

Inteligência Artificial

BCC35G

Diego Bertolini

diegobertolini@utfpr.edu.br

<http://www.inf.ufpr.br/diegob/>

Aula 009

- **Aula Anterior:**
 - kNN ;
- **Aula de Hoje:**
 - **Aprendizagem Supervisionada – Decision Trees**

Objetivo

O que vocês devem saber ao final da aula:

Algoritmo ID3 e J48 – Decision Trees

Formas de Aprendizado

Aprendizado Supervisionado

- K-Nearest Neighbor (KNN). (ok)
- **Árvores de Decisão.**
- Support Vector Machines (SVM)
- Redes Neurais.

Aprendizado Não-Supervisionado

- K-means

Aprendizado Por Reforço

Árvores de Decisão

- Uma das formas de algoritmo de aprendizado mais simples e de maior sucesso.
- Uma árvore de decisão tem como entrada um objeto ou situação descritos por um conjunto de atributos e como saída uma “decisão” (previsão do valor de saída dada a entrada).
- Uma árvore de decisão toma as suas decisões através de uma sequência de testes.

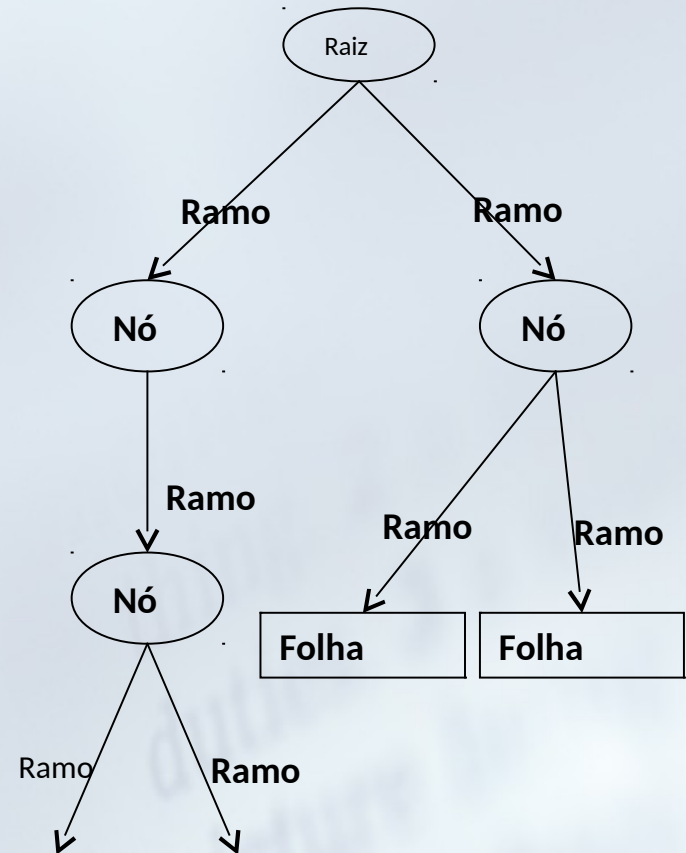
Árvores de Decisão

- Cada nó interno da árvore corresponde a um teste do valor de uma propriedade.

Os ramos dos nós são rotulados com os resultados possíveis do teste.

Cada nó folha da árvore especifica o valor a ser retornado se aquela folha for alcançada.

A representação de uma árvore de decisão é bem natural para os seres humanos.



Exemplo – Jogar Tênis

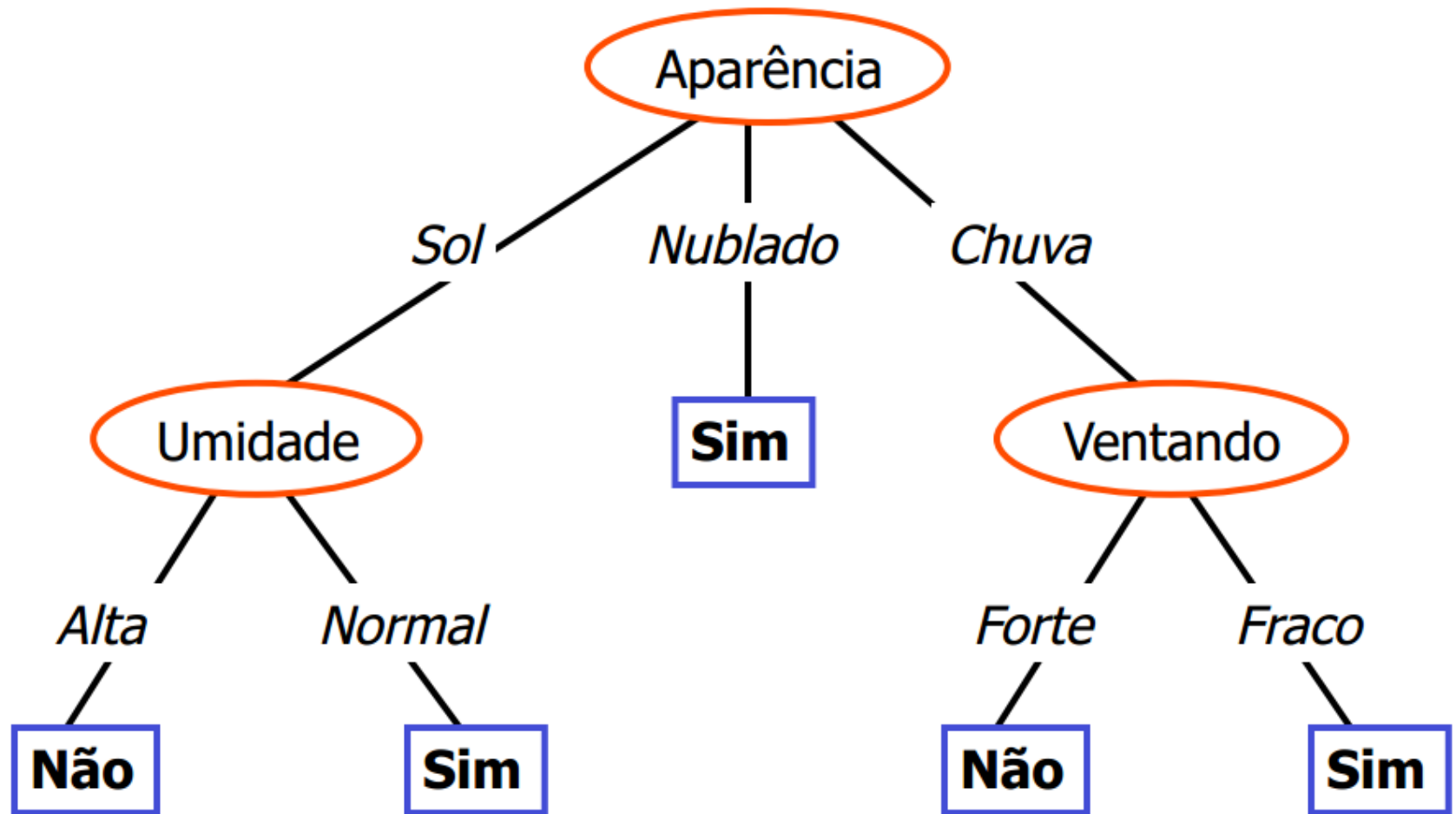
- Problema: Dada algumas condições, predizer se é possível jogar tênis ou não.
- O objetivo é aprender uma definição para o predicado “jogar tênis”.
- Primeiramente é necessário definir quais atributos estão disponíveis para descrever alguns exemplos nesse domínio.

