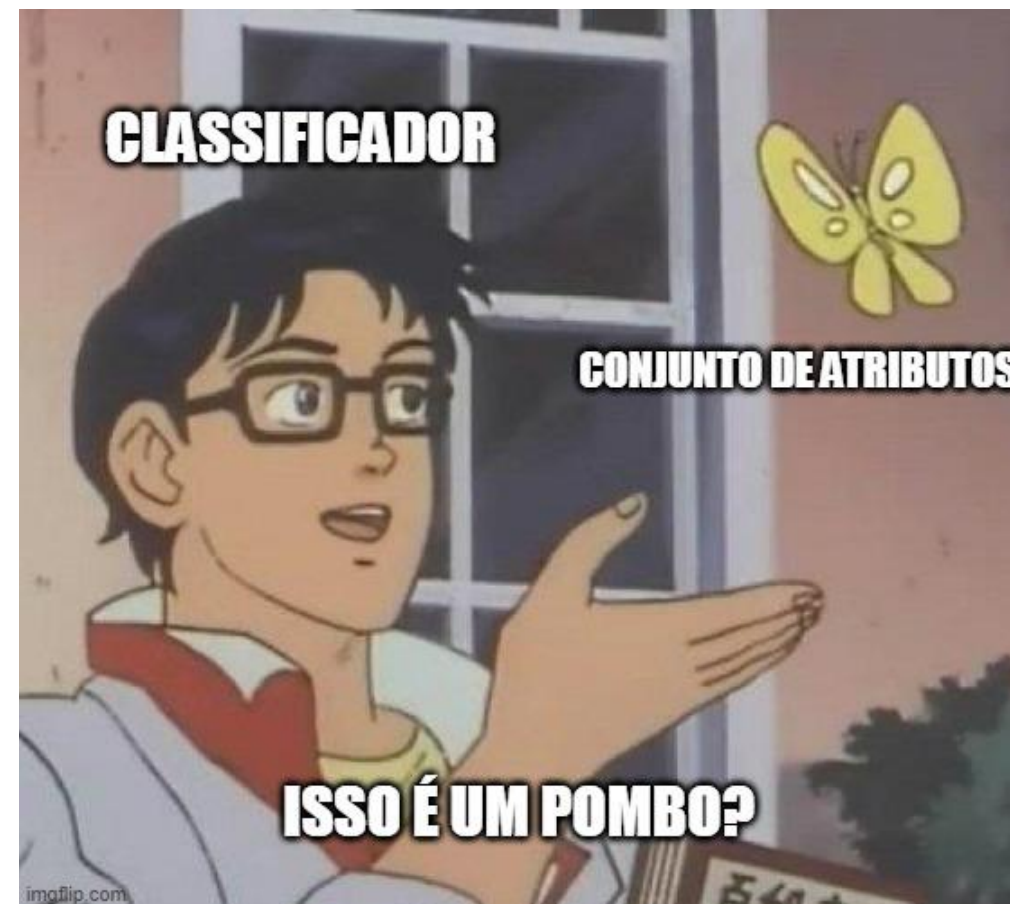


CLASSIFICADORES ELEMENTARES

Prof. André Backes | @progdescomplicada

Classificação

- Usa de exemplos conhecidos (já classificados) para aprender como categorizar as novas amostras com base nos seus atributos



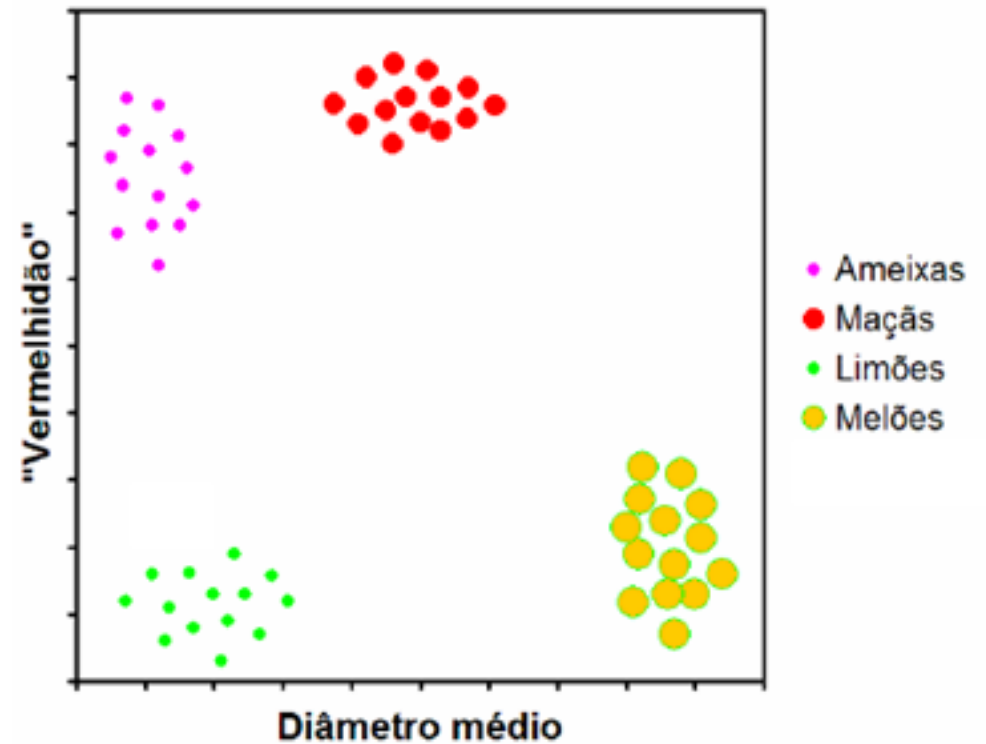
Classificação

- Consiste em tentar discriminar em diferentes classes um conjunto de objetos com características mensuráveis
 - Exemplo: classificação de frutas
 - Forma, cor, sabor, etc



Classificação

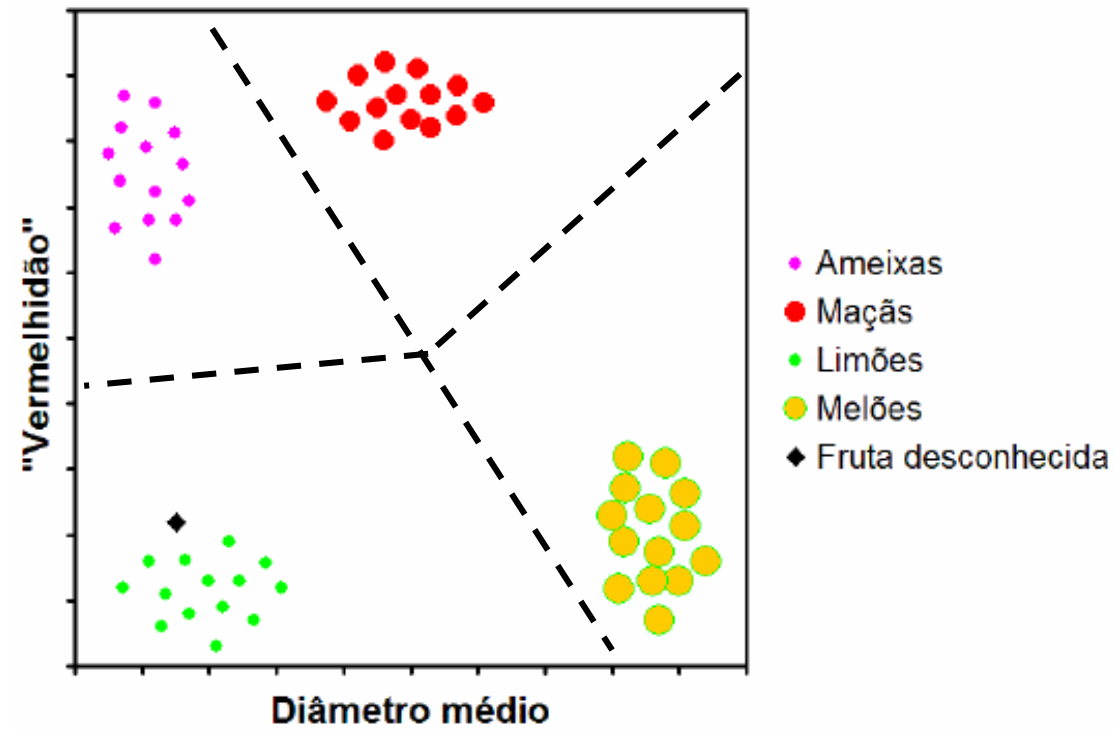
- Essas características, ou atributos, do objeto formam um espaço multidimensional
 - espaço de características
- Cada objeto é então representado como sendo um ponto nesse espaço



Classificador

- Trata-se de uma função discriminante, que, baseada na informação do conjunto de treinamento (amostras conhecidas) determina a semelhança da nova amostra a uma das classes
 - Nesse espaço de característica é que atua o classificador
 - Um objeto desconhecido é atribuído a uma classe a partir do seu posicionamento no espaço de característica

