

Reconhecimento de Padrões

Feature Selection

Profa: Deborah Magalhães



“

*Feature engineering é o processo de transformar dados em features que melhor representam o problema tratado, resultando na melhoria de **desempenho** do algoritmo de aprendizado de máquina.*

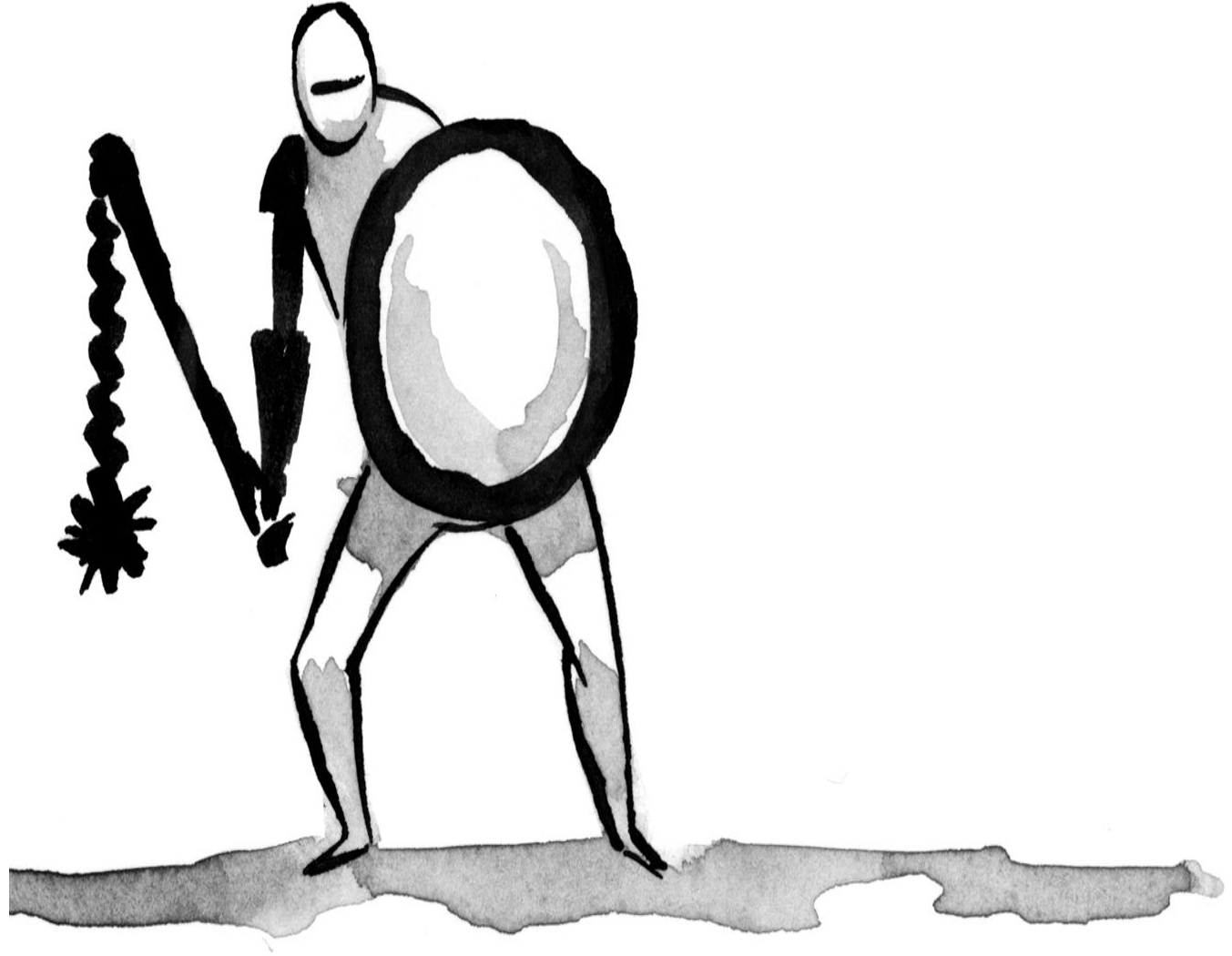
“

Atributo é geralmente o termo dado a uma coluna de uma tabela de dados, enquanto **característica** (*feature*) se refere apenas ao atributo que contribui para o sucesso do algoritmos de aprendizado de máquina.