Exemplo – Jogar Tênis

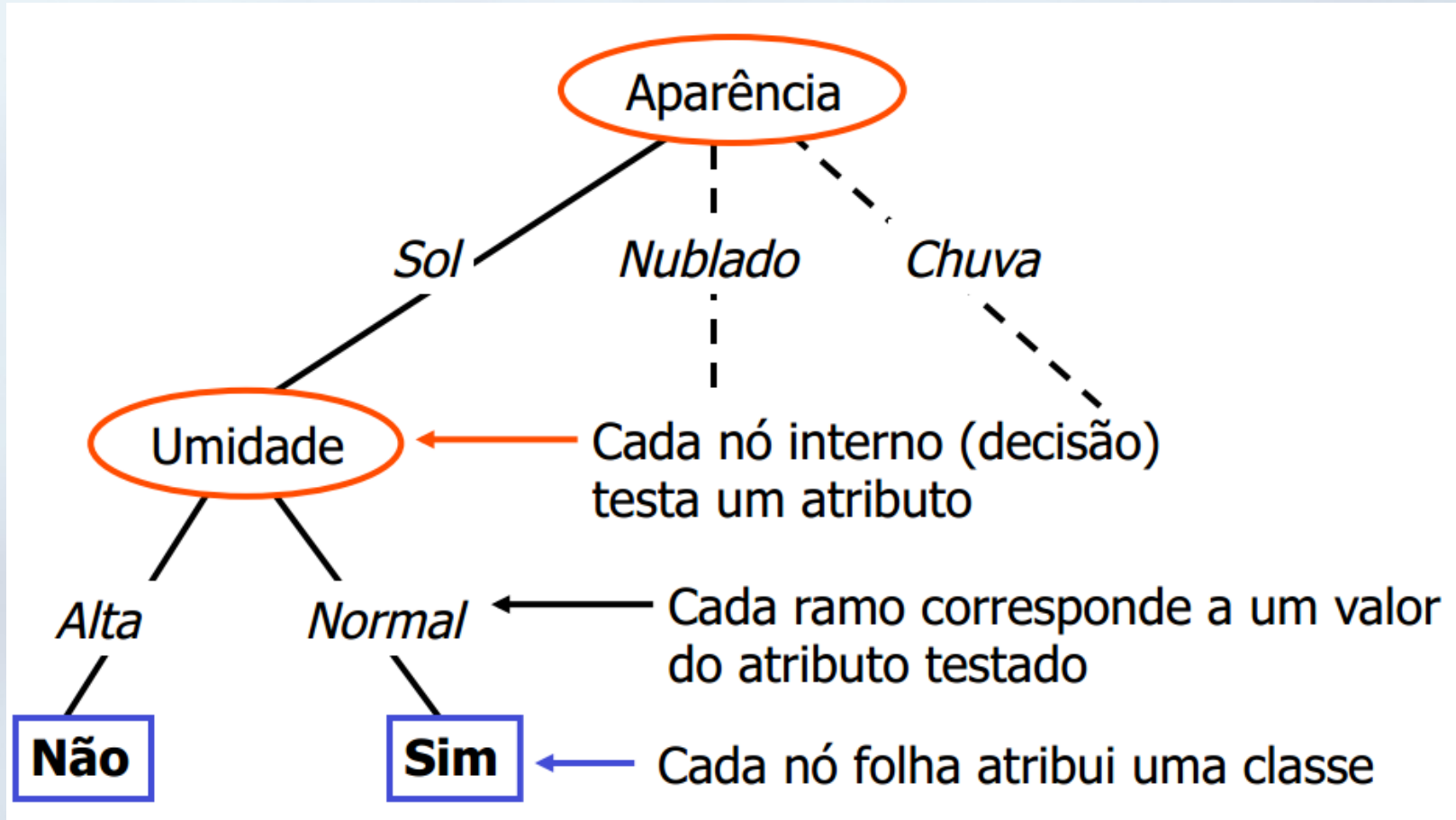
Atributos:

- **Aparência:** Sol, Nublado, Chuva ;
- **Umidade:** Alta, Normal ;
- **Vento:** Forte, Fraco ;
- **Temperatura:** Quente, Media, Fria ;
- **Classe (Conceito Alvo) jogar tênis:** Sim, NÃO ;

Exemplo – Jogar Tênis



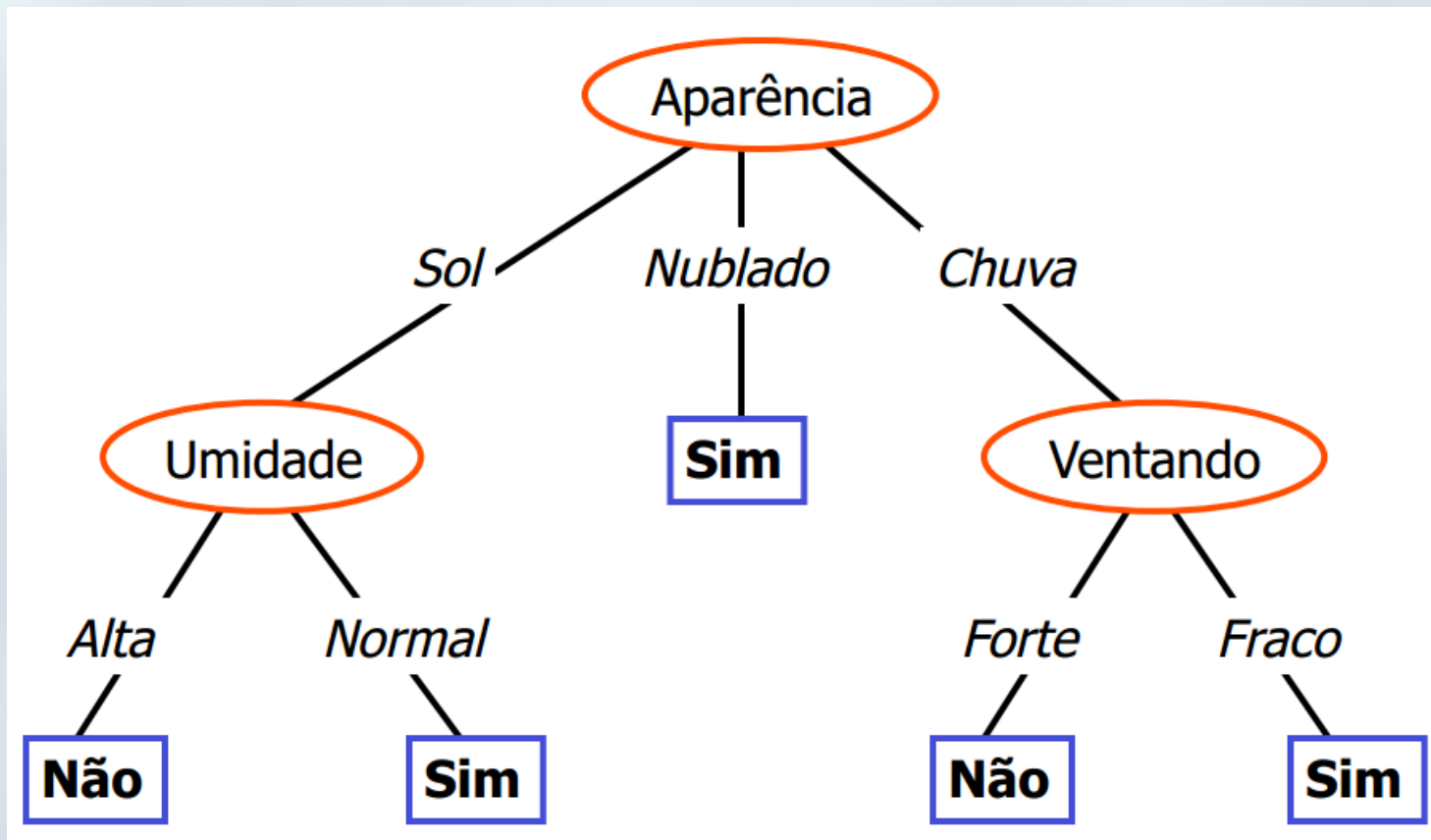
Exemplo – Jogar Tênis



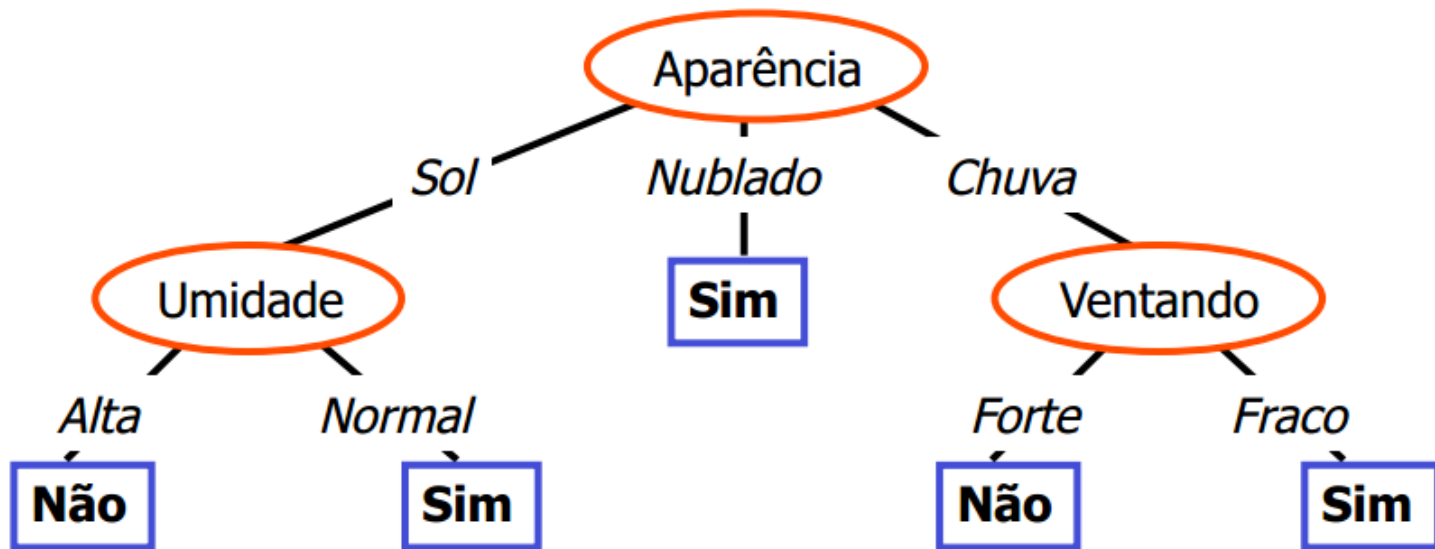
Exemplo – Jogar Tênis

Aparência: SOL, Umidade: ALTA, Ventando: FRACO

Jogar Tênis: ????