Classificador

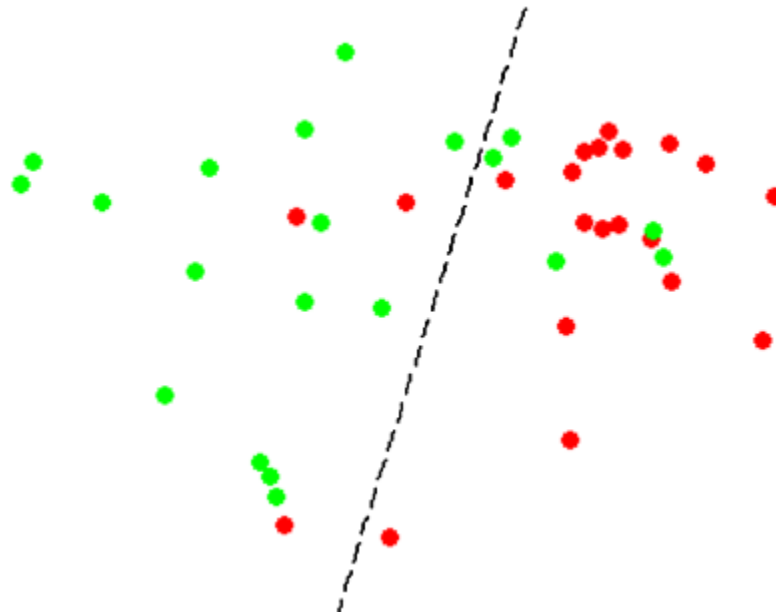


Classificadores paramétricos

- Parte da premissa de que as classes dos objetos no espaço de classificação seguem uma distribuição conhecida
 - Muitas vezes se considera a distribuição normal
- Esta hipótese conduz a superfícies de decisão com formas simples
 - Linear ou quadrática
 - Essa classificação depende da função discriminante

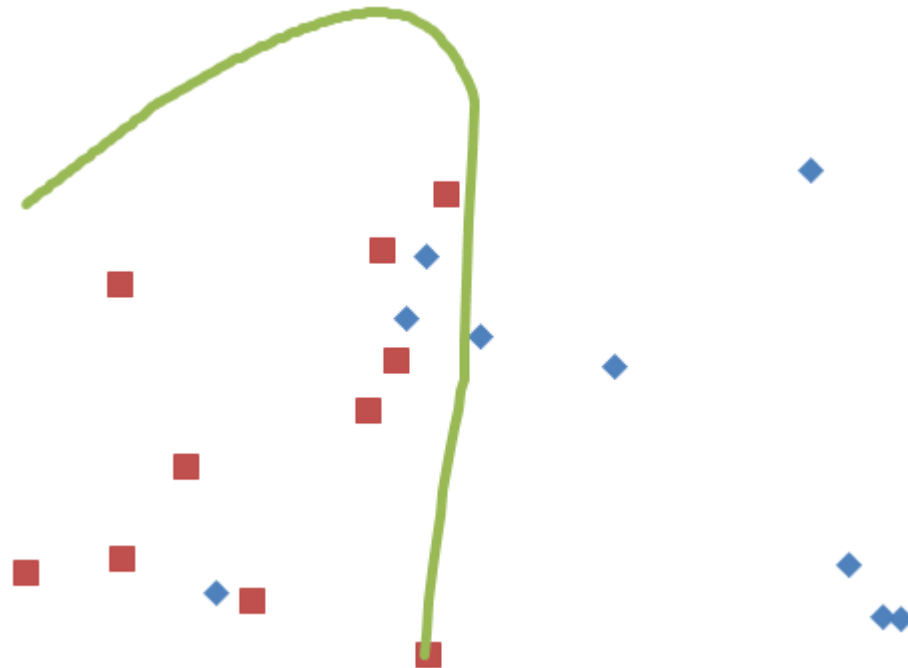
Classificadores paramétricos

- Classificador linear
 - Uma superfície de decisão (ou separação) que separa as amostras é obtida a partir da combinação linear dos atributos



Classificadores paramétricos

- Classificador quadrático
 - Uma superfície de decisão (ou separação) que separa as amostras é obtida a partir da combinação quadrática dos atributos



Classificadores paramétricos

- Considerar a distribuição normal tem uma série de vantagens
 - a regra de decisão resultante é robusta
 - se a hipótese for violada a degradação do desempenho do classificador é gradual.

Classificadores paramétricos

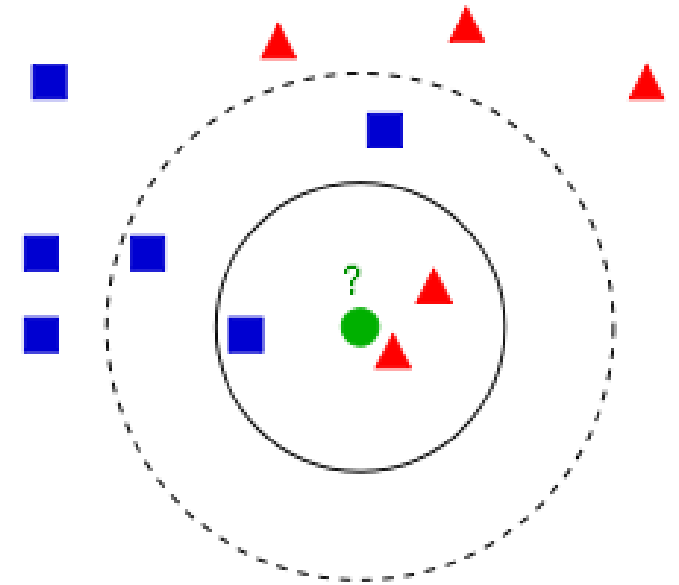
- No entanto, é difícil verificar se os dados multivariados possuem distribuição normal
 - Estima-se os dois parâmetros que caracterizam uma função gaussiana a partir dos dados de treino
 - média
 - desvio padrão

Classificadores não paramétricos

- São considerados os classificadores mais simples e intuitivos.
 - Não se conhece o comportamento estatístico dos dados
 - Não existe informação a priori sobre a distribuição de probabilidade ou as funções discriminantes

Classificadores não paramétricos

- Em geral, utilizam uma função de distância como medida de similaridade entre as amostras
 - Um objeto desconhecido é atribuído a classe a que menos se distancia
 - Uso da distância Euclidiana ou Mahalanobis



Classificadores não paramétricos

- Dados não precisam seguir uma distribuição normal
- Um dos algoritmos mais conhecidos é o *K-NN*
 - *k-NN* (*k-nearest neighbor* ou *k-vizinhos mais próximos*)

Classificadores não paramétricos

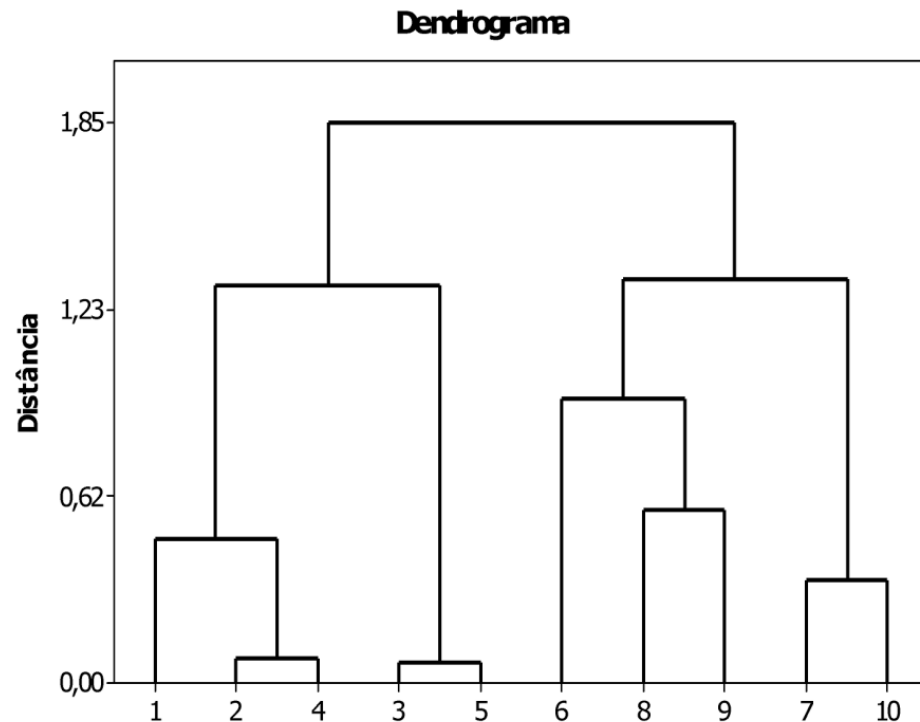
- Paramétrico ou não, qual usar?
 - Isso depende da credibilidade do modelo paramétrico
 - Com baixo número de amostras de treino as técnicas não paramétricas conduzem a melhores desempenhos do que as técnicas paramétricas, mesmo quando o modelo é correto

