Machine Learning com Python

Prof. Luciano Galdino

Etapas Fundamentais para construção do algoritmo

1) Análise do problema.

2) Exploração, Tratamento e Análise dos dados.

3) Pré processamento dos dados.

4) Escolha do grupo de algoritmos que podem ser utilizados.

5) Criação dos algoritmos de Machine Learning.

6) Comparação e escolha do melhor algoritmo.

Repositórios de dados

INEP: https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados

Google dataset Search: https://datasetsearch.research.google.com/

Portal brasileiro de dados abertos: www.dados.gov.br

Kaggle (competições Machine Learning): www.kaggle.com

UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/index.php

OMS: https://www.who.int/

Paho (organização panamericana de saúde): https://www.paho.org/en

DrivenData (competições Ciência de Dados): https://www.drivendata.org/

Separação de Dados de Treino e Teste

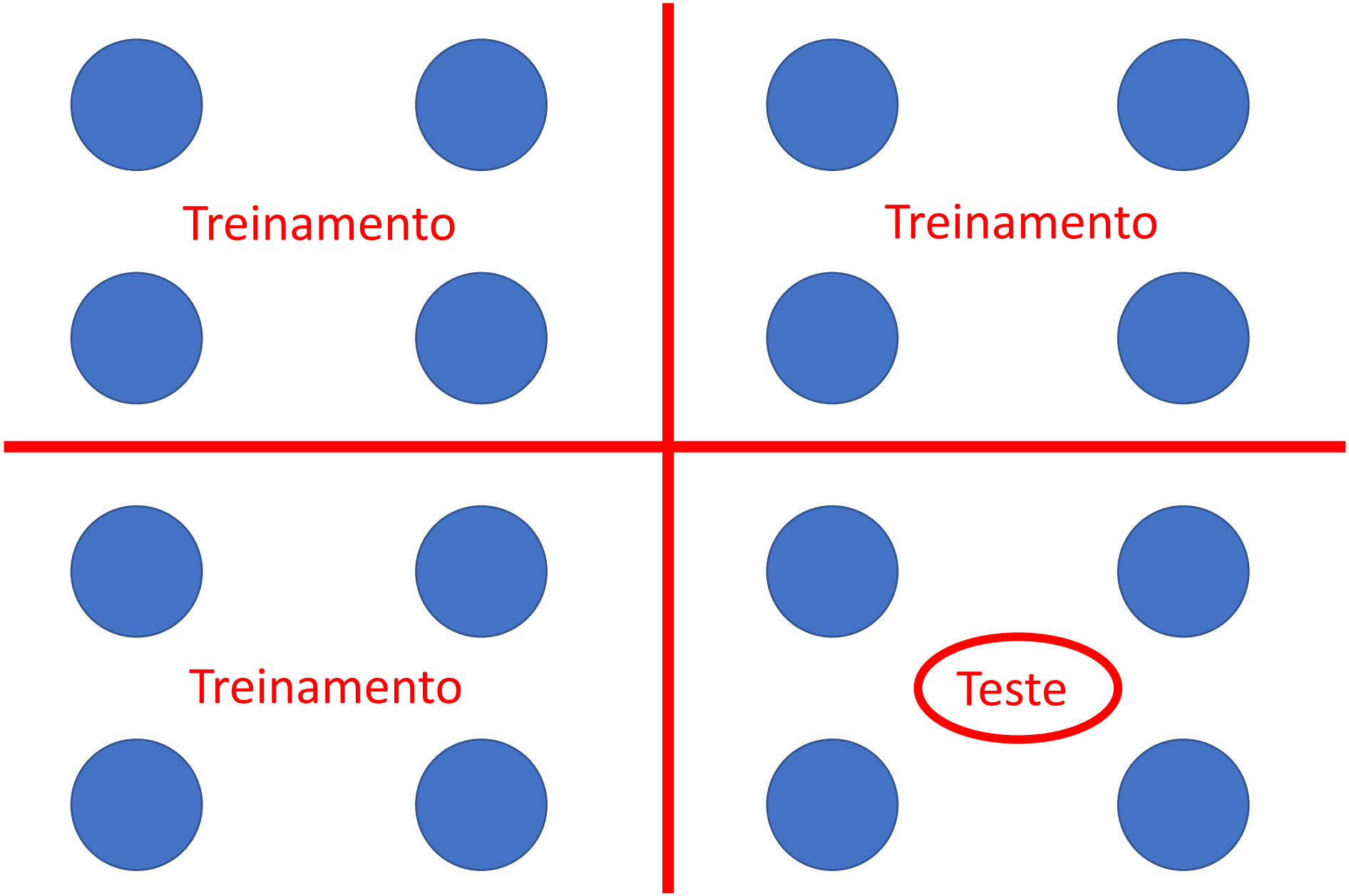
Dados de treino: Certa quantidade dos dados (aproximadamente 70%) destinada para treinar o algoritmo.

