

APRENDIZADO DE MÁQUINA

COMBINANDO MODELOS

PROF. RONALDO CRISTIANO PRATI

ronaldo.prati@ufabc.edu.br

Bloco A, sala 513-2

POR QUE COMBINAR MODELOS?

- Quando precisamos decidir sobre uma questão crítica, usualmente consultamos vários *experts* da área ao invés de confiarem no julgamento de um único consultor
- Em aprendizado de máquina, um modelo pode ser considerado como um “expert”. Então combina-los é uma boa idéia?

VANTAGENS

- Experimentalmente tem sido mostrado que modelos combinados apresentam melhores desempenhos do que um sistema decisório único
- Melhor do que o melhor modelo selecionado usando cross validation.
- Neutraliza ou minimiza drasticamente a instabilidade inerente dos algoritmos de aprendizagem.

VANTAGENS

- Sistemas combinados reduzem a variância (decomposição bias-variance)
- Em geral, quanto maior for o número de classificadores combinados, maior a redução da variância

DESVANTAGENS

- Apesar de normalmente os sistemas combinados apresentarem melhores resultados, não há garantias que isto ocorrerá sempre.
- Ainda é uma área de pesquisa com muito pontos para serem confirmados teoricamente.
- Modelos combinados são mais difíceis de analisar.
- Custo computacional

COMO COMBINAR?

- Três aspectos a serem analisados na combinação de modelos
 - A escolha da estrutura do sistema
 - A escolha dos componentes do sistema
 - A escolha do método de combinação

COMO COMBINAR?

- A estrutura do sistema
 - A maneira como os componentes estão organizados dentro do sistema
 - Quantos métodos serão necessários e como organizá-los?
- Tipos:
 - Ensemble
 - Modular

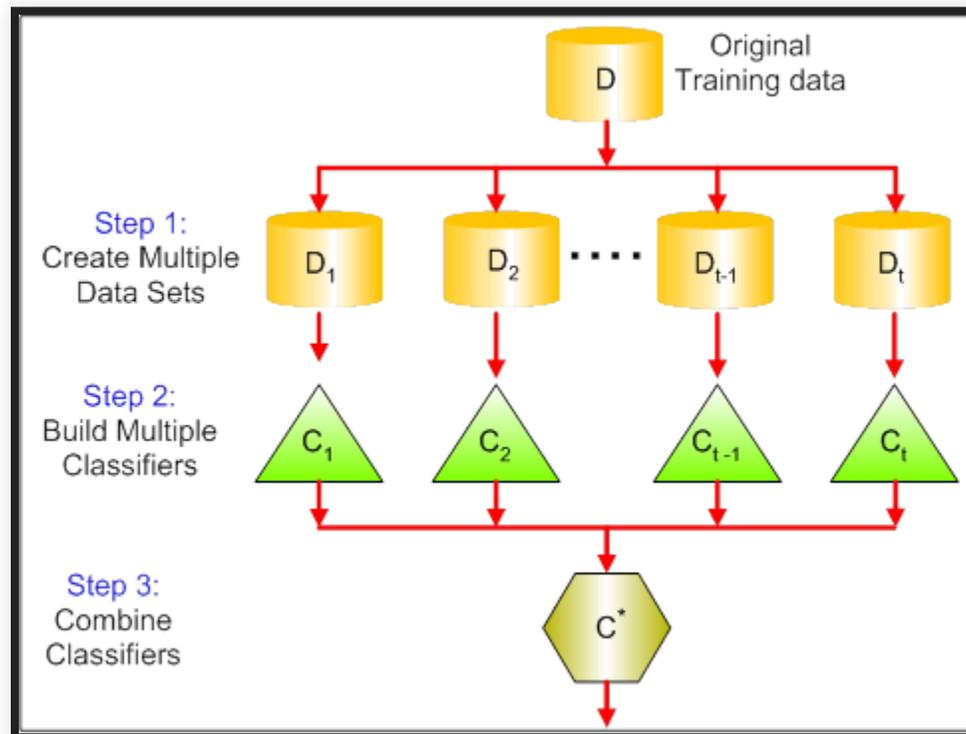
ENSEMBLES

- Também conhecido por vários outros nomes:
 - Multiple classifier systems, committee of classifiers, ou mixture of experts.
- Tem sido utilizado com sucesso em problemas onde um único modelo não funciona bem.
- Bons resultados são encontrados em várias aplicações em uma larga variedade de cenários