Métodos Hierárquicos e Não-Hierárquicos

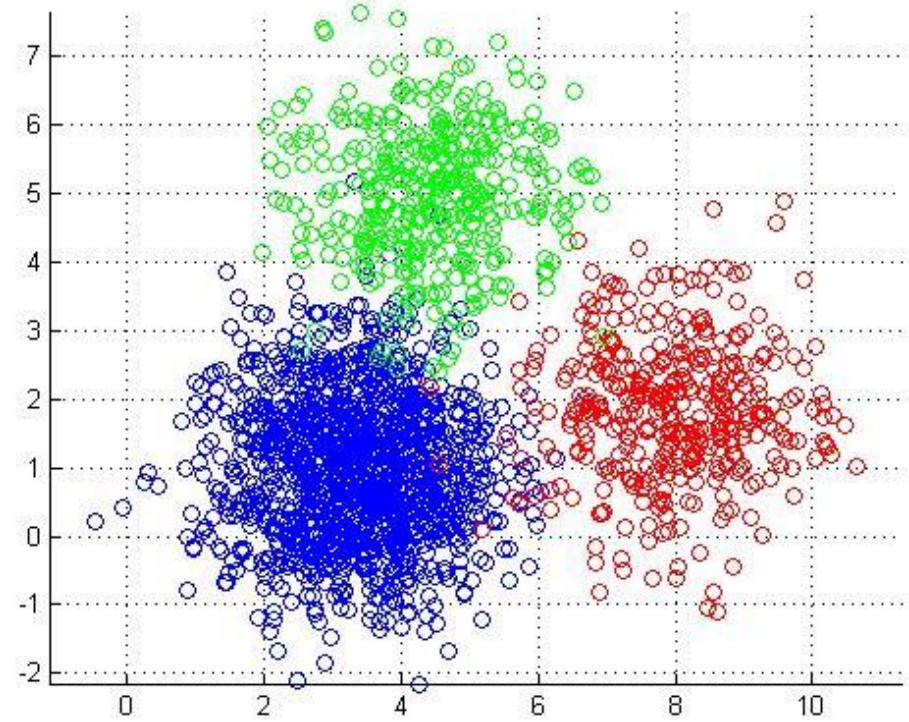
- Se referem principalmente a maneira como os dados são divididos e/ou organizados
 - Métodos Hierárquicos: constroem uma hierarquia de partições
 - Métodos Não-Hierárquicos ou Particionais: constroem uma partição dos dados

Métodos Hierárquicos e Não-Hierárquicos

Método Hierárquico



Método Não-Hierárquico

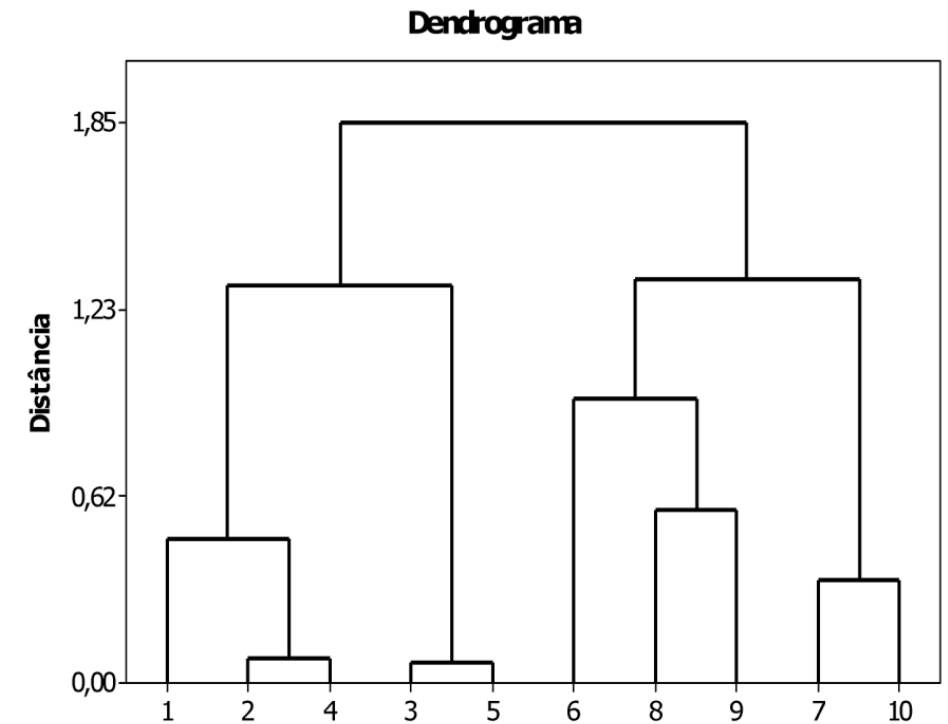


Métodos Hierárquicos e Não-Hierárquicos

- Algoritmos Hierárquicos
 - Criam uma hierarquia de relacionamentos entre os elementos.
 - Uso de uma medida de distância
 - Muito populares na área de bioinformática
 - Bom funcionamento
 - Apesar de não terem nenhuma justificativa teórica baseada em estatística ou teoria da informação, constituindo uma técnica ad-hoc de alta efetividade.

Métodos Hierárquicos e Não-Hierárquicos

- Algoritmos Hierárquicos
 - Dendrograma é talvez o algoritmo mais comum
 - Semelhante a uma árvore
 - Exemplo: relações evolutivas entre diferentes grupo de organismos biológicos (árvore filogenética)



Métodos Hierárquicos e Não-Hierárquicos

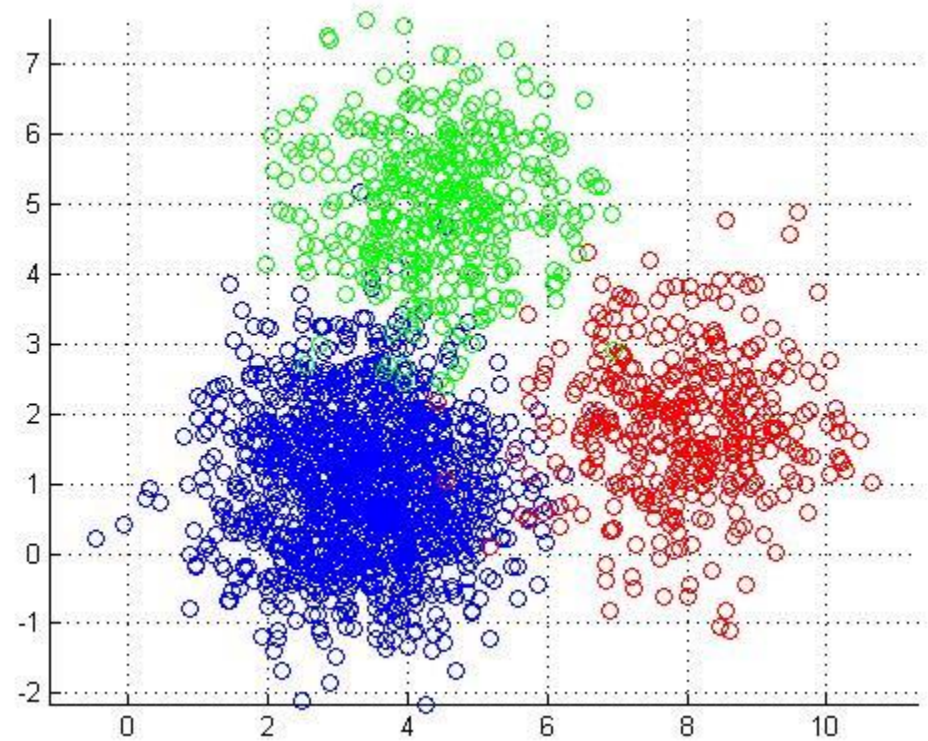
- Algoritmos Não-Hierárquicos
 - Separam os objetos em grupos baseando-se nas características que estes objetos possuem
 - Uso de uma medida de similaridade
 - Consistem de técnicas de análise de agrupamento ou *clustering*

Métodos Hierárquicos e Não-Hierárquicos

- Algoritmos Não-Hierárquicos
 - Normalmente dependem de uma série de fatores que são determinados de forma arbitrária pelo usuário
 - Número de conjuntos
 - Número de seeds de cada conjunto.
 - Esses parâmetros podem causar impacto negativo na qualidade das partições geradas

Métodos Hierárquicos e Não-Hierárquicos

- Algoritmos Não-Hierárquicos
 - K-means é talvez o algoritmo mais comum
 - Busca particionar n observações em k clusters (agrupamentos)
 - Cada observação pertence ao cluster com a média mais próxima



Template matching

- Técnica que utiliza uma abordagem natural na classificação de objetos ou padrões
 - Baseada na comparação de um objeto com um molde
- Muito utilizada em processamento de imagens
 - Encontrar as áreas de uma imagem que combinam (são semelhantes) a uma imagem modelo.
 - Comparação pixel a pixel

Template matching

- Essa técnica é frequentemente usada para identificar os caracteres impressos, números e outros pequenos objetos simples.
 - Contar o número de concordâncias
 - Máxima correlação
 - Contar o número de não concordâncias
 - Mínimo erro



