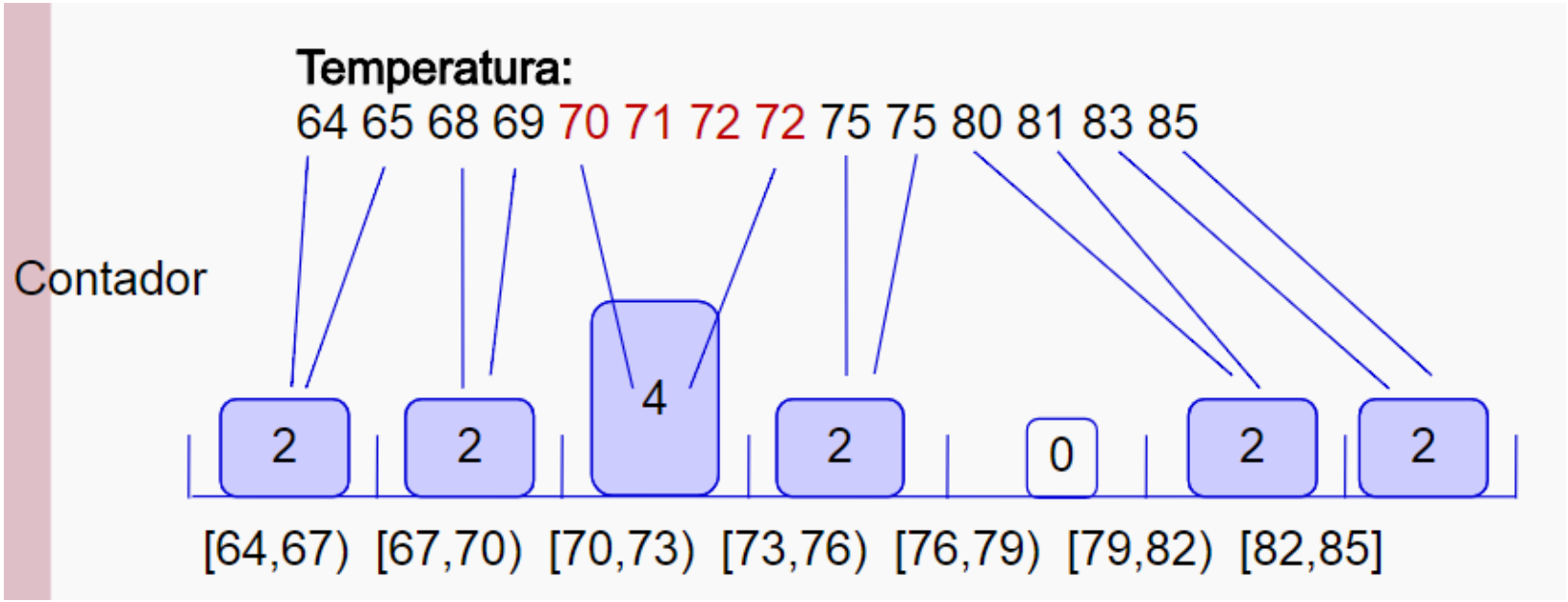
Preparação dos Dados

Transformação de Dados – **Conversão Numérico/Simbólico**

- Algumas **estratégias** utilizadas **para mapear os intervalos** pelos diferentes métodos **são**:
  - **Larguras iguais**: **divide o intervalo original de valores em subintervalos com mesma largura**. O **desempenho** dessa estratégia **pode ser afetado** pela **presença de *outliers***.
  - **Frequências iguais**: **atribui o mesmo número de objetos a cada subintervalo**. Essa estratégia **pode gerar intervalos de tamanhos muito diferentes**.
  - Uso de **algoritmos de agrupamento de dados**.
  - **Inspeção visual**.