Dados de teste: Quantidade restante dos dados (aproximadamente 30%) para analisar o desempenho do algoritmo.

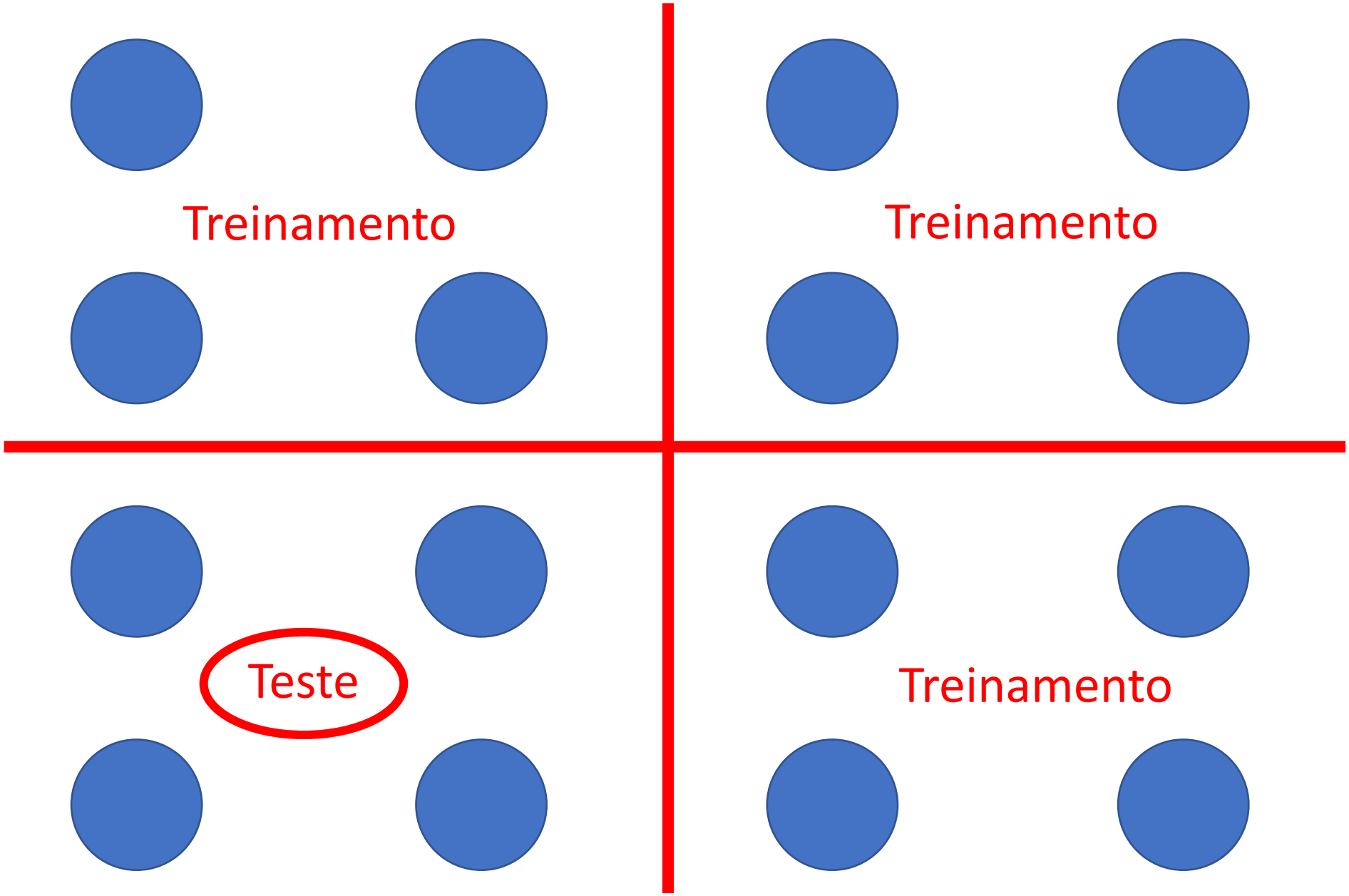
Essa separação deve ocorrer de maneira aleatória para evitar problemas nos modelos criados (Exemplo: ter uma quantidade de dados que aparecem em pequena quantidade ou nem aparecem nos dados de teste).