Template matching

- O processo é simples e funciona quando os padrões são bem comportados
- Se o objeto é maior que o modelo, percorremos ele todo em busca de onde a combinação é máxima
 - Não funciona bem com distorções do tipo rotação, escala, etc



modelo



objeto

Template matching

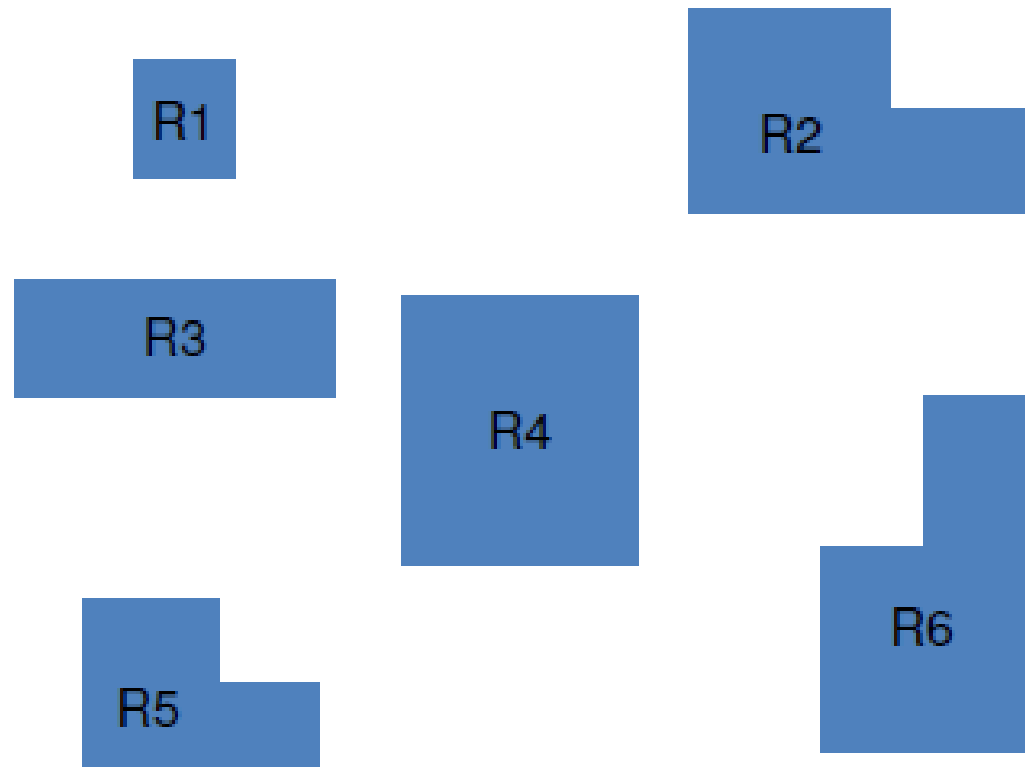
- A técnica pode ser utilizada se o desvio padrão do modelo em comparação com o objeto analisado for suficientemente pequeno.



Feature matching

- É uma variante do *template matching*
 - Ao invés de fazer a comparação pixel a pixel, a comparação se dá no nível das características
- Exemplo: momentos de HU
 - Característica Global e Invariante
 - Medidas puramente estatísticas da distribuição dos pontos
 - São invariantes a rotação, translação e escala

Feature matching | Exemplo



Feature matching | Exemplo

- Momentos de HU

Momento	R1	R2	R3	R4	R5	R6
1	1.67E-01	1.94E-01	2.08E-01	1.67E-01	1.94E-01	1.94E-01
2	0.00E+00	6.53E-03	1.56E-02	0.00E+00	6.53E-03	6.53E-03
3	0.00E+00	1.02E-03	0.00E+00	0.00E+00	1.02E-03	1.02E-03
4	0.00E+00	4.56E+05	0.00E+00	0.00E+00	4.56E+05	4.56E+05
5	0.00E+00	4.25E-09	0.00E+00	0.00E+00	4.25E-09	4.25E-09
6	0.00E+00	1.70E+06	0.00E+00	0.00E+00	1.70E+06	1.70E+06
7	0.00E+00	-8.85E+09	0.00E+00	0.00E+00	-8.85E+09	-8.85E+09

Regras de Decisão

- Forma de representação do conhecimento
 - Relaciona informações ou fatos a alguma ação ou resultado
- Definem um conjunto de condições que devem ocorrer para que uma determinada ação seja executada

Regras de Decisão

- Trata-se de um sistema especialista
 - O especialista detém conhecimento aprofundado e grande experiência prática no domínio do problema
 - Emula a estratégia de tomada de decisão de especialistas humanos
 - Atuam em um domínio restrito
 - Tem dificuldade de lidar com ambiguidades, pela presença de regras conflitantes

Regras de Decisão

- Componentes de um sistema especialista
 - Conjunto de regras: a base de conhecimento
 - Máquina de inferência: Infere conclusões a partir dos dados de entrada e da base de conhecimento.

Regras de Decisão

- Conjunto de regras: Tabela de Decisão
 - São fáceis de criar e de interpretar
 - Qual o resultado a operação “A ou B” para A = 1?
 - E para B = 0?

A	B	A ou B
0	0	0
0	1	1
1	0	1
1	1	1

Árvores de decisão (AD)

- São algoritmos que utilizam a estratégia de “dividir para conquistar”, chegando assim a classificação ou tomada de decisão sobre uma amostra
 - Divide problemas difíceis em problemas mais simples
 - “Imita” o processo de raciocínio humano

Árvores de decisão (AD)

- A idéia básica de uma AD é pegar um problema complexo e decompô-lo em sub-problemas menores, de modo que os novos problemas tenham uma complexidade menor em relação ao anterior
 - Essa estratégia é então aplicada recursivamente a cada sub-problema

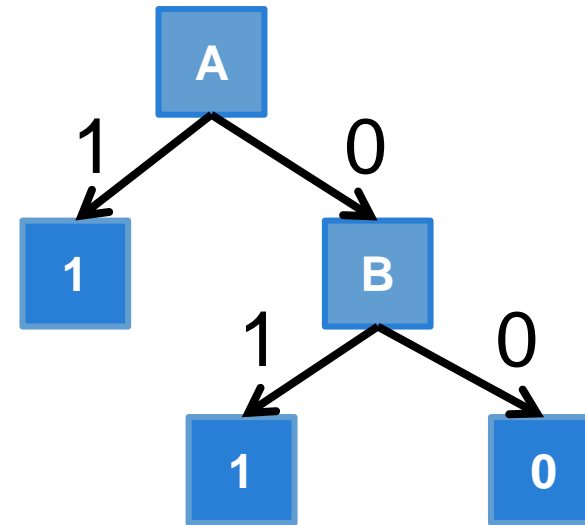
Árvores de decisão (AD)

- É um classificador baseado em um conjunto regras de decisão (ou tabela de decisão)
 - SE (condição) ENTÃO (resultado).
- É uma representação de uma tabela de decisão sob a forma de uma árvore
 - Tem a mesma utilidade da tabela de decisão.
 - É uma representação alternativa das mesmas regras obtidas com a construção da tabela.

Árvores de decisão (AD)

- Tabela de decisão e sua árvore de decisão

A	B	A ou B
0	0	0
0	1	1
1	0	1
1	1	1



Árvores de decisão (AD)

- AD é uma das técnicas de classificação mais utilizadas
 - Eficaz, eficiente e produz modelos interpretáveis
 - As condições de uma regra, em geral, envolvem intervalos para os atributos
 - A aplicação da sequência de regras vai classificando os objetos em classes cada vez menos abrangentes.