Representam Disjunções e Conjunções



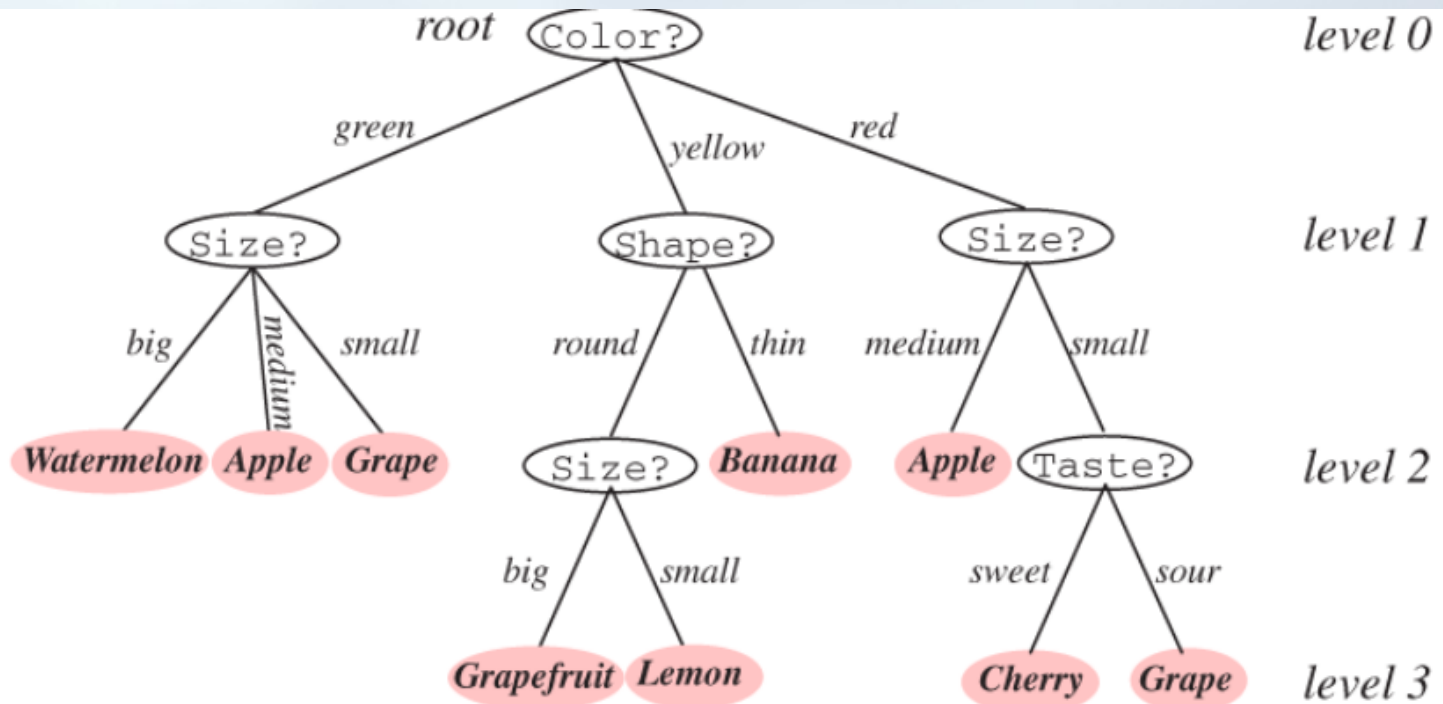
$(\text{Aparência}=\text{Sol} \wedge \text{Umidade}=\text{Normal}) \vee$
 $(\text{Aparência}=\text{Nublado}) \vee$
 $(\text{Aparência}=\text{Chuva} \wedge \text{Ventando}=\text{Fraco})$

Sim

$(\text{Aparência}=\text{Sol} \wedge \text{Umidade}=\text{Alta}) \vee$
 $(\text{Aparência}=\text{Chuva} \wedge \text{Ventando}=\text{Forte})$

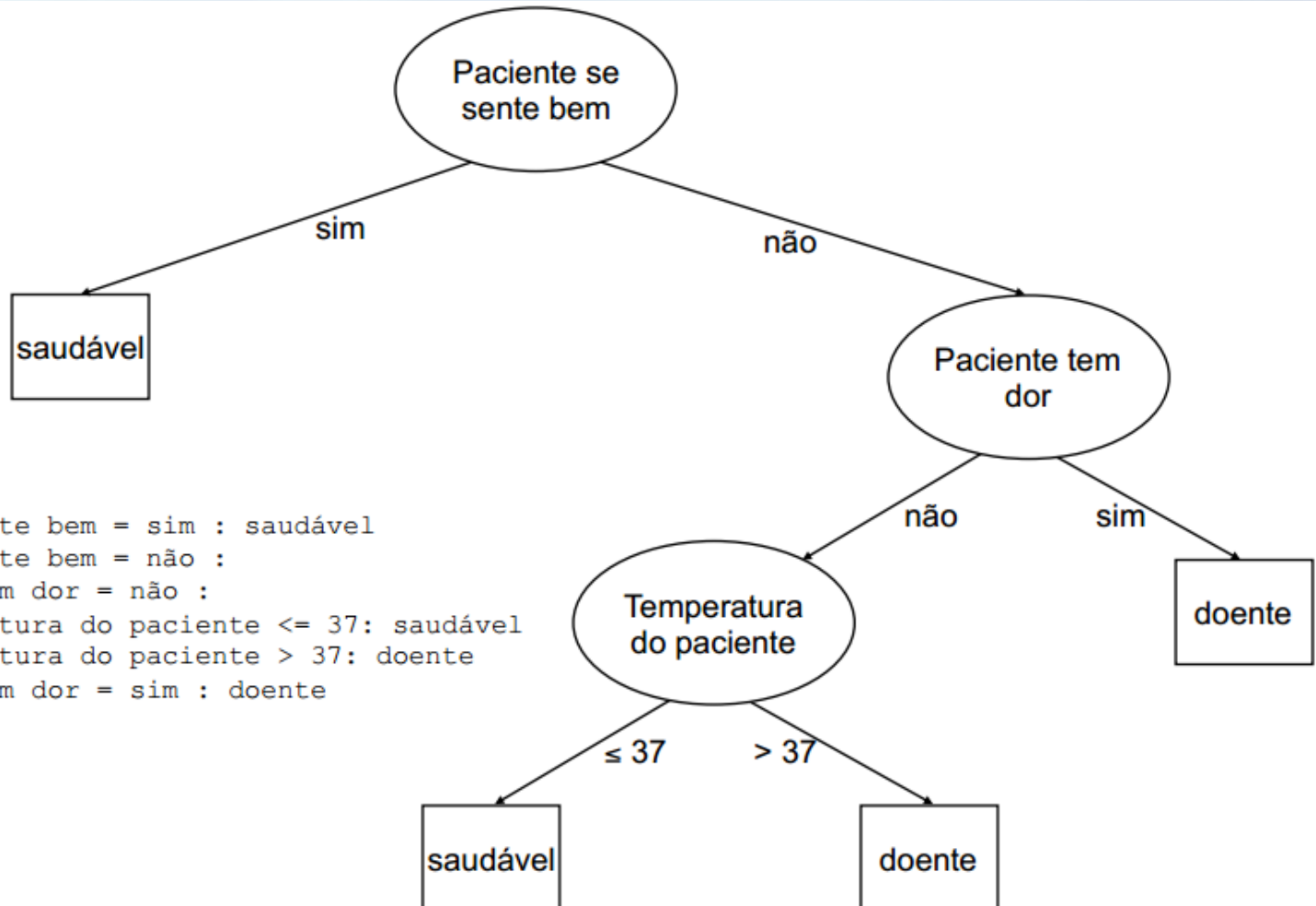
Não

Outro Exemplo – Frutas.



Classification in a basic decision tree proceeds from top to bottom. The questions asked at each node concern a particular property of the pattern, and the downward links correspond to the possible values. Successive nodes are visited until a terminal or leaf node is reached, where the category label is read. Note that the same question, **Size?**, appears in different places in the tree and that different questions can have different numbers of branches. Moreover, different leaf nodes, shown in pink, can be labeled by the same category (e.g., **Apple**).