COMITÊS

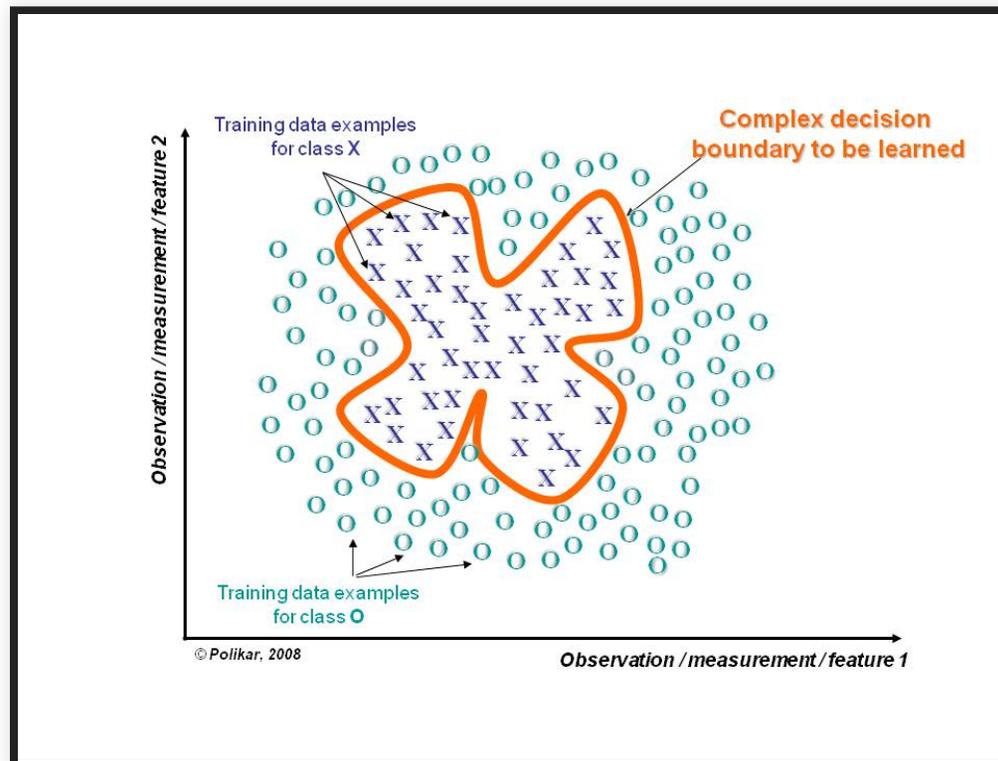
- Abordagem: redundante ou paralela de combinação
- Modelos: treinados com a mesma tarefa
- Suas respostas são combinadas para produzir estimativas mais confiáveis

ENSEMBLES



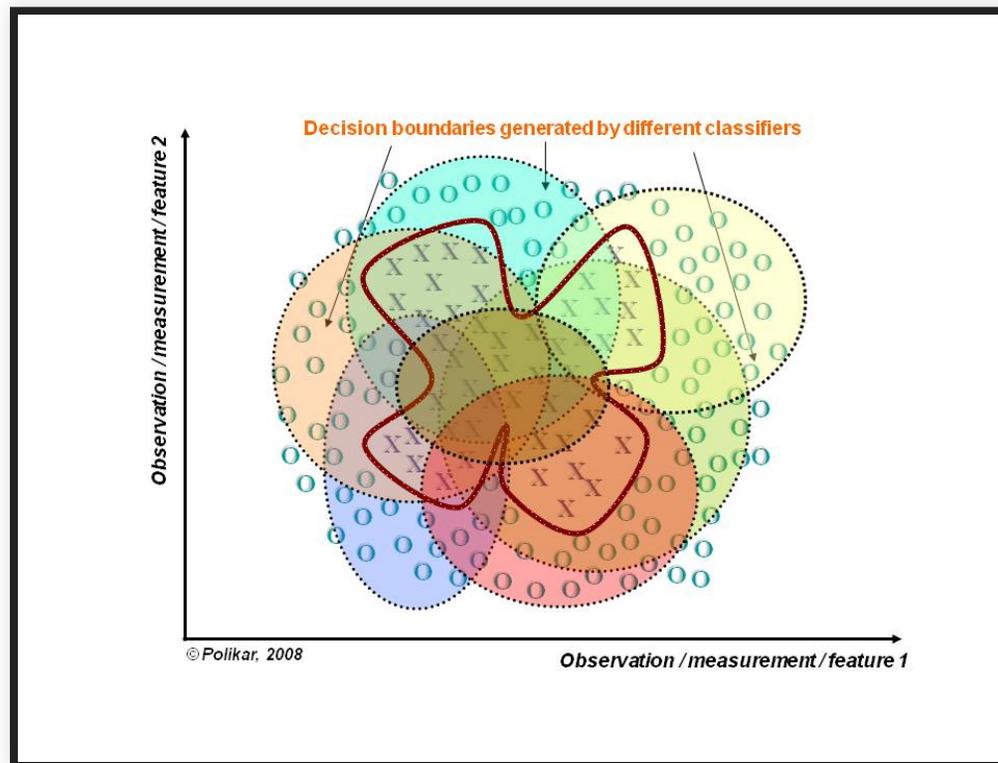
DIVIDIR E CONQUISTAR

- Independente da quantidade de dados alguns problemas são muito difíceis de serem resolvidos por um dado classificador



