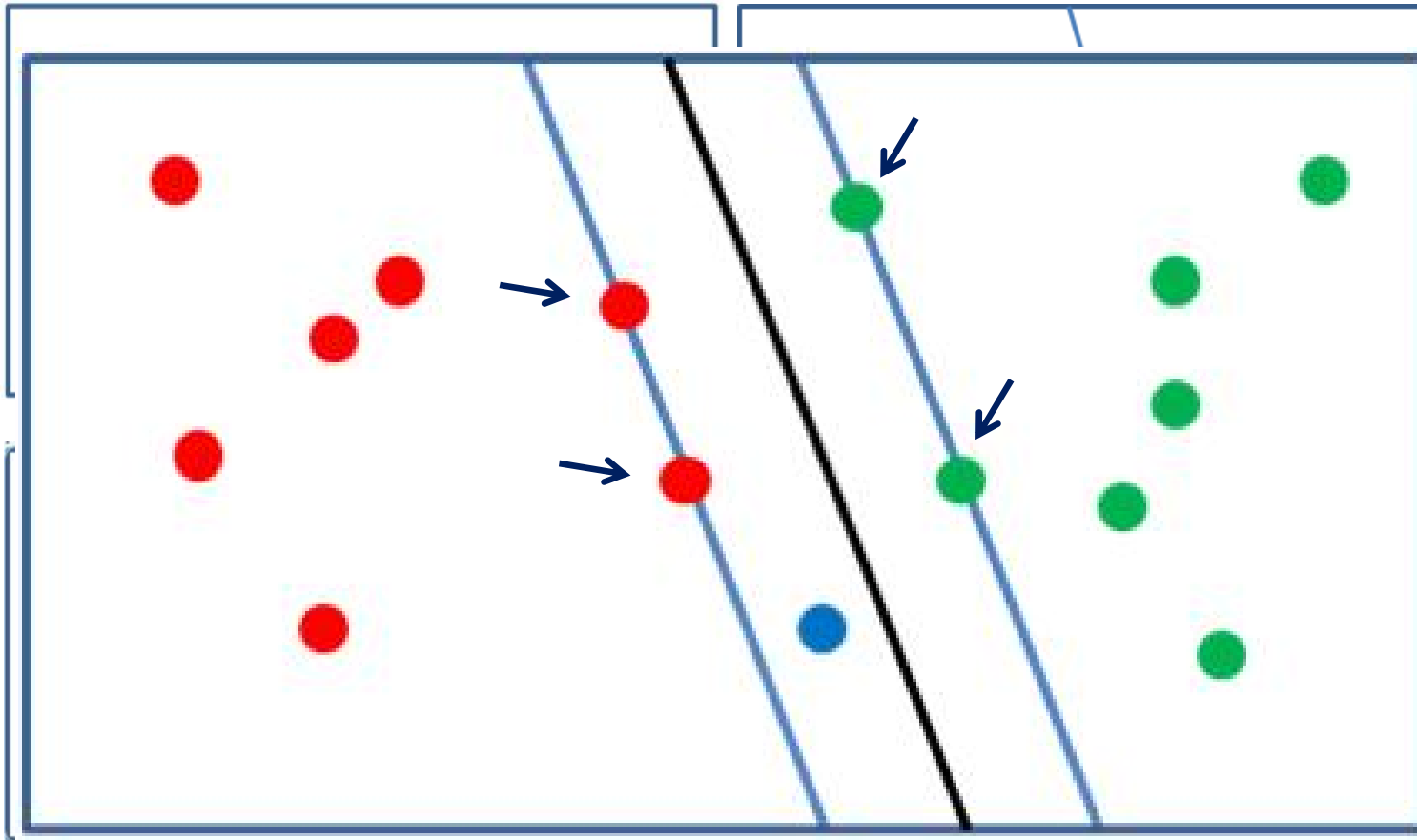




Classificação

Formação Cientista de Dados

Maquina de Vetor de Suporte



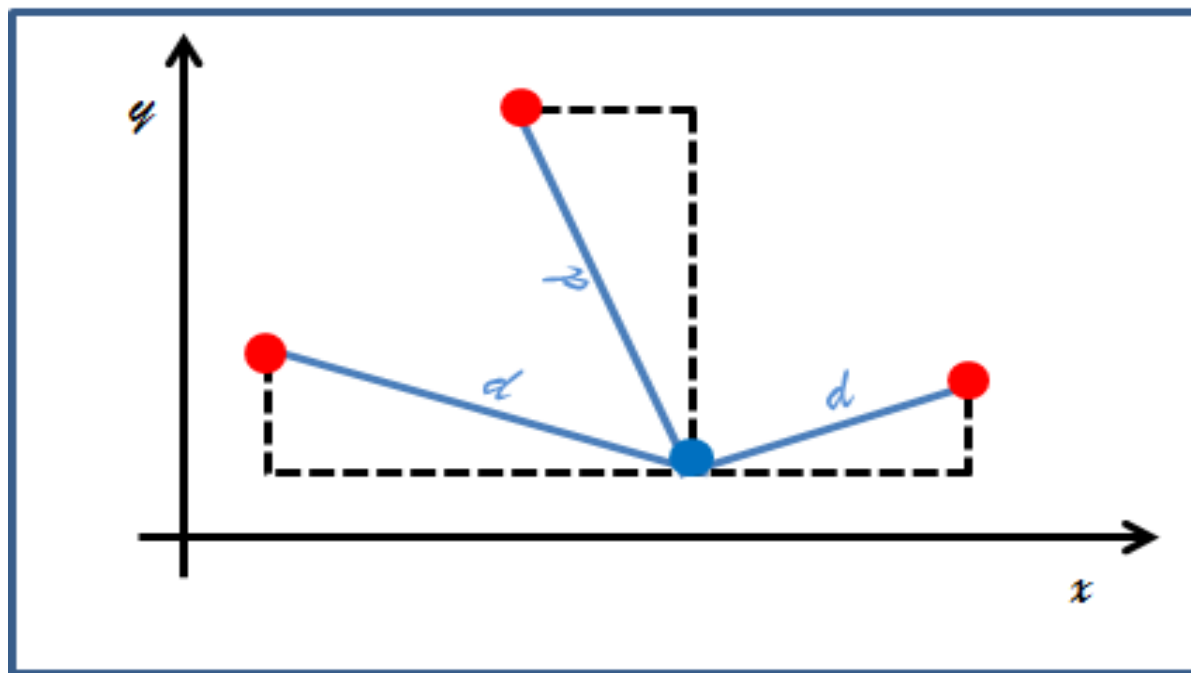


Métodos de Grupos

- 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 automaticamente
-

Problemas de Atributos Desconhecidos

- No treino: regiões “Sul”, “Sudeste”, “Centro-Oeste” e “Norte”
- Na produção: região “Nordeste”
- Como classificar?