Exemplo – Paciente Doente.



```
Paciente se sente bem = sim : saudável
Paciente se sente bem = não :
:...Paciente tem dor = não :
:....Temperatura do paciente ≤ 37: saudável
:   Temperatura do paciente > 37: doente
Paciente tem dor = sim : doente
```

Gerando Árvores de Decisão a partir de Exemplos

- Uma DT pode ser usada para classificar novos exemplos (nunca vistos)
- A partir da raiz basta descer através dos nós de decisão até encontrar um nó folha: a classe correspondente a esse nó folha é a classe do novo exemplo ;
- Um exemplo (sem valores desconhecidos) é classificado apenas por uma regra (sub-árvore) ;

Exemplo (adaptado de Quinlan, 93)

- *Neste exemplo, vamos considerar um conjunto de exemplos que contem medições diárias sobre condições meteorológicas;*
- *Atributos*
 - **Aparência:** ensolarado, nublado, chuvoso;
 - **Temperatura:** temperatura em graus Celsius
 - **Umidade:** umidade relativa do ar
 - **Ventando:** Sim ou não ;
- *Cada exemplo foi rotulado como “bom” se nas condições meteorológicas daquele dia são aconselhável fazer uma viagem a fazenda e “ruim” , caso contrario*

Exemplo (adaptado de Quinlan, 93)

Exemplo	Aparência	Temperatura	Umidade	Ventando	Viajar
E ₁	sol	25	72	sim	bom
E ₂	sol	28	91	sim	ruim
E ₃	sol	22	70	não	bom
E ₄	sol	23	95	não	ruim
E ₅	sol	30	85	não	ruim
E ₆	nublado	23	90	sim	bom
E ₇	nublado	29	78	não	bom
E ₈	nublado	19	65	sim	ruim
E ₉	nublado	26	75	não	bom
E ₁₀	nublado	20	87	sim	bom
E ₁₁	chuva	22	95	não	bom
E ₁₂	chuva	19	70	sim	ruim
E ₁₃	chuva	23	80	sim	ruim
E ₁₄	chuva	25	81	não	bom
E ₁₅	chuva	21	80	não	bom

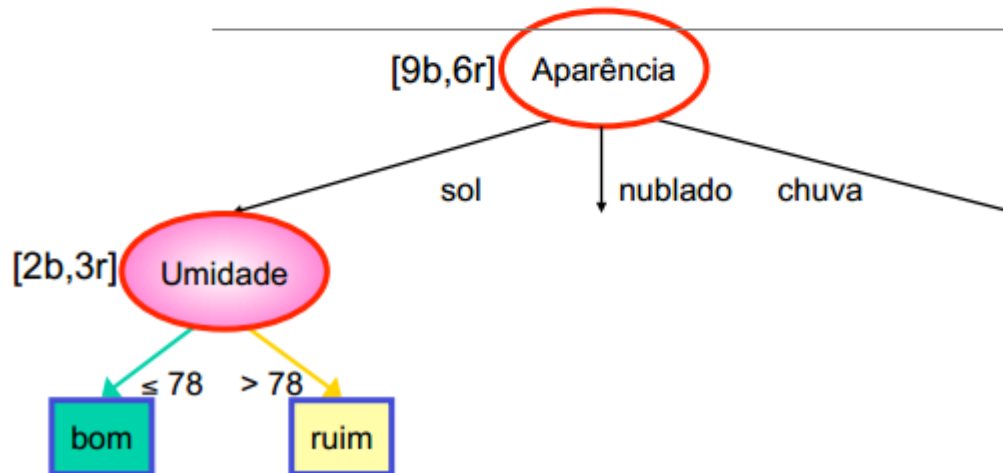
Exemplo (adaptado de Quinlan, 93)

Exemplo	Aparência	Temperatura	Umidade	Ventando	Viajar
E ₁	sol	25	72	sim	bom
E ₂	sol	28	91	sim	ruim
E ₃	sol	22	70	não	bom
E ₄	sol	23	95	não	ruim
E ₅	sol	30	85	não	ruim
E ₆	nublado	23	90	sim	bom
E ₇	nublado	29	78	não	bom
E ₈	nublado	19	65	sim	ruim
E ₉	nublado	26	75	não	bom
E ₁₀	nublado	20	87	sim	bom
E ₁₁	chuva	22	95	não	bom
E ₁₂	chuva	19	70	sim	ruim
E ₁₃	chuva	23	80	sim	ruim
E ₁₄	chuva	25	81	não	bom
E ₁₅	chuva	21	80	não	bom



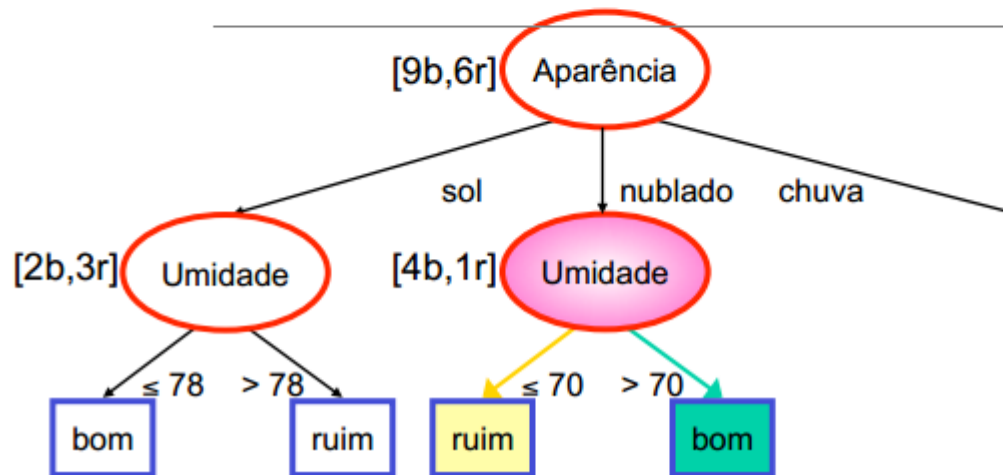
Exemplo (adaptado de Quinlan, 93)

Exemplo	Aparência	Temperatura	Umidade	Ventando	Viajar
E ₁	sol	25	72	sim	bom
E ₂	sol	28	91	sim	ruim
E ₃	sol	22	70	não	bom
E ₄	sol	23	95	não	ruim
E ₅	sol	30	85	não	ruim
E ₆	nublado	23	90	sim	bom
E ₇	nublado	29	78	não	bom
E ₈	nublado	19	65	sim	ruim
E ₉	nublado	26	75	não	bom
E ₁₀	nublado	20	87	sim	bom
E ₁₁	chuva	22	95	não	bom
E ₁₂	chuva	19	70	sim	ruim
E ₁₃	chuva	23	80	sim	ruim
E ₁₄	chuva	25	81	não	bom
E ₁₅	chuva	21	80	não	bom



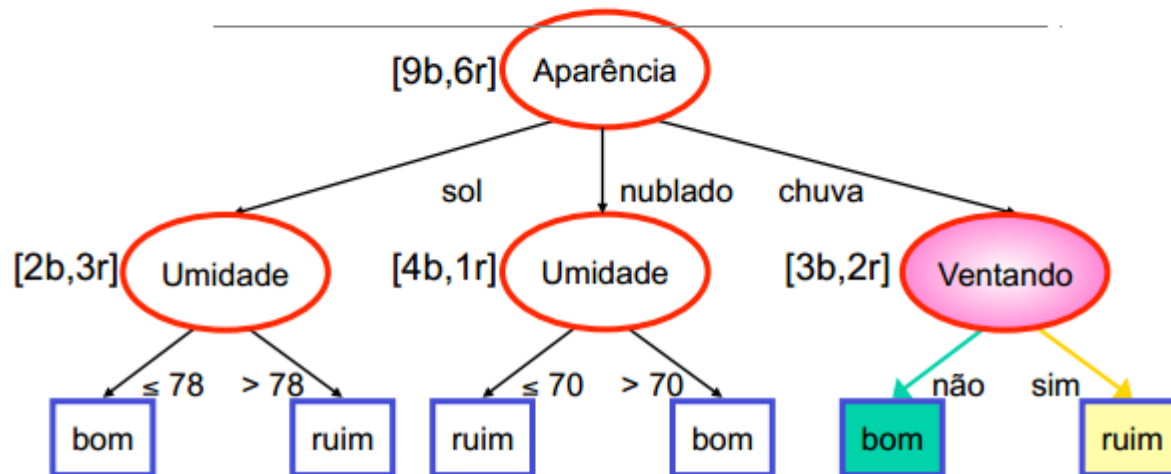
Exemplo (adaptado de Quinlan, 93)