Árvores de decisão (AD)

- Um AD é composta por
 - Nós Intermediários ou raiz: definem o parâmetro que será avaliado
 - Arestas: definem a transição entre nós de acordo com os valores do atributo
 - Nós terminais: definem a classificação final da amostra

Árvores de decisão (AD)

- Com relação aos nós e arestas
 - Temos apenas um nó raiz
 - Nenhuma aresta de entrada e $n \geq 0$ arestas de saída
 - Vários nós intermediários
 - Possuem 1 aresta de entrada e $n > 1$ arestas de saída
 - Vários nós terminais
 - 1 aresta de entrada e nenhuma aresta de saída

Árvores de decisão (AD)

- Algoritmo Rudimentar (1 Rule - 1R)
 - Todas as regras usam somente um atributo
 - Escolher o atributo com a menor taxa de erro de classificação
 - Um ramo para cada valor do atributo
 - Para cada ramo
 - Atribuir a classe mais frequente
 - Calcular a taxa de erro de classificação:

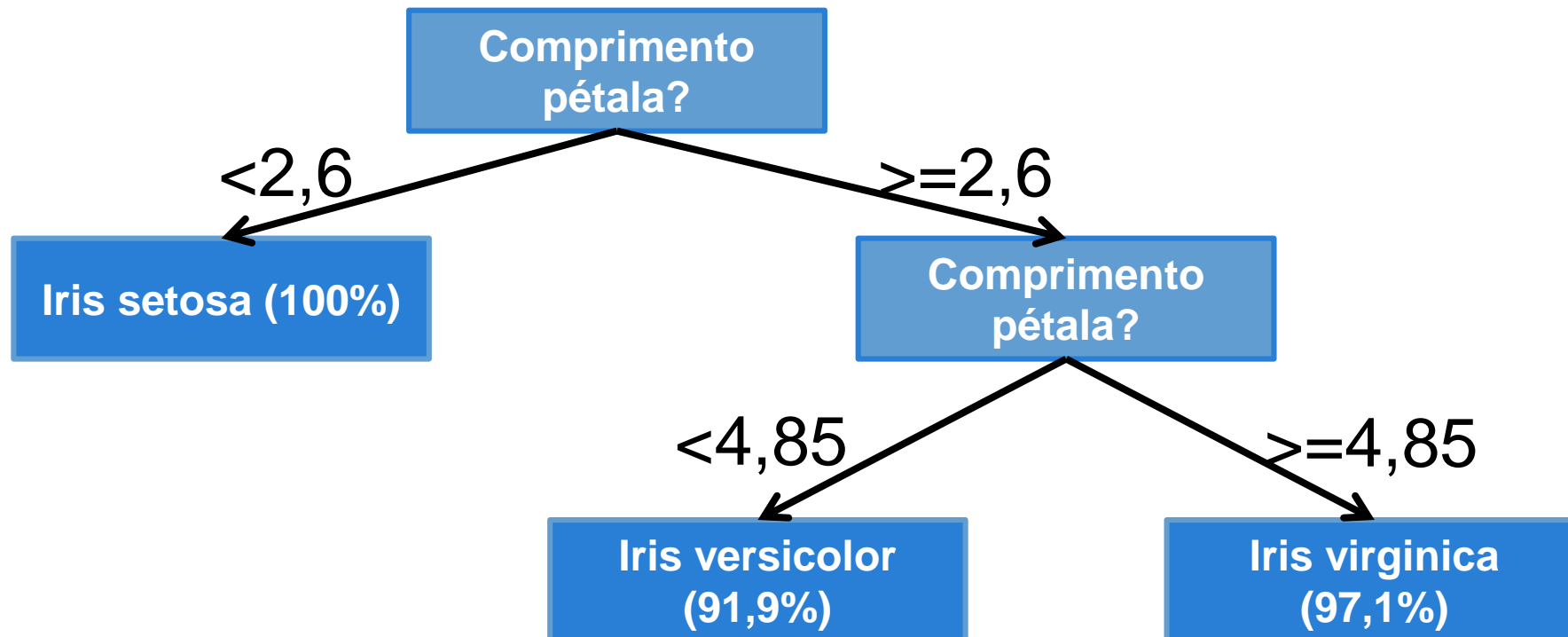
Árvores de decisão (AD)

- Exemplo: precisamos classificar 3 tipos de flores
 - Iris setosa
 - Iris virginica
 - Iris versicolor
- Como poderia ser sua AD?



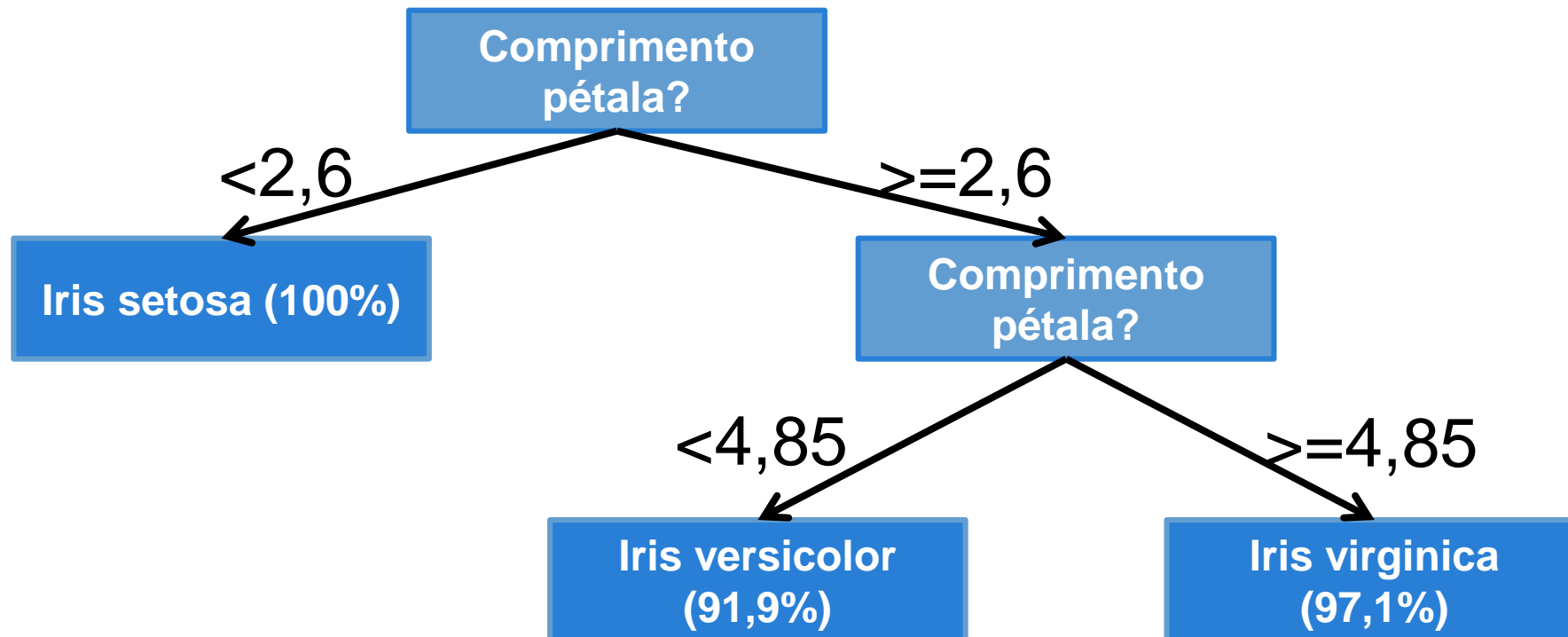
Árvores de decisão (AD)

- Exemplo de AD usada para classificar os 3 tipos de flores: Iris setosa, Iris virginica e Iris versicolor



Árvores de decisão (AD)

- Como ficaria classificada uma amostra com comprimento de pétala igual a **3,75**?
 - **3,75: Iris versicolor**

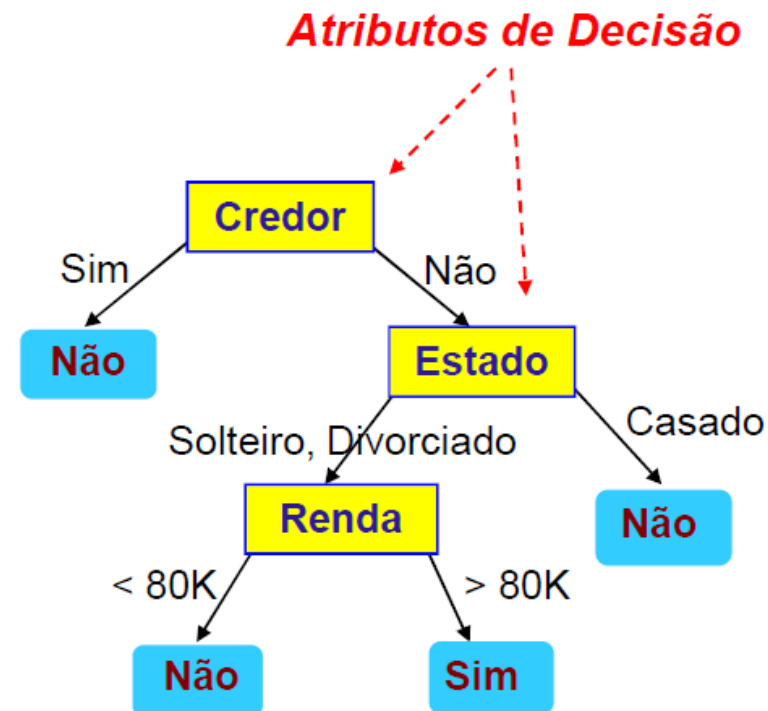


Árvores de decisão (AD)

- Exemplo de árvore de decisão com mais de 1 regra

<i>Id</i>	<i>E Credor</i>	<i>Estado Civil</i>	<i>Salário</i>	<i>Calote</i>
1	Sim	Solteiro	125K	Não
2	Não	Casado	100K	Não
3	Não	Solteiro	70K	Não
4	Sim	Casado	120K	Não
5	Não	Divorciado	95K	Sim
6	Não	Casado	60K	Não
7	Sim	Divorciado	220K	Não
8	Não	Solteiro	85K	Sim
9	Não	Casado	75K	Não
10	Não	Solteiro	90K	Sim

Dados de Treinamento



Modelo: Árvore de Decisão

Árvores de decisão (AD)

- Vantagens de se usar uma árvore de decisão
 - Uma vez construída, o seu uso é imediato
 - Em termos computacionais, é uma ferramenta muito rápida
 - Possui fácil interpretação

Árvores de decisão (AD)

- Desvantagens

- A construção de uma árvore de decisão é feita por processo de indução, pode ser uma tarefa de alta demanda computacional.
 - Cansativo quando se tem muitos atributos
- Bias indutivo: o processo de indução possui preferência de uma hipótese sobre outras, supondo a existência de hipóteses que são igualmente consistentes.

