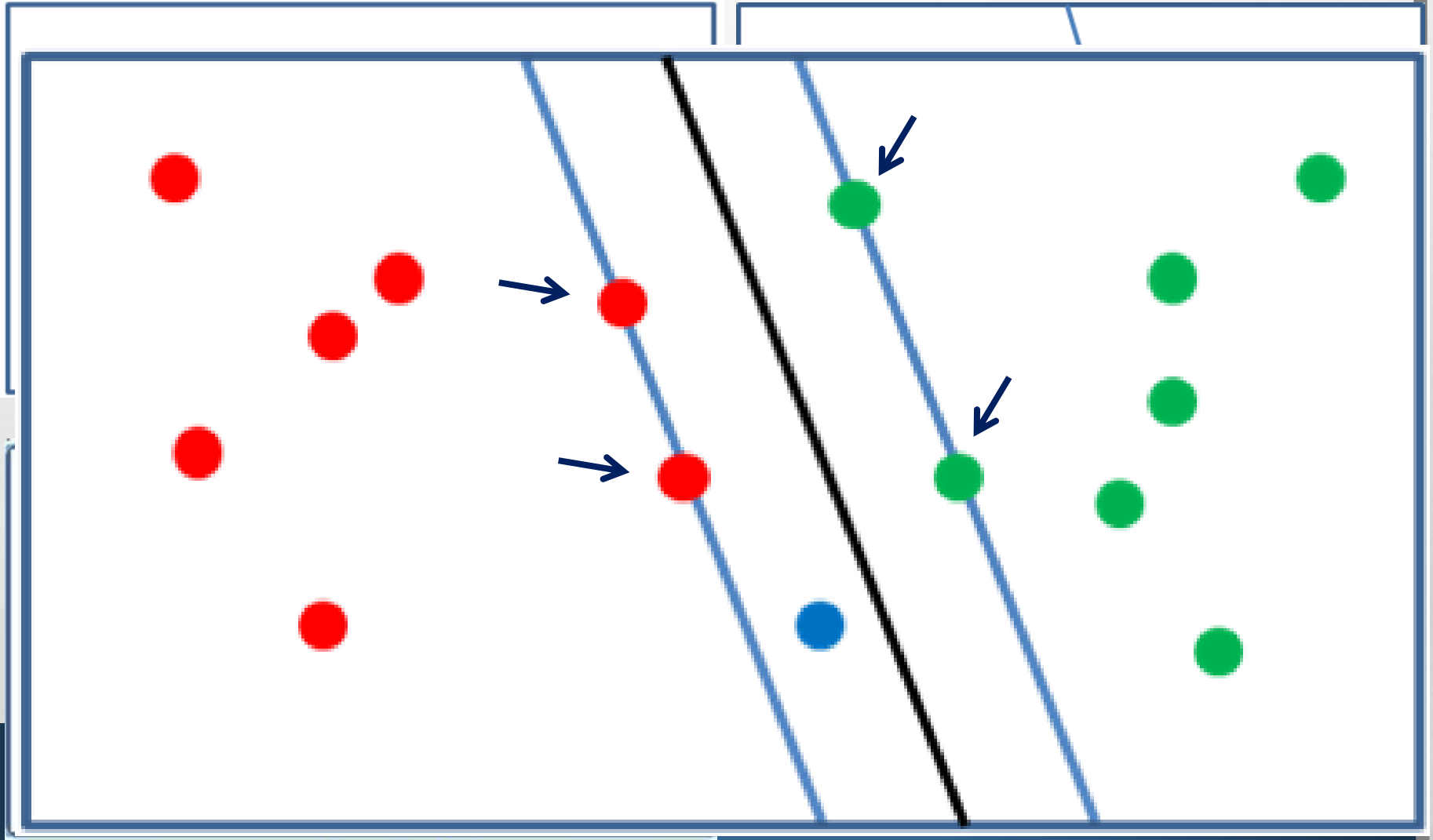


# FORMAÇÃO CIENTISTA DE DADOS

CLASSIFICAÇÃO



# MAQUINA DE VETOR DE SUPORTE

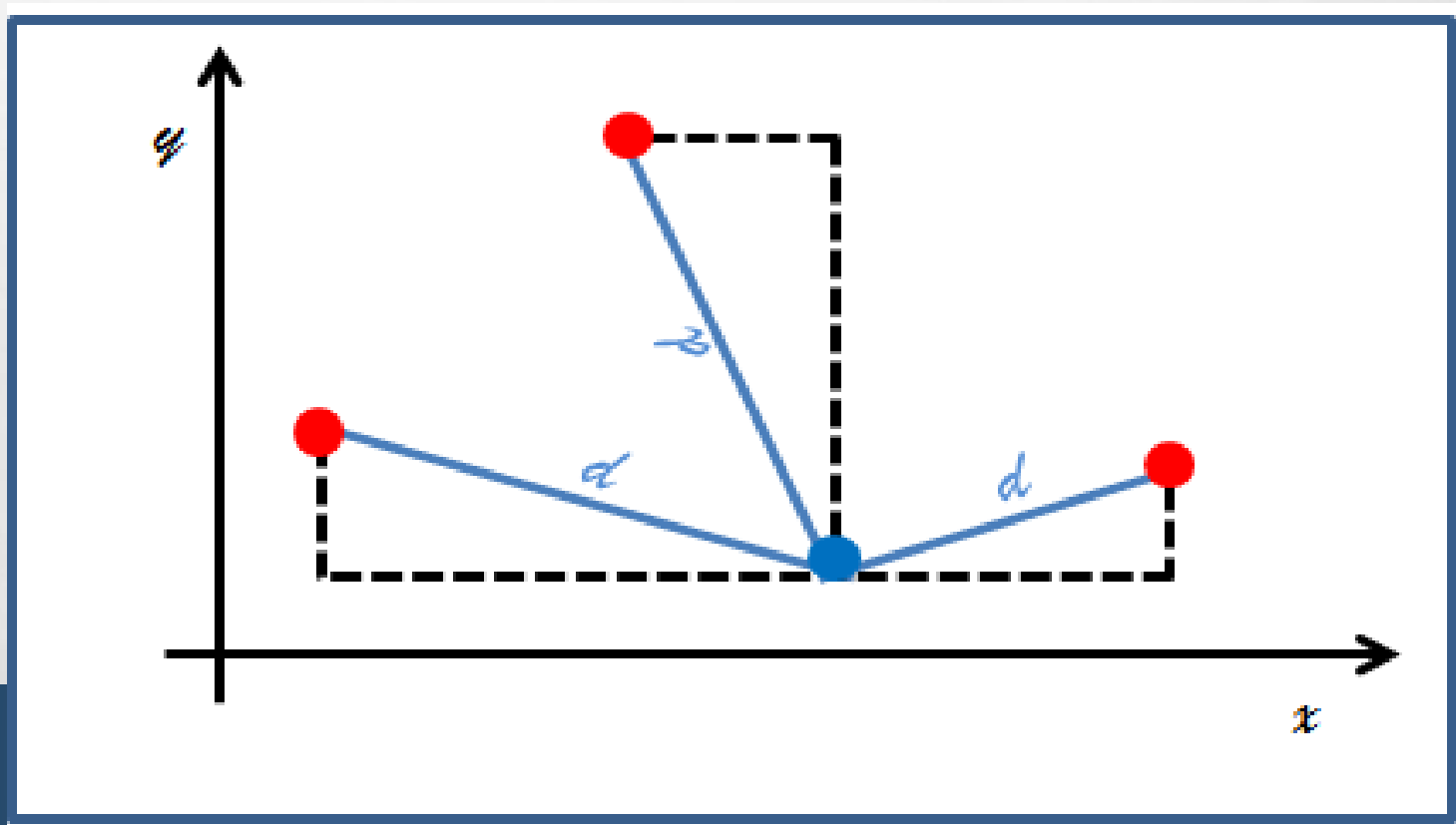


# MÉTODOS DE GRUPOS

- **FLORESTAS ALEATÓRIAS**
- **BOOSTING**

# APRENDIZADO BASEADO EM INSTÂNCIA

**Classificador do vizinho mais próximo (Nearest-neighbor)**



# SELEÇÃO DE ATRIBUTOS

- **QUANTOS ATRIBUTOS UM BOM MODELO DEVE TER?**
- **QUANTO MAIS MELHOR?**

**Fraude: Modelo com 10  
atributos: Taxa de Acerto 75%**

**Fraude: Modelo com 50  
atributos: Taxa de Acerto 68% !!!**

# MALDIÇÃO DA DIMENSIONALIDADE

- **A INCLUSÃO DE MUITOS (OU MAIS) ATRIBUTOS EM UM MODELO, DEGRADAM SUA PERFORMANCE!**
- **SE EXISTEM MUITOS ATRIBUTOS, COMO SABER QUAIS SÃO MAIS RELEVANTES PARA O MODELO?**
- **SELEÇÃO DE ATRIBUTOS:**
  - **PODE-SE FAZER “MANUALMENTE”**
  - **ALGUNS ALGORITMOS JÁ A FAZEM AUTOMÁTICAMENTE**