Exemplo	Aparência	Temperatura	Umidade	Ventando	Viajar
E ₁	sol	25	72	sim	bom
E ₂	sol	28	91	sim	ruim
E ₃	sol	22	70	não	bom
E ₄	sol	23	95	não	ruim
E ₅	sol	30	85	não	ruim
E ₆	nublado	23	90	sim	bom
E ₇	nublado	29	78	não	bom
E ₈	nublado	19	65	sim	ruim
E ₉	nublado	26	75	não	bom
E ₁₀	nublado	20	87	sim	bom
E ₁₁	chuva	22	95	não	bom
E ₁₂	chuva	19	70	sim	ruim
E ₁₃	chuva	23	80	sim	ruim
E ₁₄	chuva	25	81	não	bom
E ₁₅	chuva	21	80	não	bom



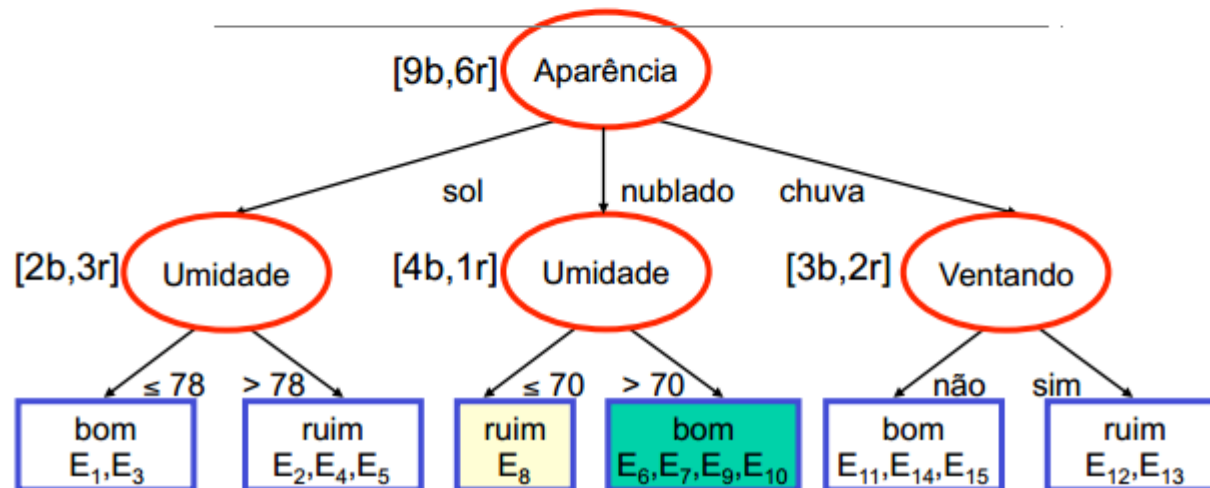
Exemplo (adaptado de Quinlan, 93)

Exemplo	Aparência	Temperatura	Umidade	Ventando	Viajar
E ₁	sol	25	72	sim	bom
E ₂	sol	28	91	sim	ruim
E ₃	sol	22	70	não	bom
E ₄	sol	23	95	não	ruim
E ₅	sol	30	85	não	ruim
E ₆	nublado	23	90	sim	bom
E ₇	nublado	29	78	não	bom
E ₈	nublado	19	65	sim	ruim
E ₉	nublado	26	75	não	bom
E ₁₀	nublado	20	87	sim	bom
E ₁₁	chuva	22	95	não	bom
E ₁₂	chuva	19	70	sim	ruim
E ₁₃	chuva	23	80	sim	ruim
E ₁₄	chuva	25	81	não	bom
E ₁₅	chuva	21	80	não	bom



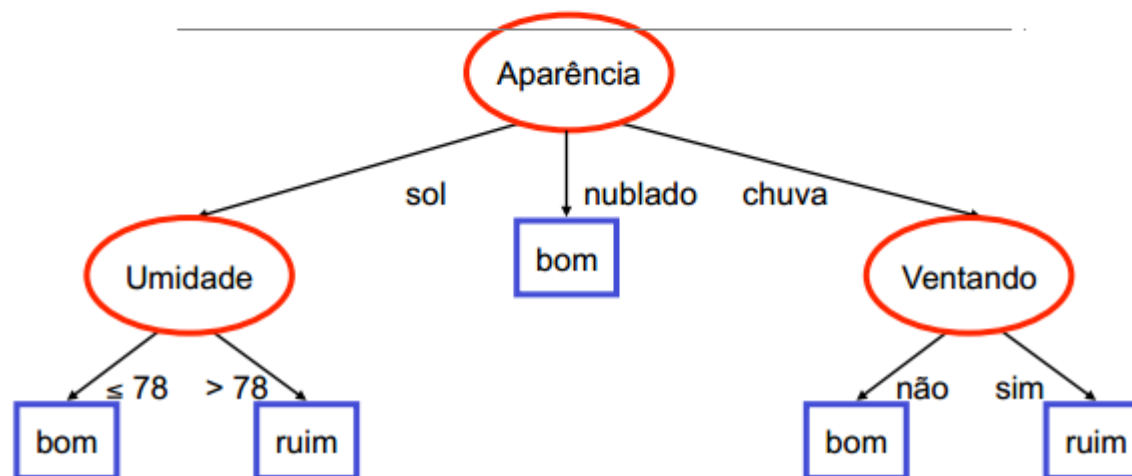
Arvore de Decisão Sem Poda

Exemplo	Aparência	Temperatura	Umidade	Ventando	Viajar
E ₁	sol	25	72	sim	bom
E ₂	sol	28	91	sim	ruim
E ₃	sol	22	70	não	bom
E ₄	sol	23	95	não	ruim
E ₅	sol	30	85	não	ruim
E ₆	nublado	23	90	sim	bom
E ₇	nublado	29	78	não	bom
E ₈	nublado	19	65	sim	ruim
E ₉	nublado	26	75	não	bom
E ₁₀	nublado	20	87	sim	bom
E ₁₁	chuva	22	95	não	bom
E ₁₂	chuva	19	70	sim	ruim
E ₁₃	chuva	23	80	sim	ruim
E ₁₄	chuva	25	81	não	bom
E ₁₅	chuva	21	80	não	bom



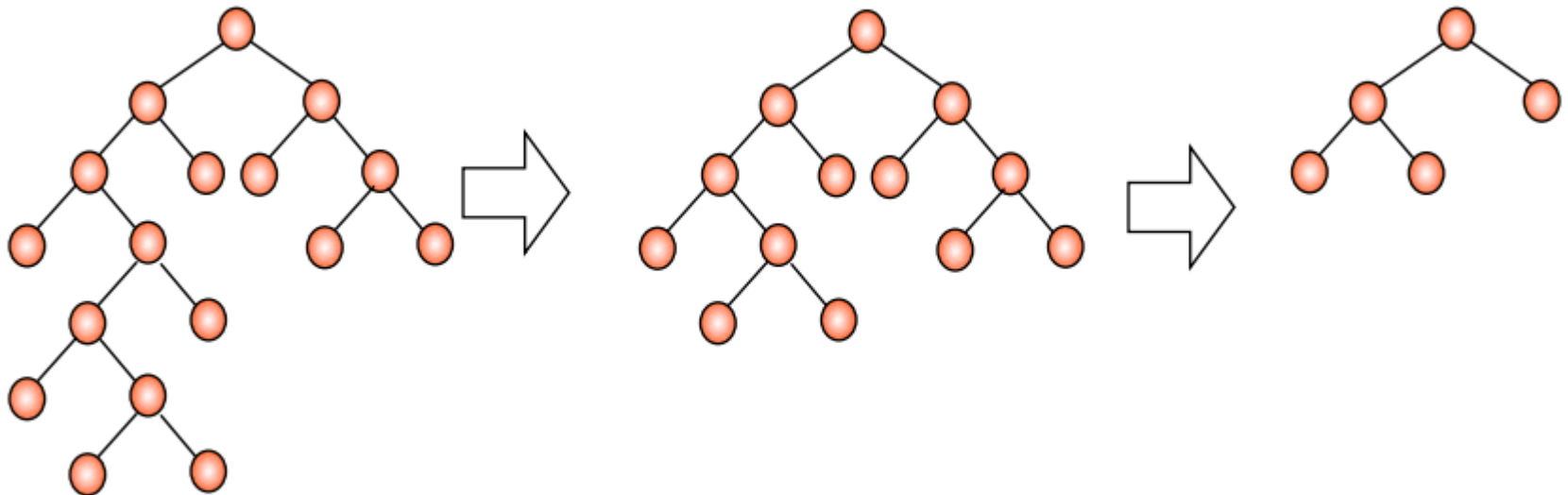
Arvore Podada

Exemplo	Aparência	Temperatura	Umidade	Ventando	Viajar
E ₁	sol	25	72	sim	bom
E ₂	sol	28	91	sim	ruim
E ₃	sol	22	70	não	bom
E ₄	sol	23	95	não	ruim
E ₅	sol	30	85	não	ruim
E ₆	nublado	23	90	sim	bom
E ₇	nublado	29	78	não	bom
E ₈	nublado	19	65	sim	ruim
E ₉	nublado	26	75	não	bom
E ₁₀	nublado	20	87	sim	bom
E ₁₁	chuva	22	95	não	bom
E ₁₂	chuva	19	70	sim	ruim
E ₁₃	chuva	23	80	sim	ruim
E ₁₄	chuva	25	81	não	bom
E ₁₅	chuva	21	80	não	bom