Árvores de decisão (AD)

- Desvantagens
 - A partir de um conjunto de atributos, podemos obter diversas árvores de decisão
 - Isso torna impraticável a busca por uma árvore de decisão ótima para um determinado problema

Árvores de decisão (AD)

- Alguns algoritmos existentes para construir uma árvore de decisão
 - Hunt's Concept Learning System
 - Um dos primeiros
 - Serve de base para vários outros
 - ID3, C4.5, J4.8, C5.0
 - CART, Random-Forest

Árvores de decisão (AD) | Algoritmo de Hunt

- Considere
 - D_t o conjunto de objetos que atingem o nó t
 - Esses objetos ainda não foram classificados em um nó folha acima na árvore
- Passo 1
 - Se todos os objetos de D_t pertencem à mesma classe c_t , então t é um nó folha rotulado como c_t

Árvores de decisão (AD) | Algoritmo de Hunt

- Passo 2: Se D_t contém objetos que pertencem a mais de uma classe, então t deve ser um nó interno
 - Escolher uma condição de teste sobre algum dos atributos que não houverem sido selecionados nos nós acima na árvore
 - Criar um nó filho para cada possível saída da condição de teste (valor do atributo). Os objetos em D_t são distribuídos neles
 - Aplicar o algoritmo recursivamente para cada nó filho criado

Árvores de decisão (AD) | Critério de Parada

- Podemos finalizar a indução da árvore quando
 - Os dados do nó atual têm o mesmo rótulo
 - Os dados do nó atual têm os mesmos valores para os atributos de entrada, mas rótulos de classes diferentes
 - Todos os atributos já foram utilizados no caminho

Avaliação da Qualidade dos Classificadores

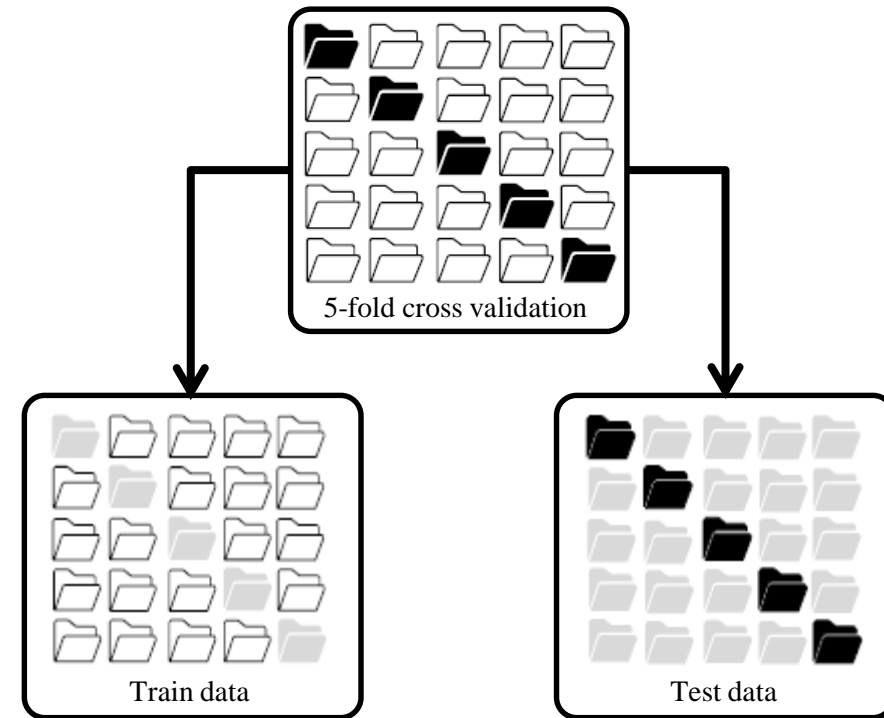
- Avaliar a qualidade de um classificador não é uma tarefa fácil
 - O que devemos avaliar?
 - Exatidão?
 - Reproducibilidade?
 - Robustez?
 - Capacidade em utilizar toda a informação disponível?
 - Aplicabilidade?
 - Objetividade?
- Nenhum classificador satisfaz tudo isso

Avaliação da Qualidade dos Classificadores

- De modo geral, espera-se que um classificador apresente desempenho adequado para dados não vistos
 - Acurácia
 - Pouca sensibilidade ao uso de diferentes amostras de dados

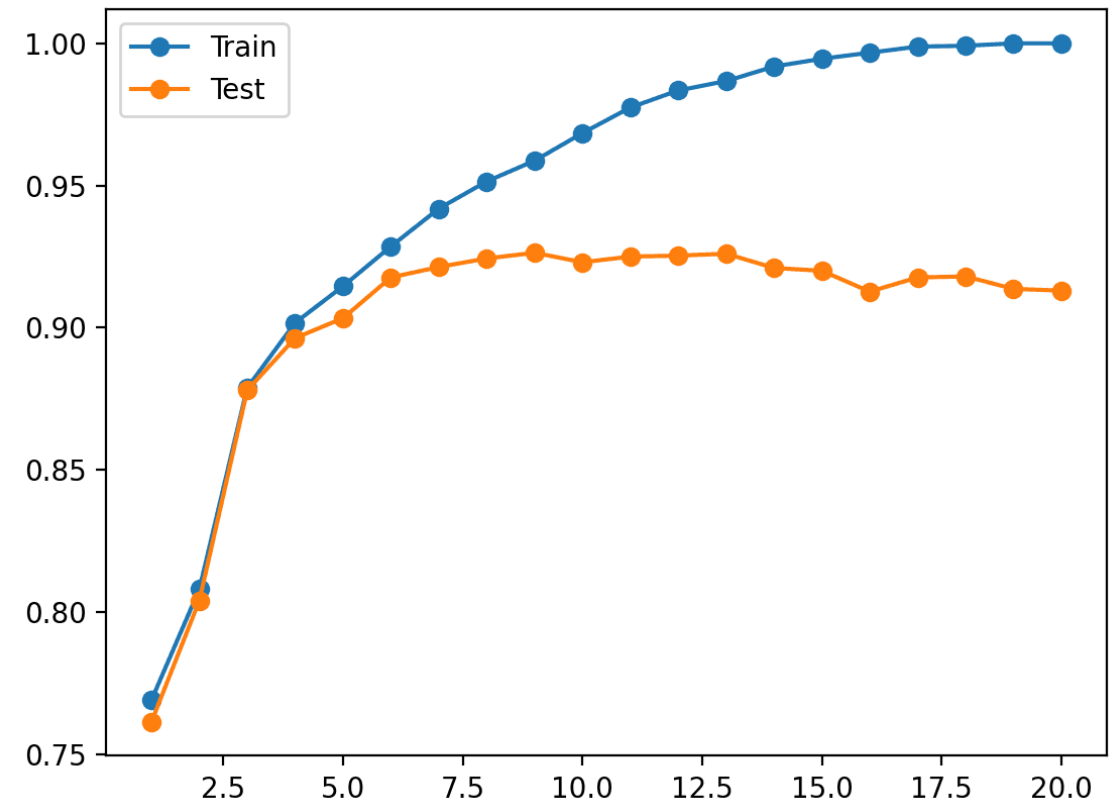
Avaliação da Qualidade dos Classificadores

- Podemos avaliar o desempenho usando conjuntos distintos de treinamento e teste
 - Treinamento
 - Usado para estimar os parâmetros do classificador.
 - Teste (ou validação)
 - Usado para validar o classificador



Avaliação da Qualidade dos Classificadores

- Vantagens de usar treinamento e teste
 - Estima a capacidade de generalização do modelo
 - Avalia a sua variância (estabilidade)