Validação cruzada (Cross validation)

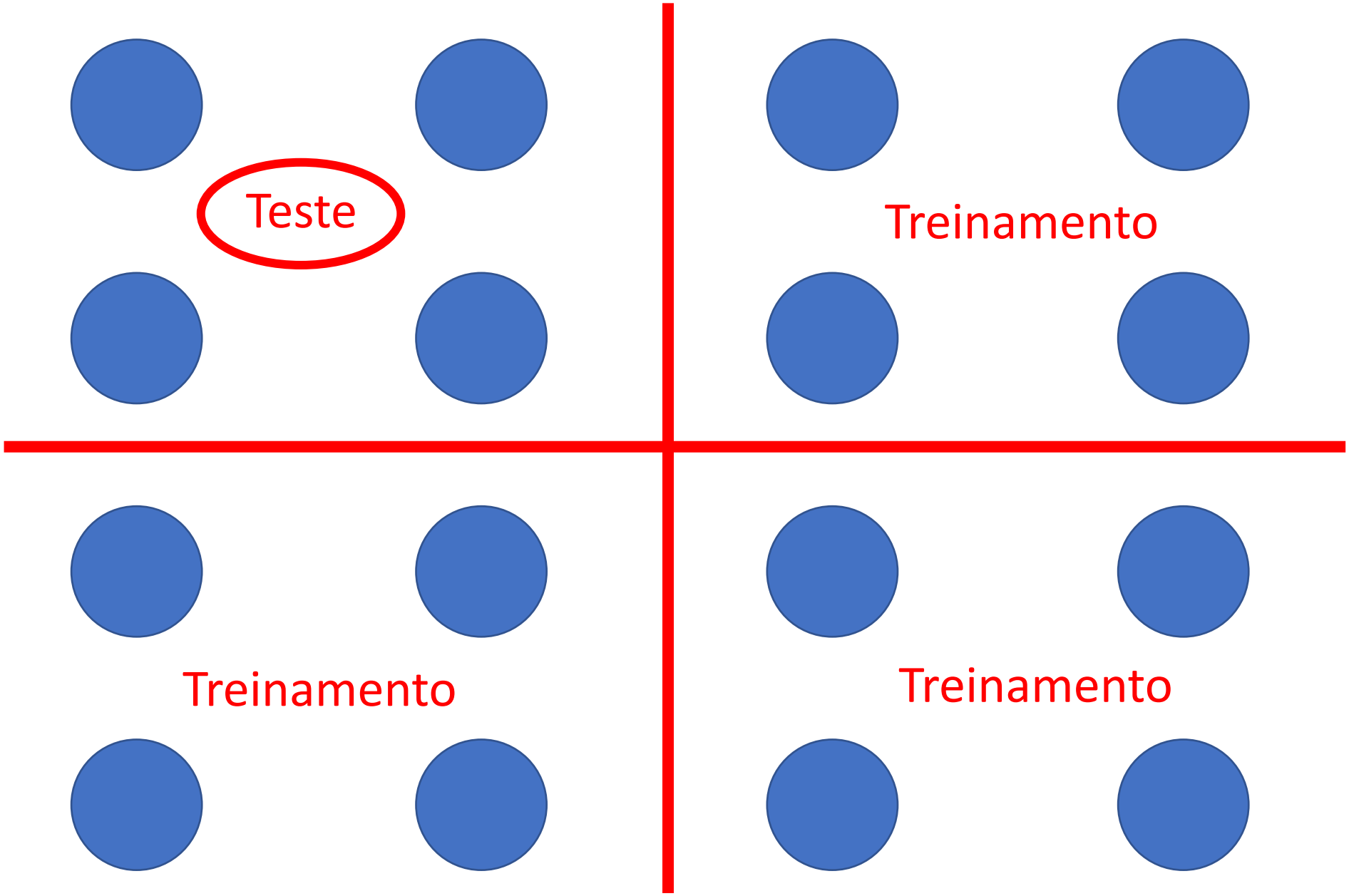
n_split = 4
(k-folds)