DIVIDIR E CONQUISTAR

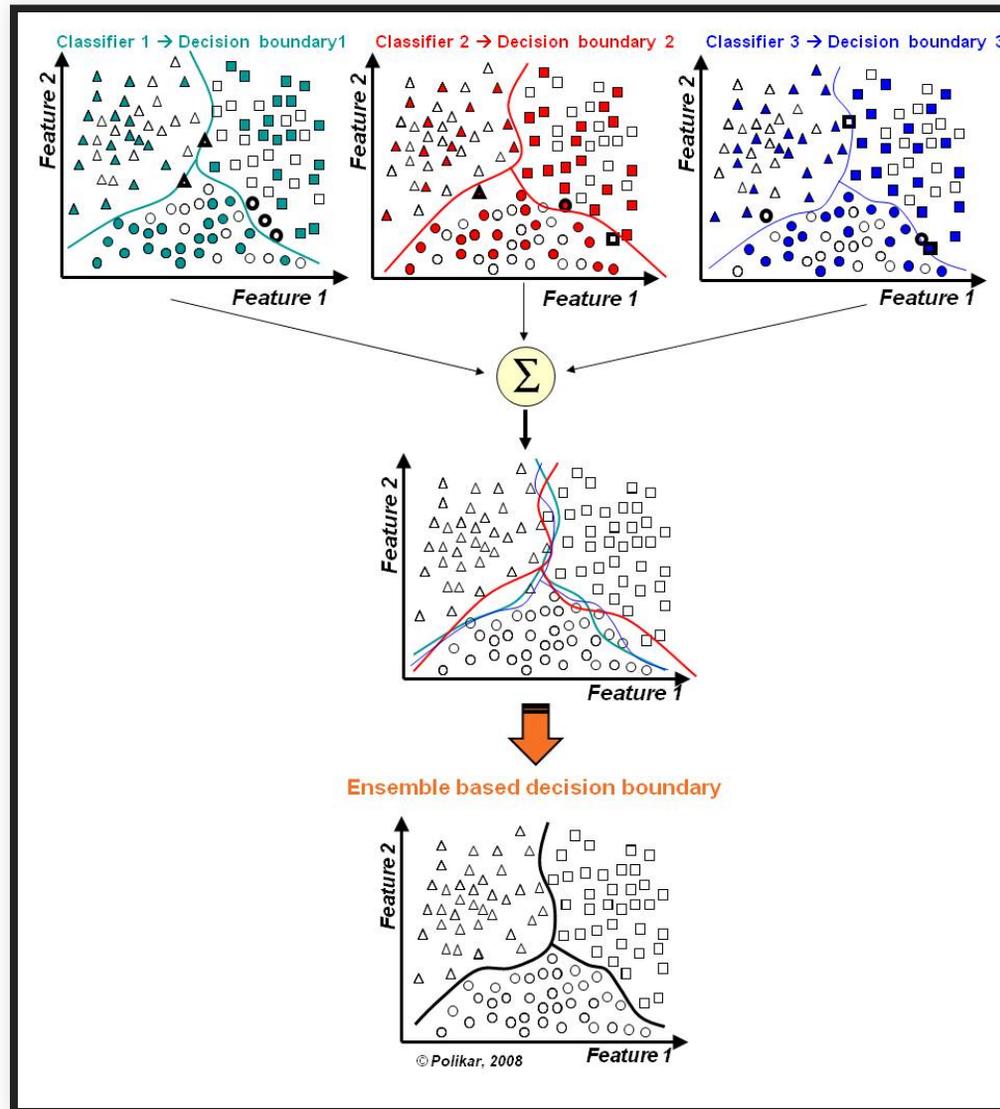
A fronteira de decisão que separa os dados de diferentes classes pode ser muito complexa ou estar fora do escopo do classificador.



DIVIDIR E CONQUISTAR

- A idéia de ensembles é que o sistema de classificação siga a abordagem dividir-para-conquistar;
- O espaço de dados é dividido em porções menores e mais “fáceis” de aprender por diferentes classificadores;
- Assim a linha base da fronteira de decisão pode ser aproximada por meio de uma combinação apropriada dos diferentes classificadores.

DIVIDIR E CONQUISTAR



CLASSIFICADORES "FRACOS"

- Muitos ensembles usam modelos "fracos" (*weak classifier*)
 - Hipóteses simples, de rápido treinamento, e que tem um desempenho um pouco superior a uma predição Aleatóriidade
 - Pequeno *bias*, alta variância.
- A combinação das saídas produzidas pelos classificadores reduz o risco de escolha por um classificador com um pobre desempenho

ENSEMBLES

- Os modelos são componentes que fornecem redundância
 - Uma solução para o mesmo problema, mesmo que usando meios diferentes
- Um aspecto importante é diversidade
 - Não há nenhuma vantagem em um ensemble com métodos idênticos

DIVERSIDADE

- O sucesso de um ensemble e a habilidade em corrigir erros de alguns de seus membros, depende fortemente da diversidade dos classificadores que o compõem;
- Cada classificador DEVE ter diferentes erros em diferentes exemplos dos dados;
- A idéia é construir muitos models e então combinar suas saídas de modo que o desempenho final seja melhor do que o desempenho de um único classificador;

DIVERSIDADE

- A diversidade de classificadores pode ser obtida de diferentes formas:
 1. Nos dados
 2. Nos modelos
 3. Nas técnicas de modelagem
 4. Na aleatoriedade

DIVERSIDADE

- Dados:
 - usar amostras diferentes (por exemplo, gerar diferentes amostras com reposição)
 - usar conjuntos de atributos diferentes (selecionar alatóriamente subconjuntos de atributos)

DIVERSIDADE

- Modelos:
 - usar parâmetros diferentes (diferentes valores de regularização, camadas em redes neurais, etc.)
 - atribuir pesos diferentes a a modelos

DIVERSIDADE

- Técnicas de modelagem:
 - Usar algoritmos diferentes
 - usar *kernels* diferentes para uma mesma base de Dados

DIVERSIDADE

- Aleatóriedade
 - Usar pesos diferentes em nas inicializações
 - Decidir empates aleatóreamente

ENSEMBLES TAMBÉM SÃO ÚTEIS:

- Grandes volumes de dados
 - A quantidade de dados é grande para ser manipulada por um único classificador.
 - Particionar os dados em sub-conjuntos e treinar diferentes classificadores com diferentes partições dos dados e então combinar as saídas com uma inteligente regra de combinação.