Preparação dos Dados

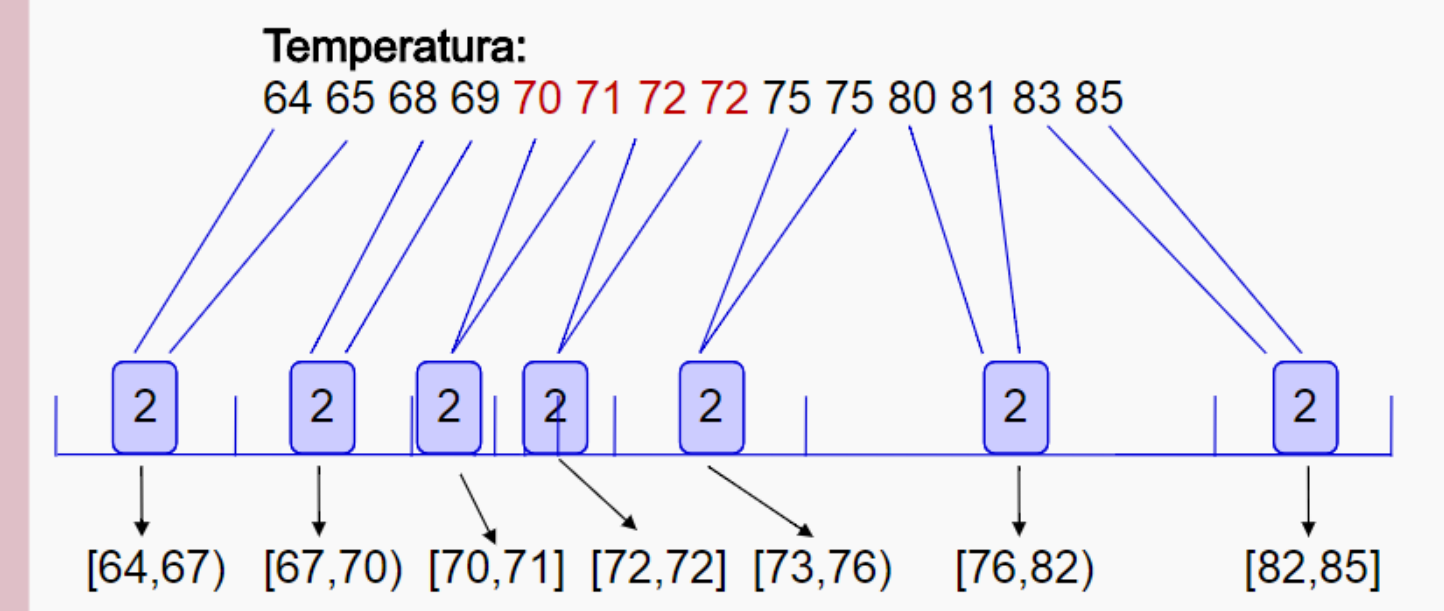
Transformação de Dados – **Conversão Numérico/Simbólico - Discretização – Larguras Iguais**



Observe que cada subintervalo tem largura 3

Preparação dos Dados

Transformação de Dados – **Conversão Numérico/Simbólico - Discretização – Frequências Iguais**



Cada intervalo tem 2 valores.

Preparação dos Dados

Transformação de Dados – **Atributos Numéricos**

- Algumas vezes, o valor numérico de um atributo precisa ser transformado em outro valor numérico.
- Isso geralmente **ocorre quando os limites inferior e superior** de valores **dos atributos são muito diferentes** ou **quando vários atributos estão em escalas diferentes**.
- É realizada para **evitar que um atributo predomine sobre outro**.
- Nesses casos, a normalização (Max-min (reescalar) e a score-z (padronização)) pode ser aplicada.

Preparação dos Dados

Transformação de Dados – **Atributos Numéricos**

- **Geralmente, é preferível padronizar (score-z) a fazer uso da reescala (max-min), pois a padronização lida melhor com outliers.**
- No entanto, **pode haver situações em que essa variação deve ser preservada** por ser importante para a indução de um bom modelo.
- Temos também a **tradução** , no qual o valor de um dado tipo é traduzido para um valor do mesmo tipo, mais facilmente manipulável.
- Por exemplo: a conversão de um atributo data de nascimento para idade, graus Celsius para Fahrenheit, ou de localização dada por um aparelho de GPS para o código postal.

Bibliografia

BÁSICA:

- AGGARWAL, Charu C. **Artificial Intelligence: A Textbook**. New York: Springer: 2021.
- CHOLLET, François. **Deep Learning with Python, 2ed.** Shelter Island: Manning, 2021.
- GÉRON, Aurélien. **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**, 2 ed**.** Sebastopol: O'Reilly, 2019.

COMPLEMENTAR:

- GOODFELLOW, Ian; BENGIO, Yoshua, COURVILLE, Aaron. **Deep Learning**. Cambridge: MIT Press, 2016.
- RASCHKA, Sebastian; MIRJALILI, Vahid. **Python Machine Learning**. 3 ed. Birmingham: Packt, 2017.
- RUSSEL, Stuart; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. 3 ed. Upper Saddle River: Pearson, 2010.
- TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introduction to Data Mining**. 2 ed. Upper Saddle River: Pearson, 2018.
- VANDERPLAS, Jake. **Python Data Science Handbook**. Sebastopol: O'Reilly, 2017.

ADICIONAIS:

- FACELI, Katti et al. **Inteligência artificial: uma abordagem de aprendizado de máquina**. 2ª Ed. Rio de Janeiro: LTC- Livros Técnicos e Científicos, 2021.
- LUGER, George F. **Inteligência Artificial** - 6ª ed. São Paulo: Pearson Education do Brasil, 2015.
- RUSSELL, Stuart; NORVIG, Peter. **Inteligência artificial: Uma Abordagem Moderna** - 4ª. Ed. GEN LTC, 2022.