Say no to bad attributes

✓ Performance

✓ Time



Baseados em estatística

- Correlação de Pearson
- Testes de Hipótese

Baseados em modelos

- Recursive Feature Elimination (RFE)
- Information Gain (IG)

**Pearson
product-moment
correlation
coefficient
(PPMCC) OU
Coeficiente de
Pearson**

$$r_{xy} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{\sum x_i^2 - n \bar{x}^2} \sqrt{\sum y_i^2 - n \bar{y}^2}},$$

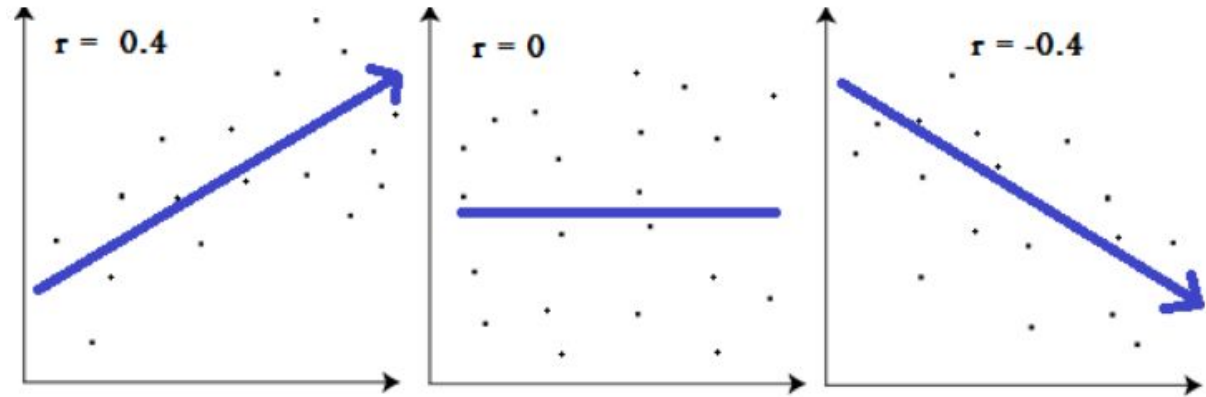
onde :

n : tamanho da amostra

x_i, y_i : valor da observacao indexado por i

\bar{x}, \bar{y} : media amostral

Coeficiente de Pearson



Força da Associação	Coeficiente (r)	
	Positiva	Negativa
Pequena	.1 a .3	-.1 a -.3
Média	.3 a .5	-.3 a -.5
Grande	.5 a 1.0	-.5 a -1.0

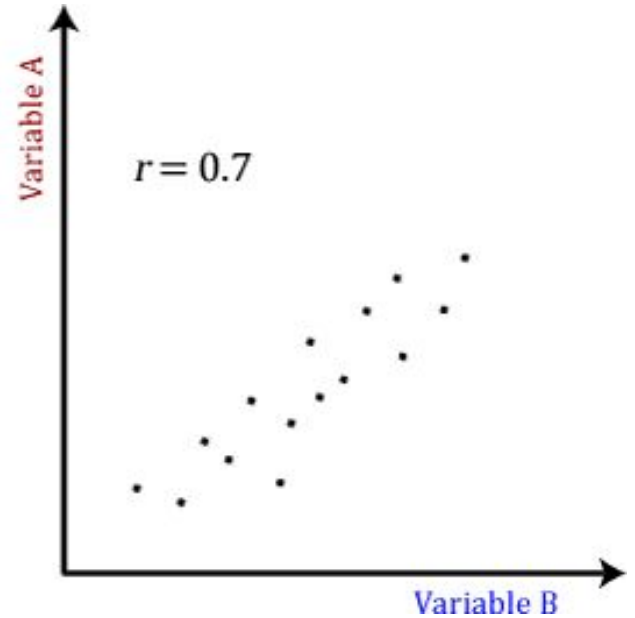
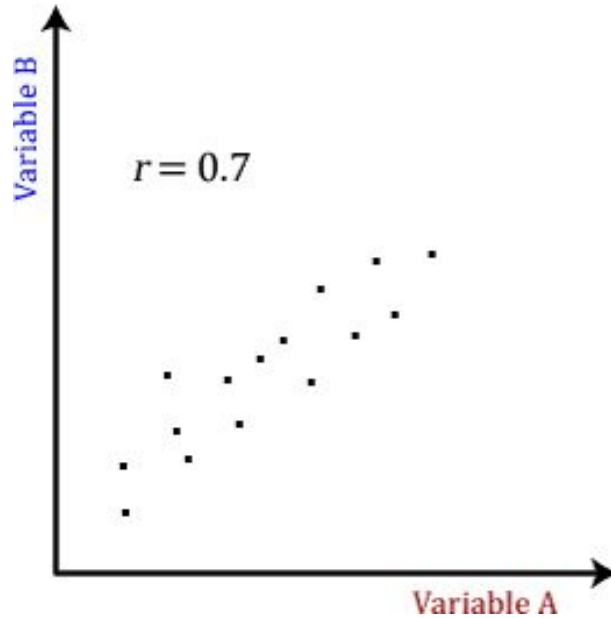
Desvantagens da Correlação de Pearson