Poda

- ❑ Uma árvore maior é induzida de forma a superajustar os exemplos e então ela é podada até obter uma árvore menor (mais simples)
- ❑ A poda evita *overfitting*



Gerando Árvores de Decisão a partir de Exemplos

- **Seguindo o principio de Ockham, devemos encontrar a menor árvore de decisão que seja consistente com os exemplos de treinamento.**
 - “Qualquer fenômeno deve assumir apenas as premissas estritamente necessárias à explicação do fenômeno e eliminar todas as que não causariam qualquer diferença aparente nas predições da hipótese ou teoria.”
- **A ideia básica do algoritmo é testar os atributos mais importantes primeiro.**
 - **O atributo mais importante é aquele que faz mais diferença para a classificação de um exemplo.**
- Dessa forma, esperamos conseguir a classificação correta com um pequeno número de testes.

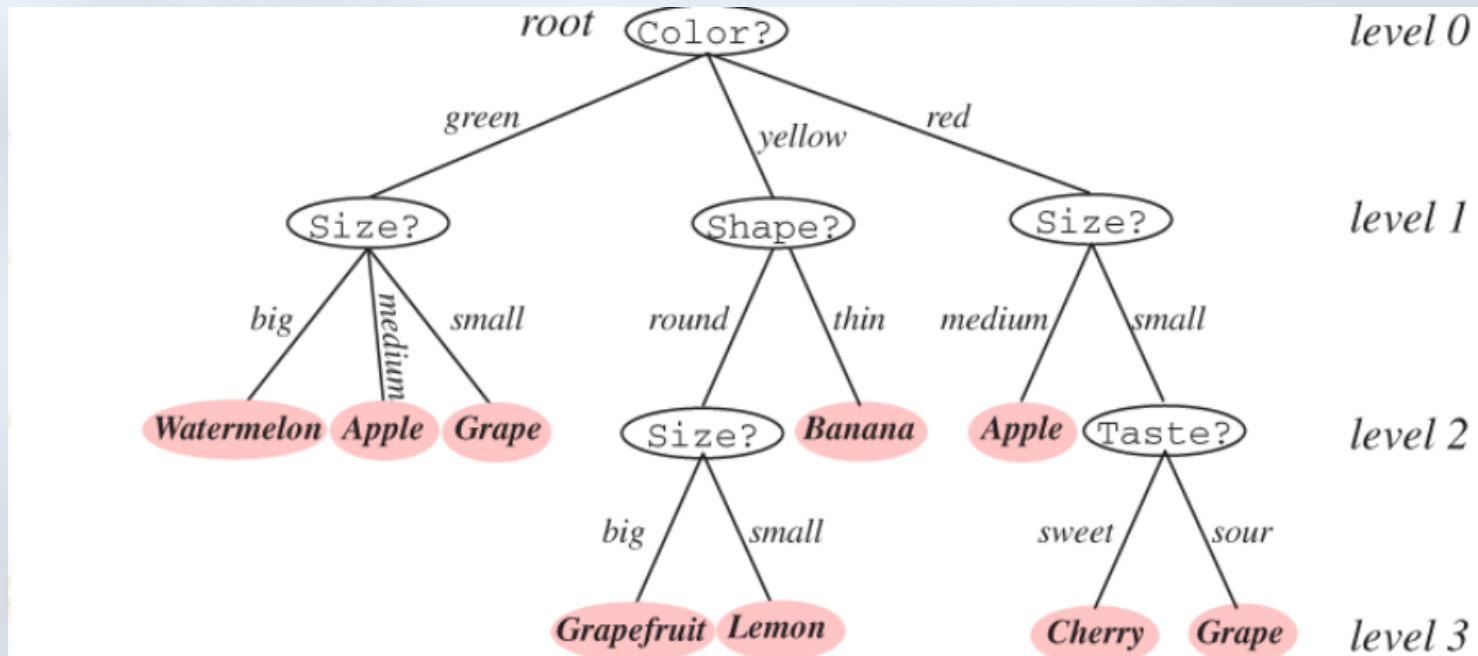
Gerando Árvores de Decisão a partir de Exemplos

Algoritmo:

- (1) Enquanto existirem exemplos positivos e negativos, deve-se escolher o melhor atributo para dividi-los.
- (2) Se todos os exemplos restantes forem positivos (ou todos negativos), então podemos responder Sim ou Não.
- (3) Se não existirem exemplos restantes, retorna um valor padrão calculado a partir da classificação da maioria dos atributos do nó pai.
- (4) Se não existirem atributo restantes, mas ainda existirem exemplos positivos e negativos temos um problema.

Gerando Árvores de Decisão a partir de Exemplos

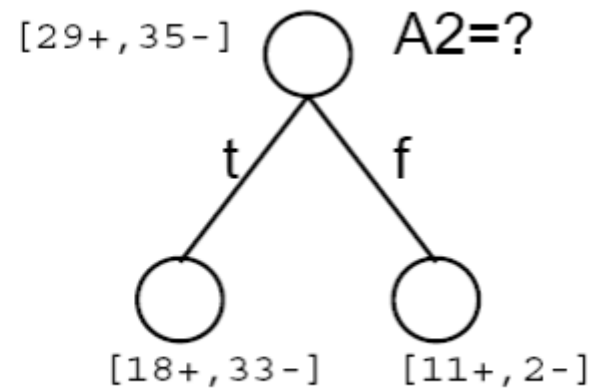
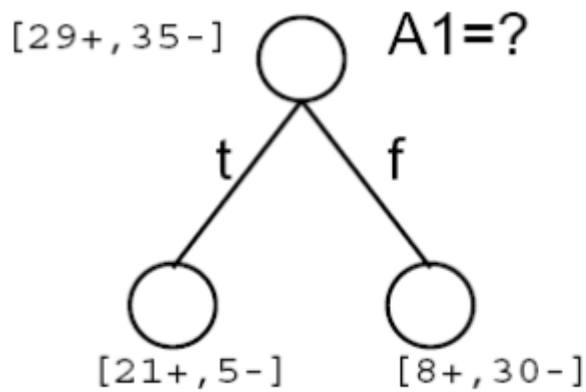
Qual o melhor atributo?



Classification in a basic decision tree proceeds from top to bottom. The questions asked at each node concern a particular property of the pattern, and the downward links correspond to the possible values. Successive nodes are visited until a terminal or leaf node is reached, where the category label is read. Note that the same question, *Size?*, appears in different places in the tree and that different questions can have different numbers of branches. Moreover, different leaf nodes, shown in pink, can be labeled by the same category (e.g., **Apple**).

Escolhendo os Melhores Atributos

Qual é o melhor atributo?



Gerando Árvores de Decisão a partir de Exemplos

- **Como medir a habilidade de um dado atributo na tarefa de discriminar as classes?**
- **Existem muitas medidas. Todas concordam em dois pontos:**
 - Uma divisão que mantém as proporções de classes em todas as partições é inútil;
 - Uma divisão na qual em cada partição todos os exemplos são da mesma classe tem utilidade máxima

Escolhendo os Melhores Atributos

Entropia

- Caracteriza a (im)pureza de uma coleção arbitrária de exemplos.
- Dado uma coleção **S** contendo exemplos positivos (+) e negativos (−) de algum conceito alvo, a entropia de **S** relativa a esta classificação booleana é:

$$\text{Entropia}(S) = -p_+ \log_2 p_+ - p_- \log_2 p_-$$

- p_+ é a proporção de exemplos positivos em S.
- p_- é a proporção de exemplos negativos em S.

Escolhendo os Melhores Atributos

Exemplo: Sendo **S** uma coleção de 14 exemplos de treinamento de algum conceito booleano, incluindo 9 exemplos positivos e 5 negativos [9+, 5-].

A entropia de S relativa a classificação é:

$$\text{Entropia}([9+, 5-]) = -\left(\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14}\right) = 0.940$$

A função entropia relativa a uma classificação varia entre 0 e 1.

