



Normalização e Discretização

Agente Educacional
Marcelo Pita

Agenda

Normalização
Discretização



Normalização

Altera a escala dos dados, normalmente para **padronizá-la**.

Tipos mais comuns:

- min-max**

- Z-score**

- Escala decimal**

min-max

define novo intervalo [nmin, nmax]

$$v' = \frac{v - \min}{\max - \min} (n\max - n\min) + n\min$$

Exemplo

se valores variam de 12000 a 98000 e queremos normalizá-los entre [0,1], então 73600 vira:

$$\{ (73600 - 12000) / (98000 - 12000) \} * (1 - 0) + 0 = 0,716$$

Z-score

distribuição normal

$$v' = \frac{v - \bar{x}}{s}$$

Escala decimal

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

j : menor inteiro de forma que $\max(|v'|) < 1$

Discretização

Conversão de **valores numéricos** em **intervalos** (categóricos).

Alguns algoritmos clássicos de ciência de dados só trabalham com **dados categóricos/nominais**

Exemplo: descoberta de regras de associação.

Os atributos numéricos contínuos são **quantizados**, i.e., alocados para intervalos, que se tornam o **novo valor do registro**.

Exemplo

O atributo idade pode ser discretizado nos seguintes intervalos fixos:

$(0, 20]$, $(20, 30]$, $(30, 40]$, $(40, 50]$, $(50, 65]$, $(65, 80]$, $(80, \infty]$

Métodos comuns de discretização

Por intervalo

- Largura fixa

- Frequência fixa

Por entropia

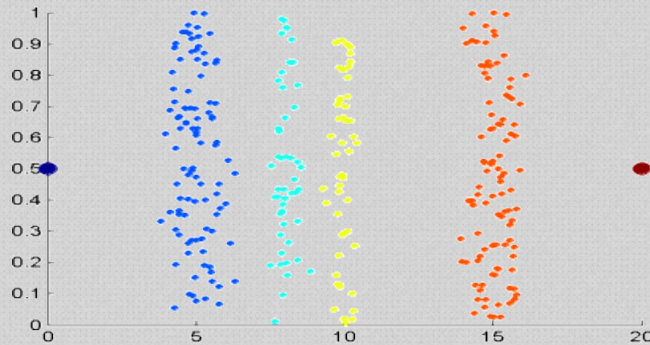
- Maximiza o ganho de informação nos intervalos

Usando técnica de agrupamento

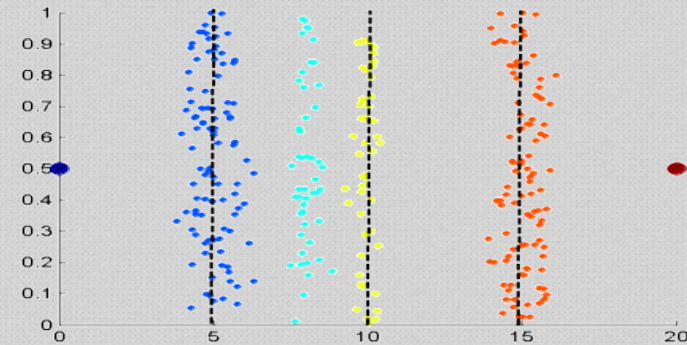
- Usando k-means, por exemplo

Normalização e Discretização

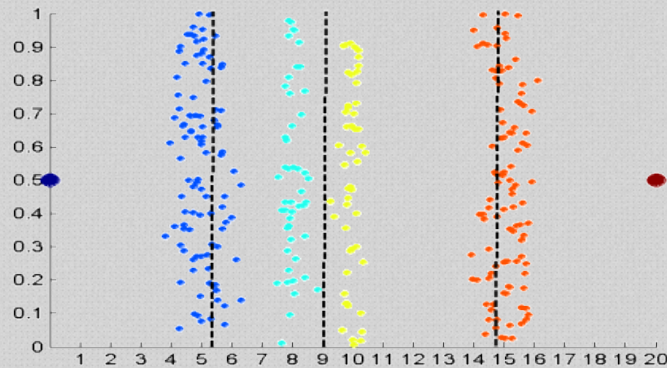
Discretização
Métodos comuns de discretização



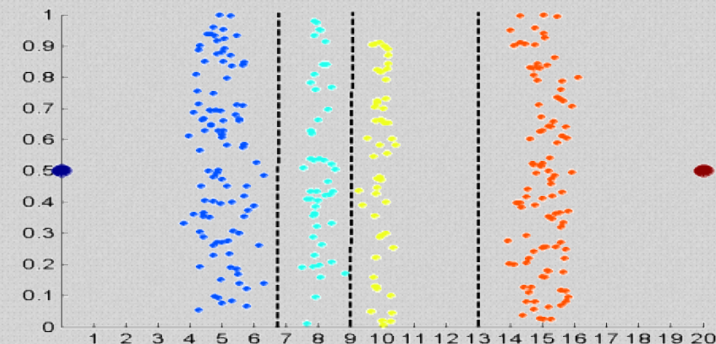
Data



Intervalos iguais



Intervalos com mesma frequência



Agrupamento com K-means

[Exercício Normalização e Discretização – KNIME]





Obrigado!

Agente Educacional

Marcelo Pita

marcelo.pita@serpro.gov.br | #81 8794

Demais agentes educacionais sobre o assunto:

Sérgio M. Dias | sergio.dias@serpro.gov.br | #31 6539

Gustavo Torres | gustavo.gamatorres@serpro.gov.br | #31 6950