ENSEMBLES TAMBÉM SÃO ÚTEIS:

- Pequenos volumes de dados
 - Quando há ausência de dados de treinamento técnicas de amostragem podem ser utilizadas para a criação de subconjuntos de dados aleatórios sobrepostos
 - Cada subconjunto é utilizado para treinar diferentes classificadores e então criar ensembles com desempenhos melhores a modelos com a base reduzida.

TÉCNICAS DE CRIAÇÃO DE ENSEMBLES

- As técnicas mais conhecidas que combinam modelos para problemas de regressão e classificação são:
 - BAGGING
 - BOOSTING

BAGGING

- Possui uma implementação simples e intuitiva;
- A diversidade é obtida com o uso de diferentes subconjuntos de dados aleatoriamente criados com reposição;
- Cada subconjunto é usado para treinar um classificador do mesmo tipo;
- As saídas dos classificadores são combinadas por meio do voto majoritário com base em suas decisões;

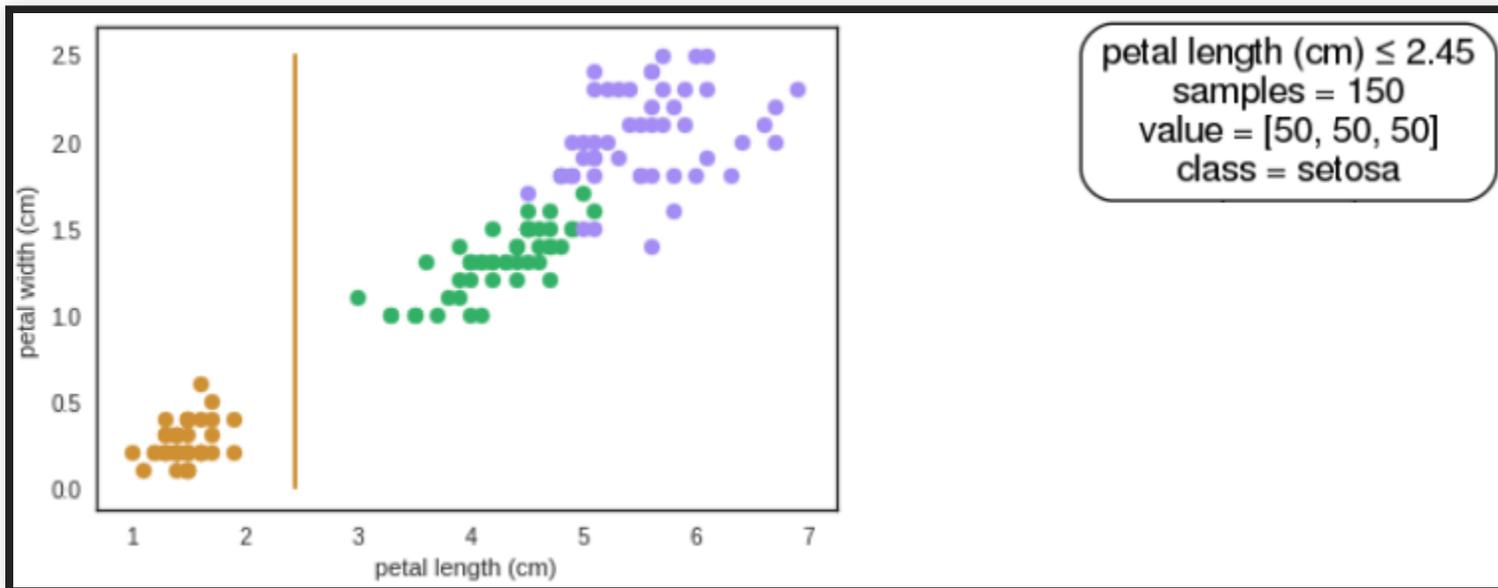
EXEMPLO: RANDOM FORESTS

- Usado para a construção de ensembles com árvores de decisão;
- Variação da quantidade de dados e atributos;
- Usando árvores de decisão com diferentes inicializações;