Avaliação da Qualidade dos Classificadores

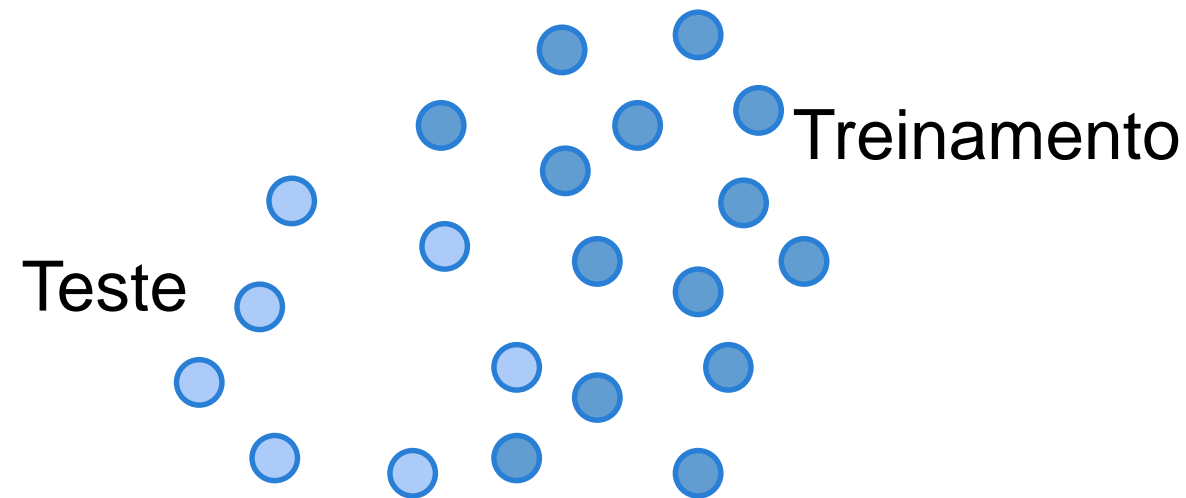
- Como obter os conjuntos de treinamento e teste?
 - Particionar o conjunto de dados conhecidos em subconjuntos mutualmente exclusivos
 - Utilizar alguns destes subconjuntos para treinamento e o restante dos subconjuntos para o teste ou validação do modelo.

Avaliação da Qualidade dos Classificadores

- Hold out (ou *Split-Sample*)
 - Uma das técnicas mais simples. Faz uma única partição da amostra
 - A maior parte dos dados é usada como dados de treinamento, e o restante é usado como dados de teste
 - Divisão dos dados
 - Treinamento: 2/3 (por exemplo)
 - Teste: dados restantes

Avaliação da Qualidade dos Classificadores

- Exemplo Hold out



Avaliação da Qualidade dos Classificadores

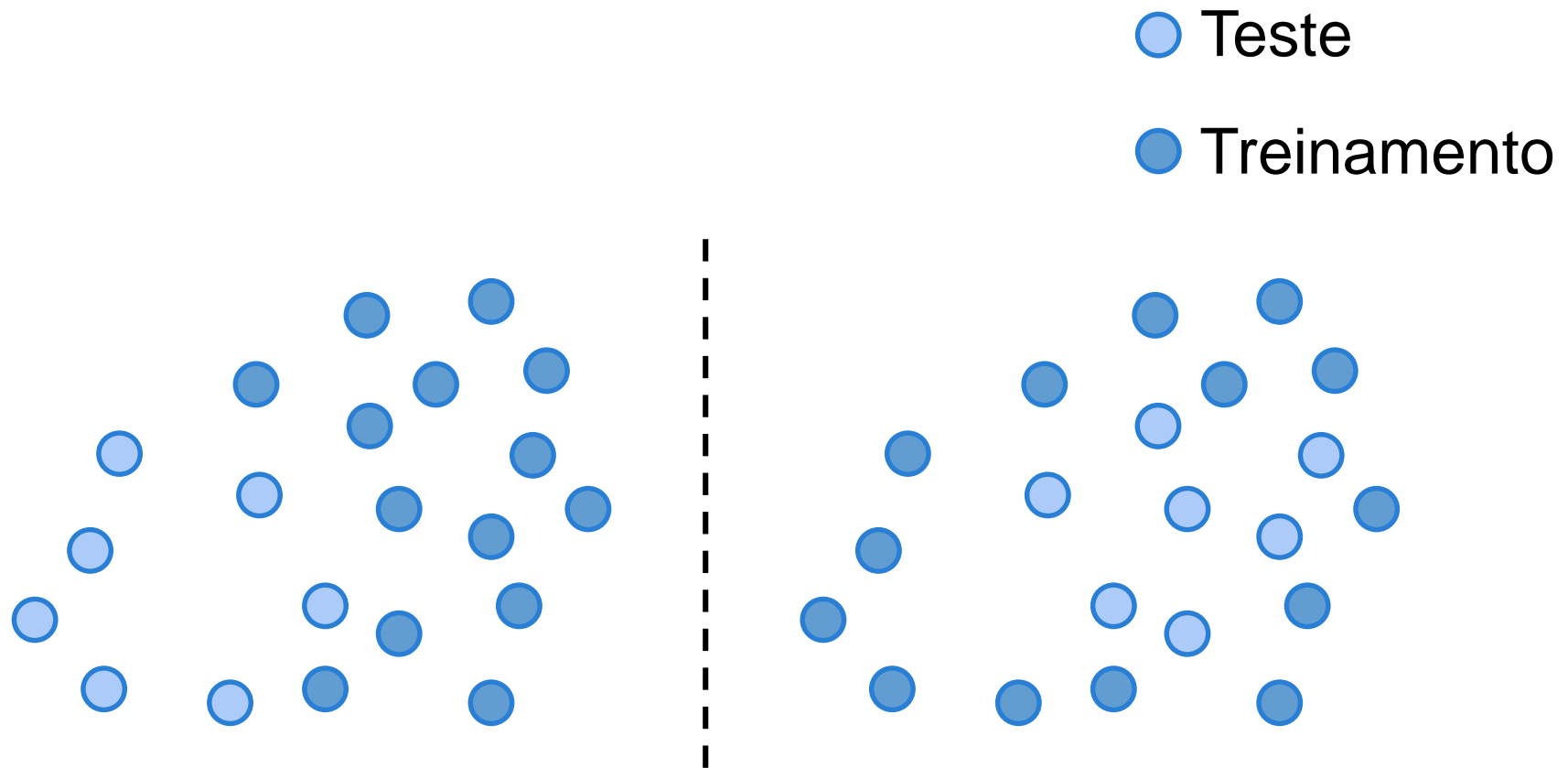
- Hold out (ou *Split-Sample*)
 - Indicado para uma grande quantidade de dados.
 - Problemas em pequena quantidade de dados
 - Poucos dados de treinamento
 - Menor o conjunto de treinamento, maior a variância do classificador (instabilidade)
 - Menor o conjunto de teste, menos confiável
 - Treinamento e teste podem não ser independentes
 - Classe sub-representada ou super-representada

Avaliação da Qualidade dos Classificadores

- Random Subsampling
 - Múltiplas execuções de Holdout, com diferentes partições treinamento-teste escolhidas de forma aleatória
 - Não pode haver interseção entre os conjuntos de teste e treinamento
 - Permite uma estimativa de erro mais precisa
 - Erro de classificação
 - Média da taxa de erro de classificação obtida para cada execução

Avaliação da Qualidade dos Classificadores

- Random Subsampling (2 execuções)



Avaliação da Qualidade dos Classificadores

- Validação cruzada (ou *cross validation*)
 - Classe de métodos de particionamento de dados
 - Usado para estimativa da taxa de erro verdadeira
 - É comumente usado quando a quantidade de dados disponível é pequena

Avaliação da Qualidade dos Classificadores

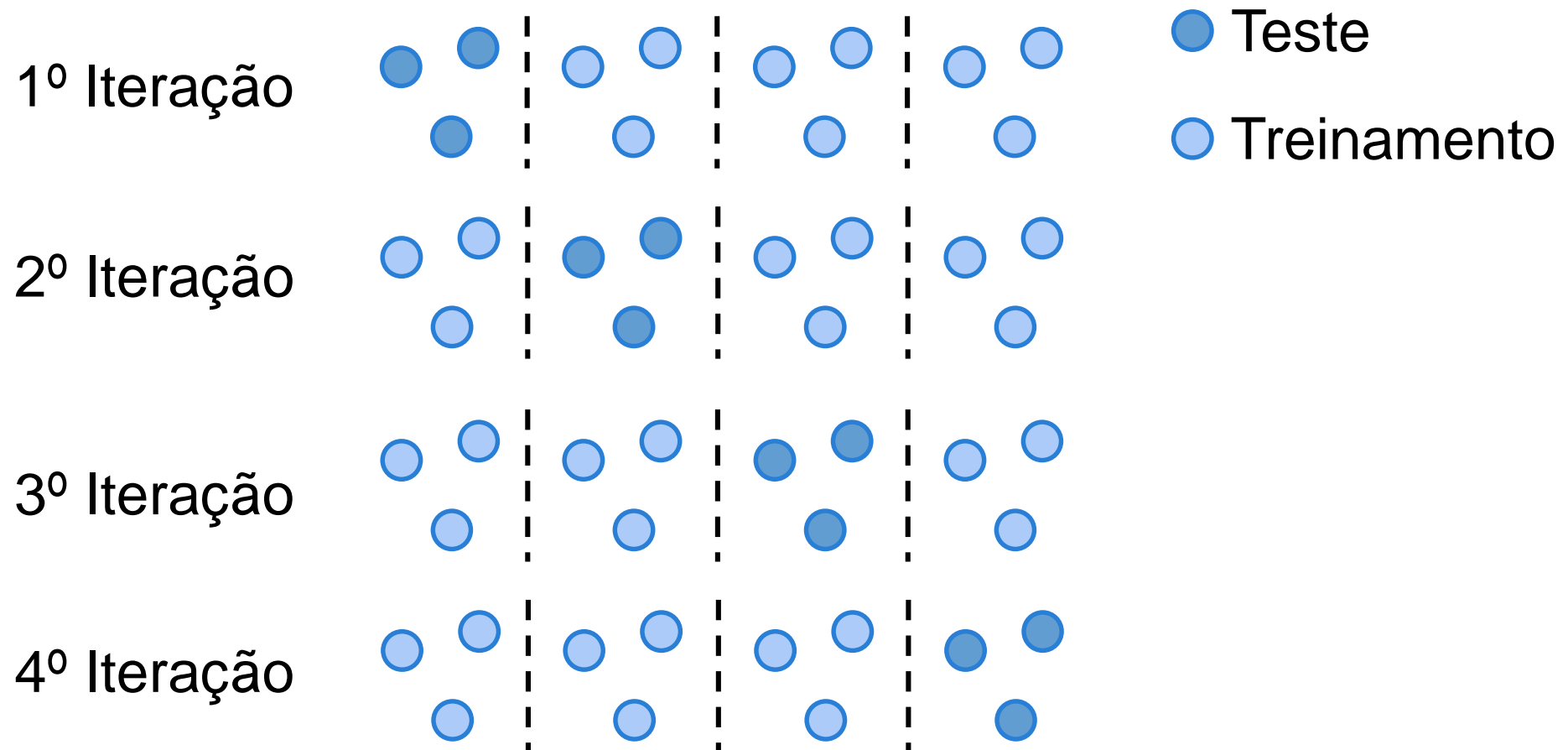
- Validação cruzada (ou *cross validation*)
 - Consiste em particionar o conjunto de dados em subconjuntos mutualmente exclusivos
 - Utiliza-se alguns subconjuntos para treinamento e o restante para teste
 - Métodos
 - k-fold
 - leave-one-out