Escolhendo os Melhores Atributos

- Generalizando para o caso de um atributo alvo aceitar **n** diferentes valores, a entropia de S relativa a esta classificação de n -classes é definida como:

$$\text{Entropia}(S) = \sum_{i=1}^n -p_i \log_2 p_i$$

Medindo Desempenho

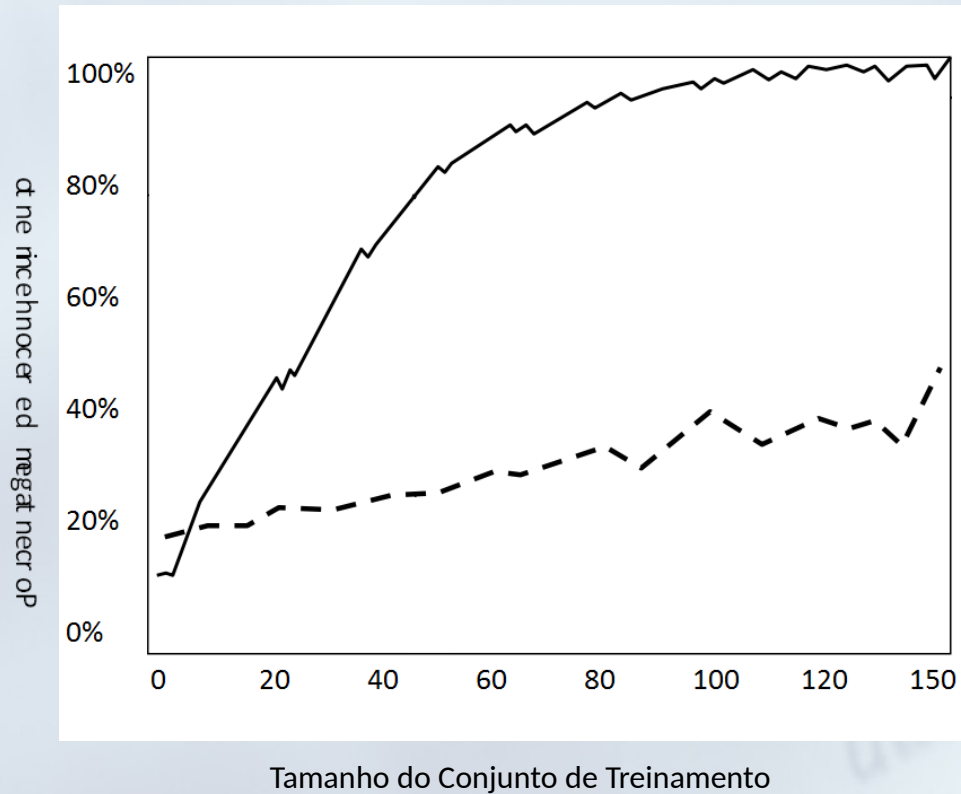
- Um algoritmo de aprendizado é bom se ele produz hipóteses que conseguem prever a classificação de exemplos não vistos.
- A maneira mais simples de medir o desempenho de um método de aprendizado é realizando a classificação de um conjunto de exemplos de teste.

Medindo Desempenho

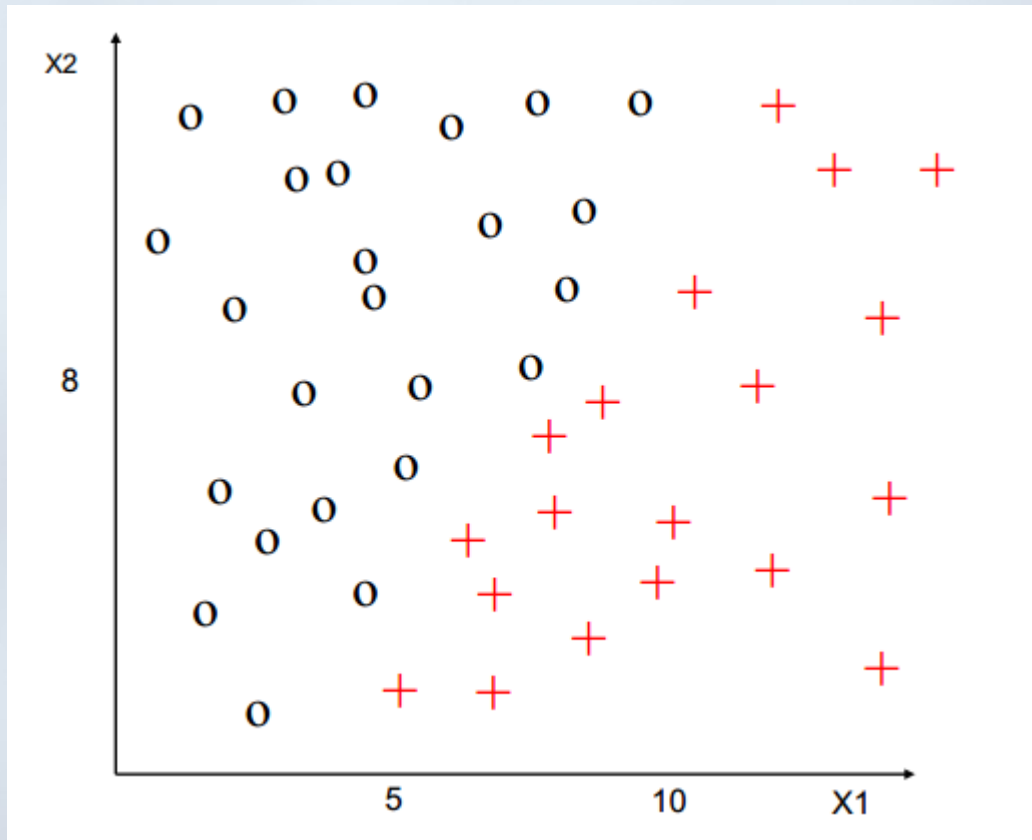
Processo de avaliação:

- (1) Divide-se o conjunto total de exemplos conhecidos em dois conjuntos:
Conjunto de Treinamento.
Conjunto de Teste.
- (2) Gera-se uma hipótese h (árvore de decisão) com base no Conjunto de Treinamento.
- (3) Para cada exemplo do Conjunto de Teste, classifica-se o exemplo utilizando a árvore de decisão criada a partir do conjunto de treinamento.
- (4) Verifica-se a quantidade de exemplos de teste classificados corretamente e calcula-se a porcentagem de acertos.
- (5) Escolhe-se aleatoriamente um novo conjunto de exemplos de treinamento (normalmente com um numero maior de exemplos) e repete-se novamente o processo.

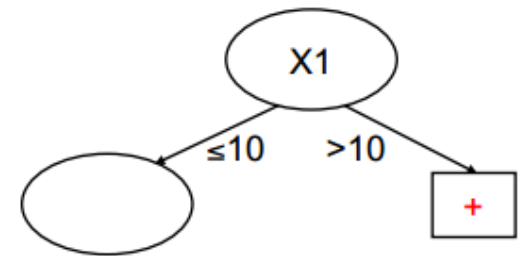
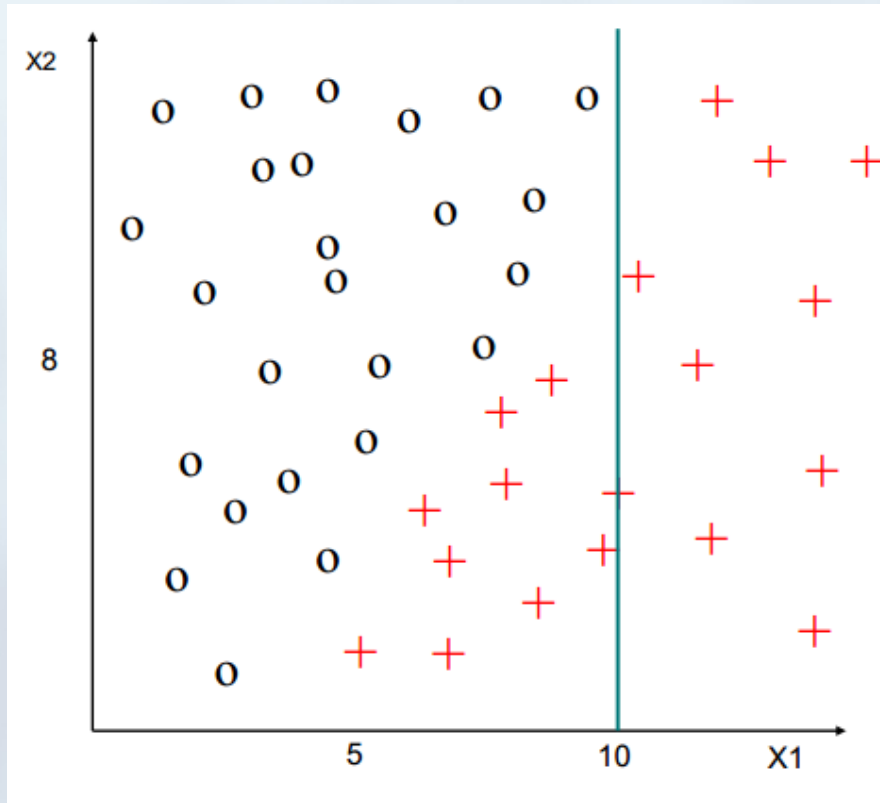
Medindo Desempenho