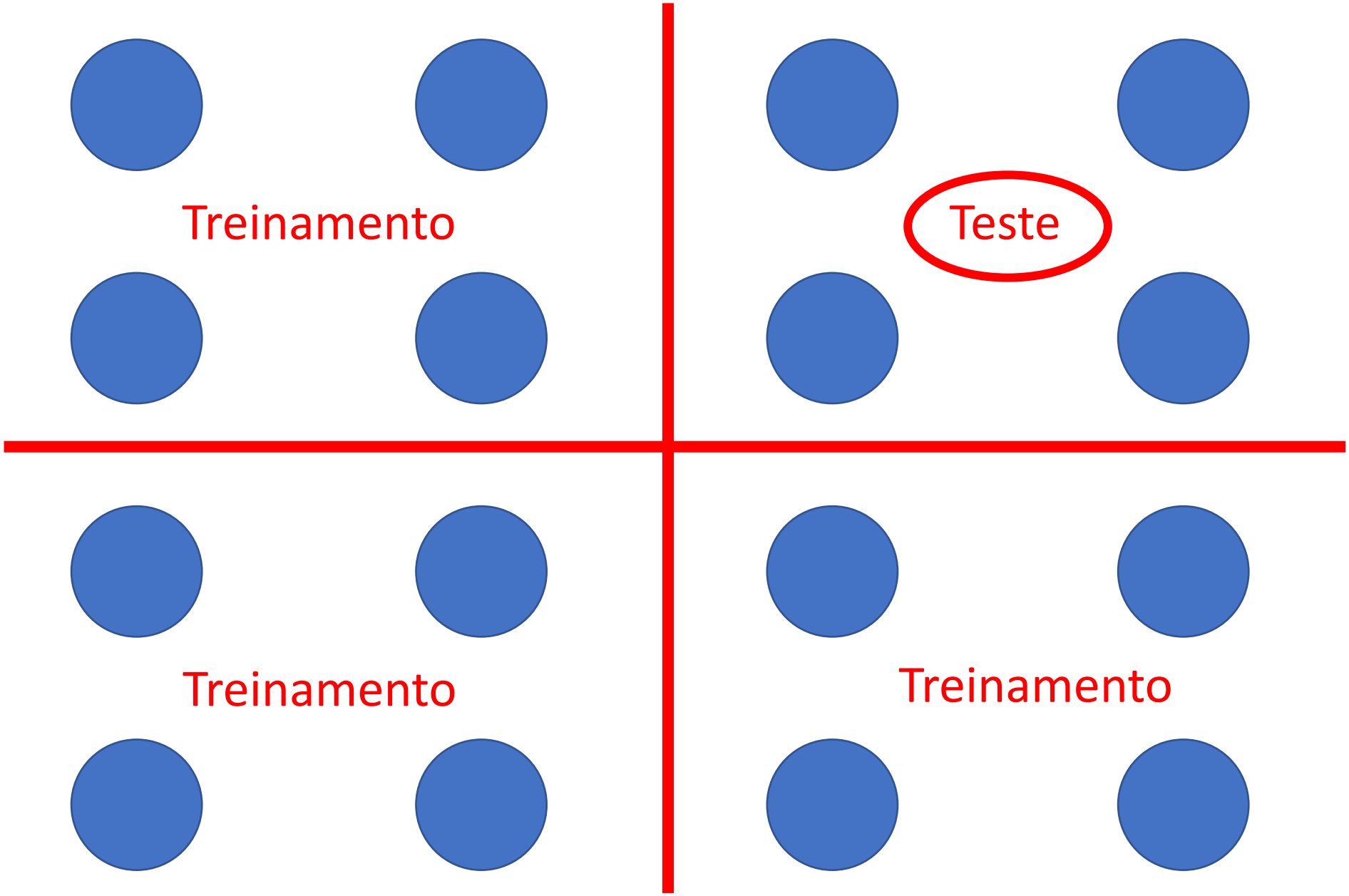
Validação cruzada



Validação cruzada



Validação cruzada



Atenção a dois problemas no treinamento		
1	Underfitting (alto viés)	Algoritmo que não se encaixa com os dados de entrada.
2	Overfitting (alta variância)	Algoritmo ótimo para os dados de entrada e ruim para dados de teste.