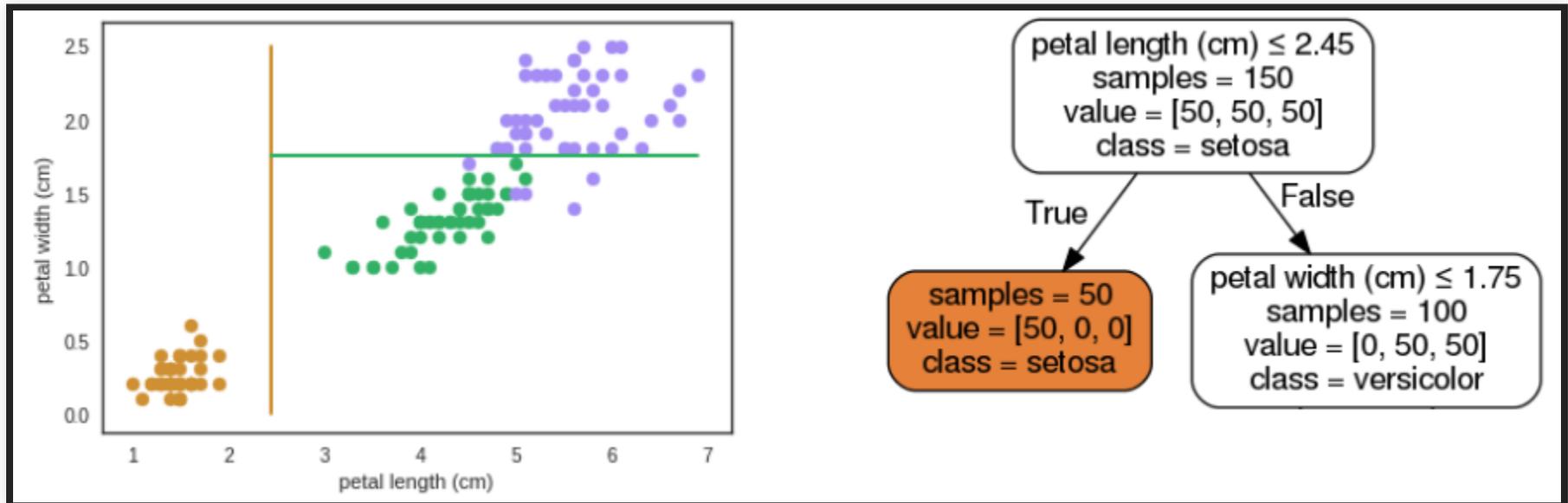
ÁRVORE DE DECISÃO

- Divide recursivamente o espaço de exemplos:
 1. Escolhe um atributo
 2. Particiona os dados de acordo com esse atributo
 3. Para cada subconjunto gerado, repete volta ao passo 1 até que o subconjunto seja homogêneo