Avaliação da Qualidade dos Classificadores

- Validação cruzada: método k-fold
 - consiste em dividir o conjunto total de dados em k subconjuntos mutuamente exclusivos do mesmo tamanho
 - A cada iteração, uma das k partições é usada para testar o modelo
 - As outras $k-1$ são usadas para treinar o modelo

Avaliação da Qualidade dos Classificadores

- Exemplo: k-fold ($k = 4$)



Avaliação da Qualidade dos Classificadores

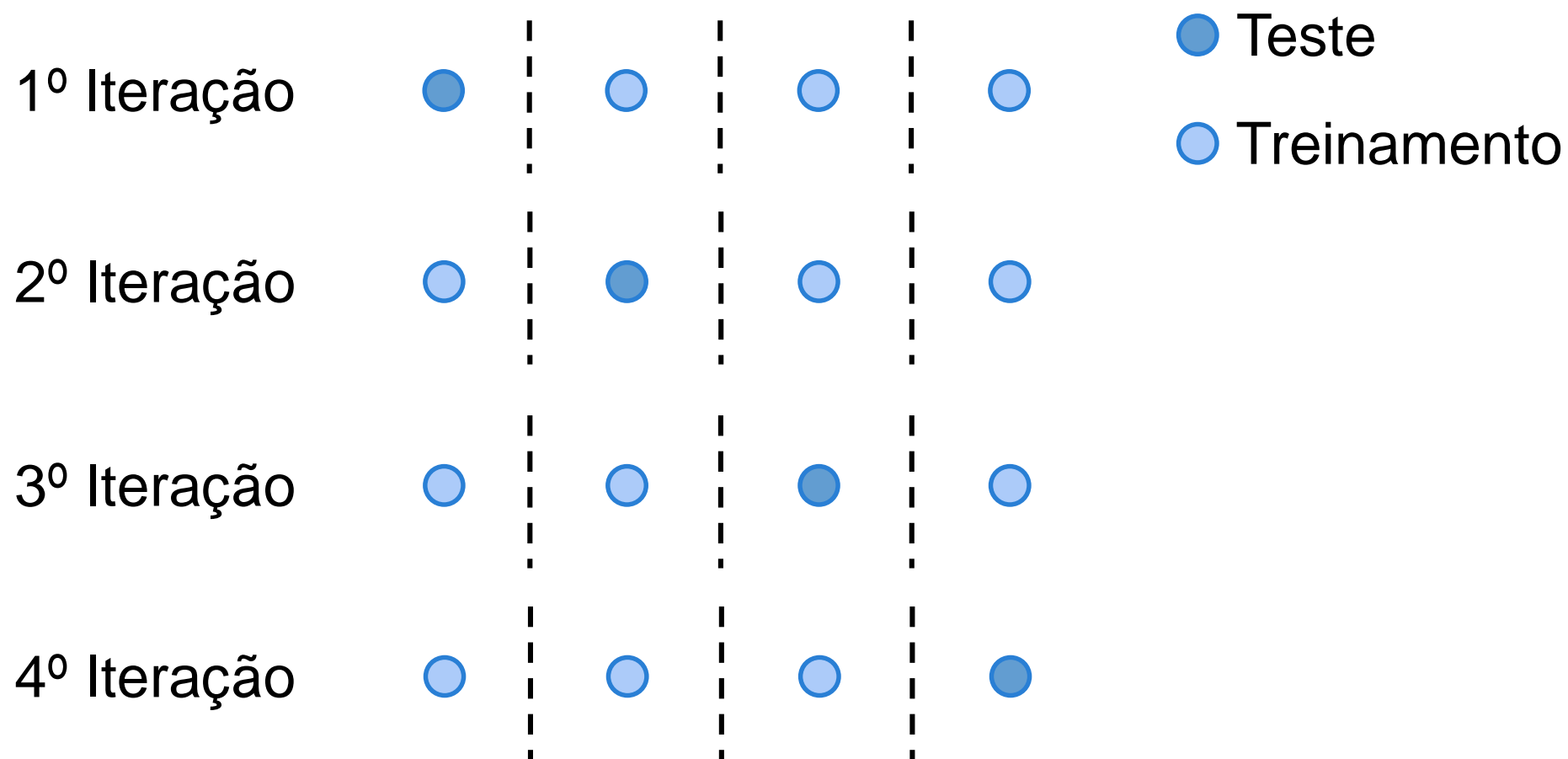
- Validação cruzada: método k-fold
 - Cada objeto participa o mesmo número de vezes do treinamento
 - **$k-1$ vezes**
 - Cada objeto participa o mesmo número de vezes do teste
 - **1 vez**
 - Taxa de erro é tomada como a média dos erros de validação das **k** partições

Avaliação da Qualidade dos Classificadores

- Validação cruzada: método leave-one-out
 - Trata-se de um caso específico do k-fold
 - Nesse caso, o valor de k é igual ao número total de dados N
 - **10-fold** se aproxima do **leave-one-out**

Avaliação da Qualidade dos Classificadores

- Exemplo: leave-one-out ($N = 4$)



Avaliação da Qualidade dos Classificadores

- Validação cruzada: método leave-one-out
 - Cada objeto participa o mesmo número de vezes do treinamento
 - **$N-1$ vezes**
 - Cada objeto participa o mesmo número de vezes do teste
 - **1 vez**
 - Taxa de erro é obtida dividindo por **N** o número total de erros de validação observados

Avaliação da Qualidade dos Classificadores

- Validação cruzada: método leave-one-out
 - Vantagem
 - Investigação completa sobre a variação do modelo em relação aos dados utilizados
 - Estimativa de erro é não tendenciosa, ou seja, tende à taxa verdadeira
 - Desvantagem
 - Alto custo computacional
 - Indicado para uma quantidade pequena de dados

Avaliação da Qualidade dos Classificadores

- Bootstrap
 - É uma técnica de reamostragem
 - Visa a obtenção de um “novo” conjunto de dados, por reamostragem do conjunto de dados original
 - Ao invés de usar sub-conjuntos dos dados, usa-se sub-amostras
 - Funciona melhor que cross-validation para conjuntos muito pequenos
 - Existem diversas variações: Bootstrap .632, etc.

Avaliação da Qualidade dos Classificadores

- Bootstrap
 - A amostragem é feita com reposição
 - Dado um conjunto com N objetos
 - Sorteia-se um objeto para compor a sub-amostra
 - Devolve-se o objeto sorteado ao conjunto de dados
 - Repete-se esse processo até compor uma sub-amostra de tamanho N

Avaliação da Qualidade dos Classificadores

- Bootstrap
 - Conjuntos de Treinamento e Teste
 - A sub-amostra gerada será o conjunto de treinamento
 - Os objetos restantes (que não fazem parte do treinamento) são o conjunto de teste
 - De modo geral, a sub-amostra tem 63,2% de objetos não repetidos
 - Processo é repetido b vezes
 - O resultado é a média dos experimentos

Agradecimentos

- Agradeço aos professores
 - Guilherme de Alencar Barreto - Universidade Federal do Ceará (UFC)
 - Prof. Ricardo J. G. B. Campello – ICMC/USP
- pelo material disponibilizado