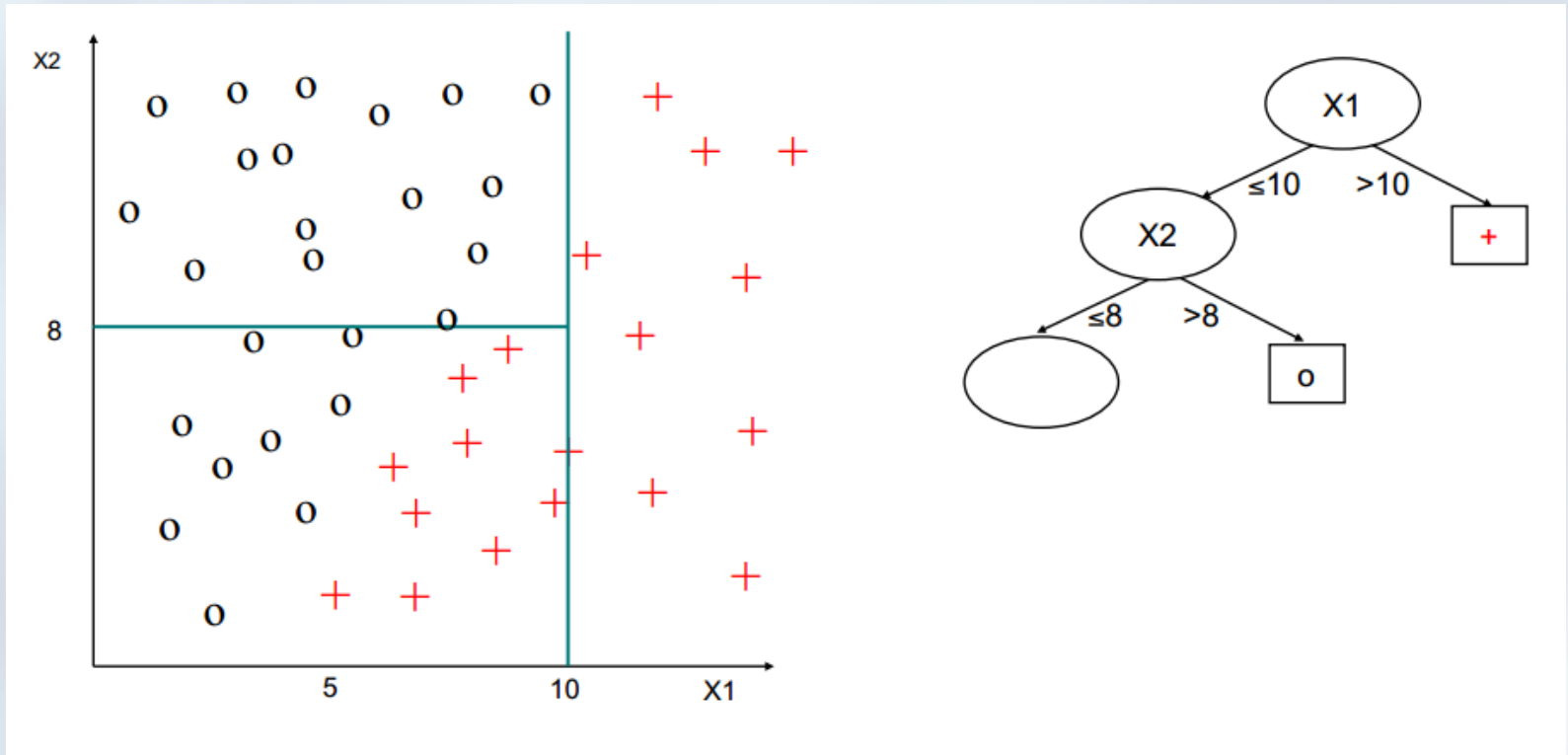
Avaliação Geométrica para DT



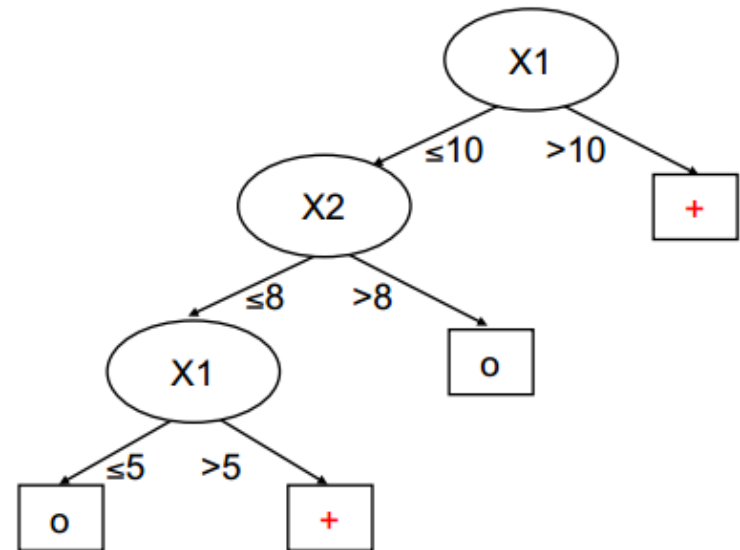
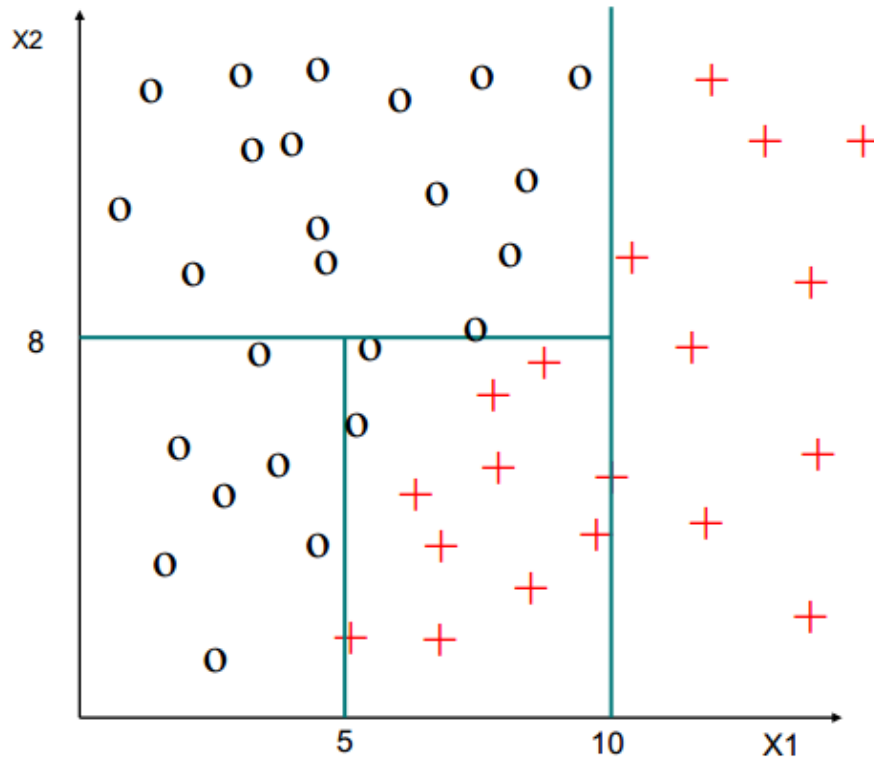
Avaliação Geométrica para DT



Avaliação Geométrica para DT



Avaliação Geométrica para DT



Resumindo...

- Árvores de Decisão em geral, possuem um tempo de aprendizado relativamente rápido ;
- Árvores de decisão permitem a classificação de conjuntos com milhões de exemplos e centenas de atributos a uma velocidade razoável ;
- Possível converter para regras de classificação, podendo ser interpretadas por seres humanos ;
- Precisão comparável a outros métodos.

Atividades

- Baixar os arquivos da base de *spam* (***conj. Treinamento e conjunto Teste***);
- Verificar o desempenho para Árvores de Decisão ;
- Utilizar e verificar o Parametro: Poda ;

Bibliografia e Materiais.

Estes slides foram adaptados do Livro:

Artificial Intelligence; Chapter 18: Learning from Examples. (18.3 Learning Decision Trees)

Adaptado das Aulas do Professor Ederley – PUC-RIO

Adaptado das Aulas do Professor Alessandro – PUC-PR

Adaptado das Aulas do George Cavalcanti – UFPE

Adaptado das Aulas do Jose Augusto – USP