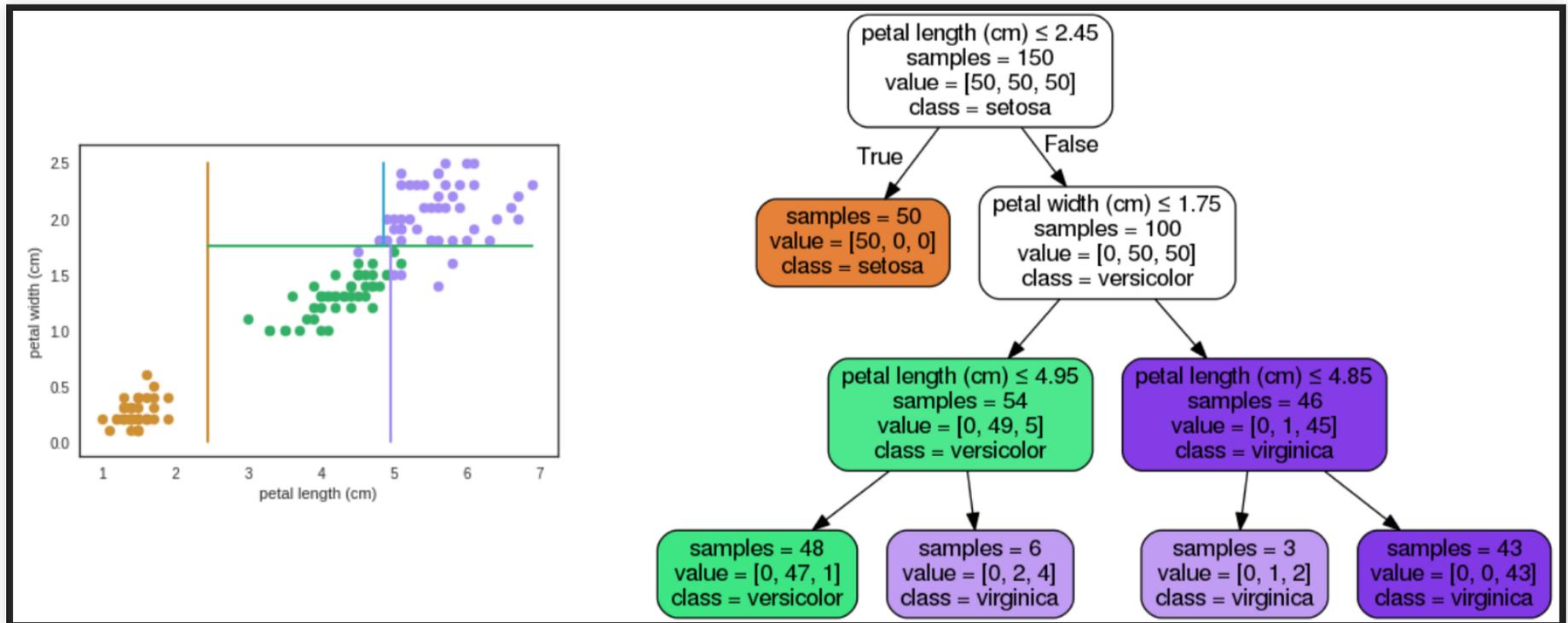
ÁRVORE DE DECISÃO



ÁRVORE DE DECISÃO



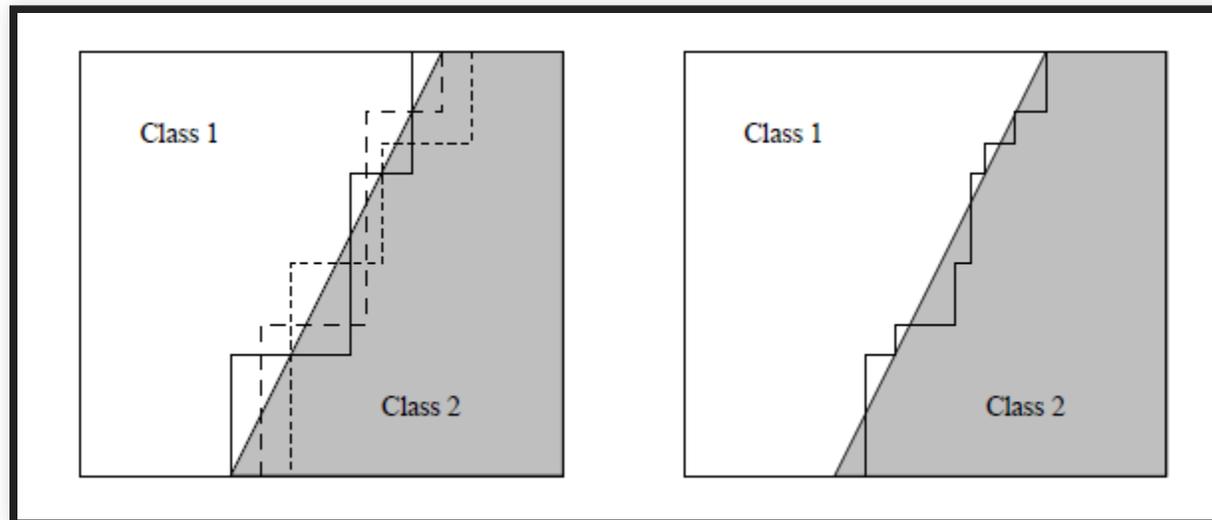
ÁRVORE DE DECISÃO



EXEMPLO: RANDOM FORESTS

- Por que usar uma floresta?
 - Uma única árvore complexa requer muitos exemplos
 - Como a escolha do atributo é heurística, um escolha ruim compromete toda a sub-árvore
 - Uma combinação de várias árvores simples pode gerar uma fronteira de decisão complexa

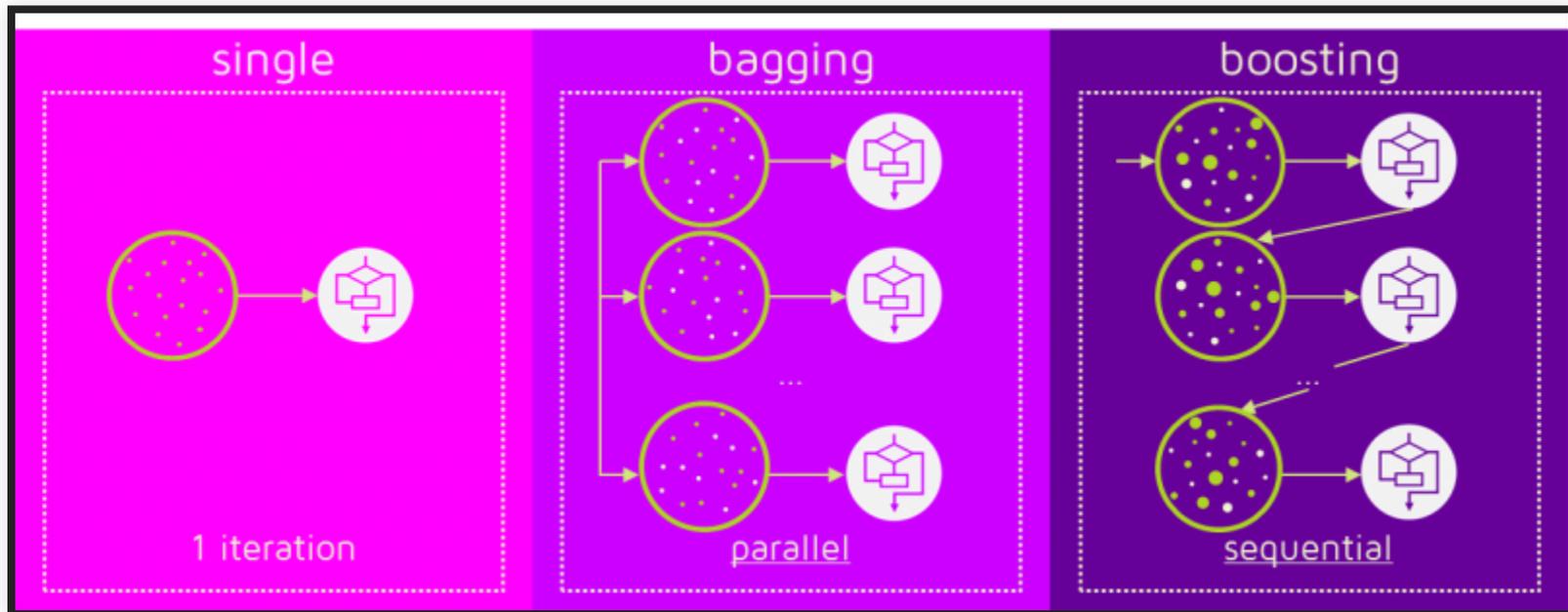
RANDOM FORESTS



BOOSTING

- Focar em exemplos mais "difíceis" de classificar
- Cria um modelo inicial, e marca os exemplos em que o modelo tem um desempenho ruim
- Iterativamente cria um novo novo modelo, atribuindo um custo maior ao exemplos incorretos na interação anterior
- Atribui um peso diferente e decrescente (para evitar overfitting) a cada modelo