#1 - Não se pode utilizar qualquer tipo de variável

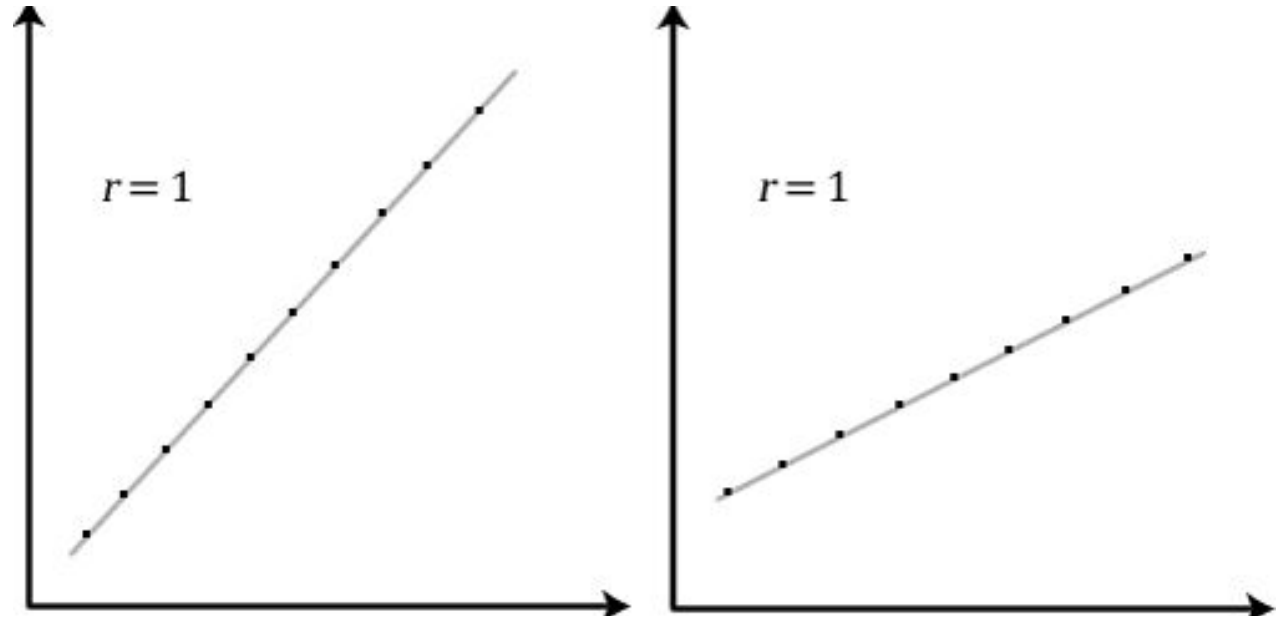
Quantitativa vs. Categoria

Quantitativa	O que os dados representam?	Exemplos
Discreta	Número contável entre quaisquer dois valores	Número de reclamações de clientes, número de falhas de uma peça
Contínua	Número infinito entre quaisquer dois valores	Comprimento, volume, saldo

#2 - Não há diferença entre variáveis dependentes e independentes



**#3 - Não
representa a
inclinação da
linha de
melhor ajuste**



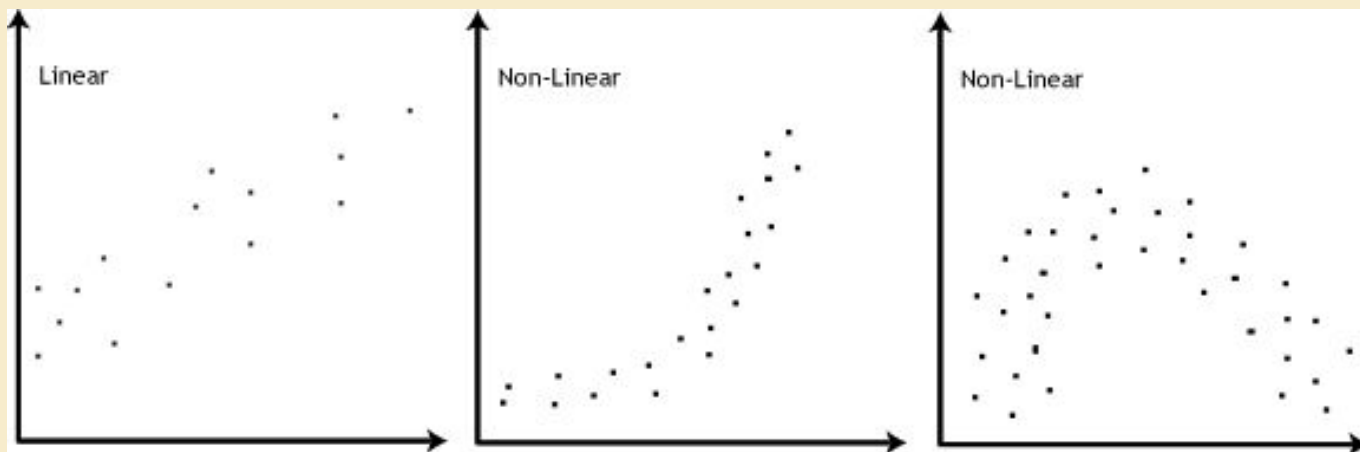
Premissas da correlação de Pearson

#1: as duas variáveis devem ser contínuas

#2: independência das observações

#3: as variáveis devem seguir a distribuição normal univariada

#4: as variáveis devem possuir uma relação linear





Muito Obrigada!

Se você tiver qualquer dúvida ou sugestão:

- deborah.vm@ufpi.edu.br