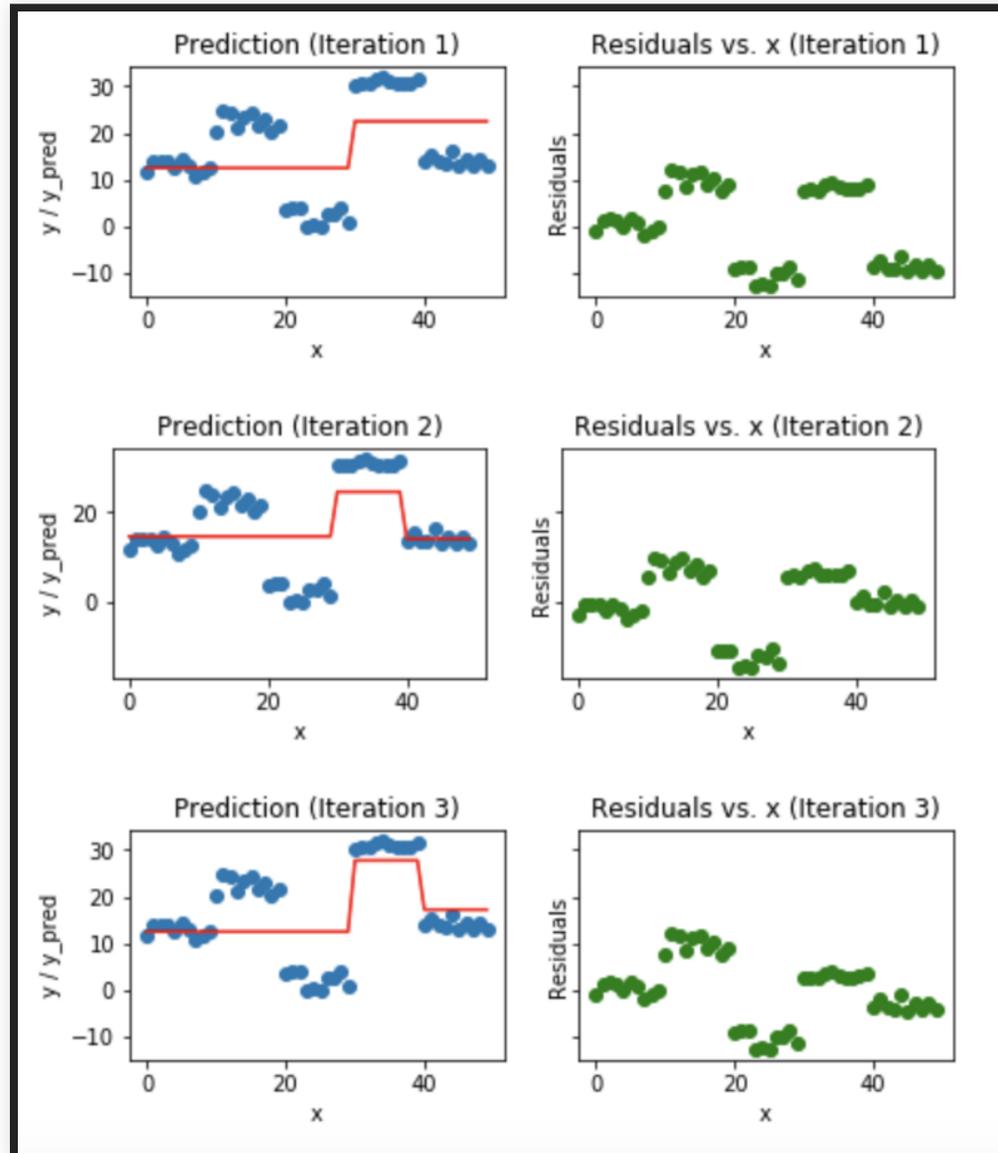
BOOSTING



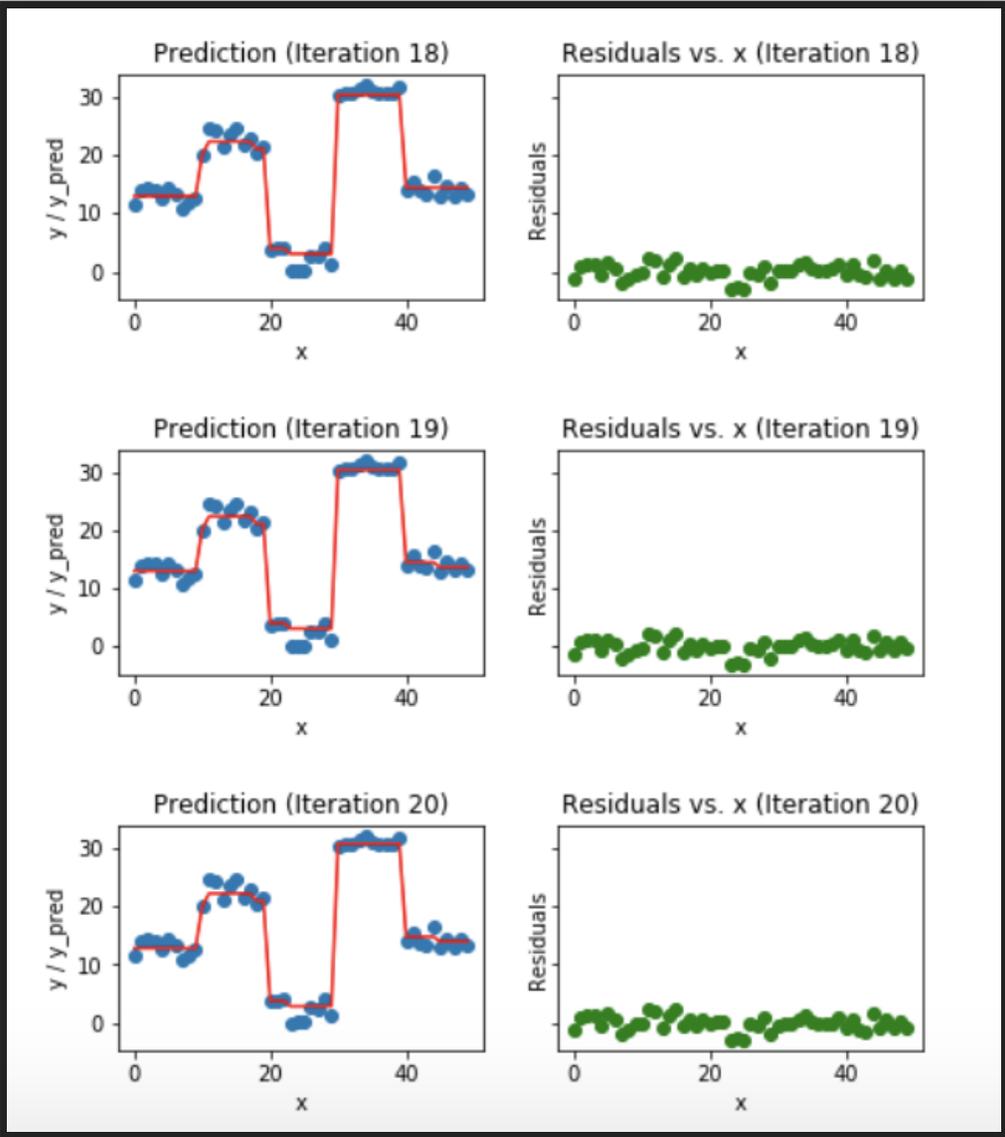
EXEMPLO: GRADIENT BOOSTING

1. Treinar um modelo simples
2. Calcular o residual (diferença entre valor predito e o real)
3. Cria um novo modelo usando o resíduo como atributo alvo
4. Adiciona o residual predito em 3. ao valor predito em 1
5. Repetir os passos 1 a 5 até convergir

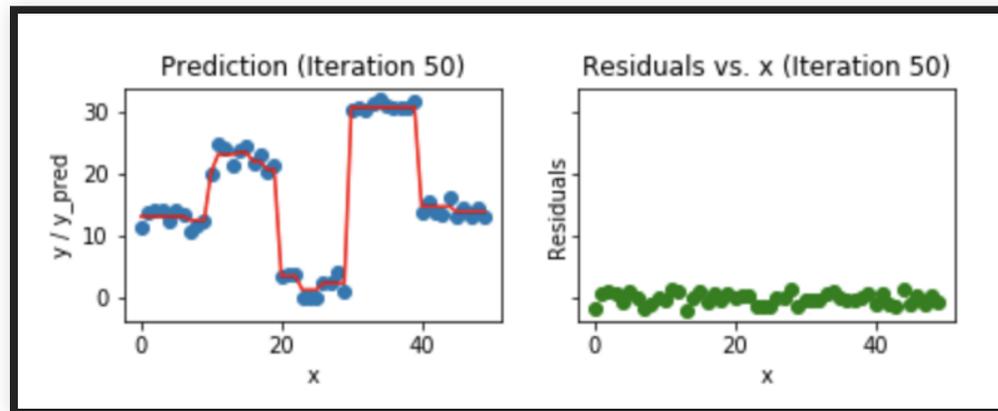
EXEMPLO: GRADIENT BOOSTING



EXEMPLO: GRADIENT BOOSTING



EXEMPLO: GRADIENT BOOSTING



BAGGING VERSUS BOOSTING

- A combinação de modelos do bagging pode reduzir overfitting. Ele atua principalmente na componente de variância do erro, e pode ser executado em paralelo.
- Já o boosting pode atuar tanto na componente de bias quando na de variância do erro. Entretanto, ele pode levar a um overfitting e é executado sequencialmente.

BAGGING VERSUS BOOSTING

