

# INF 1771 – Inteligência Artificial

## Aula 13 – Árvores de Decisão

Edirlei Soares de Lima  
<elima@inf.puc-rio.br>



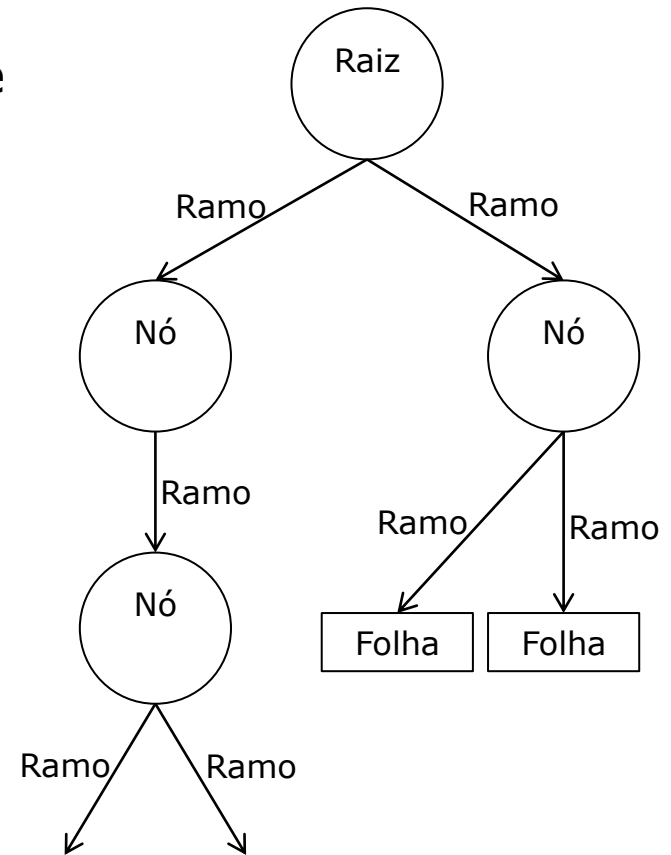
# Árvores de Decisão

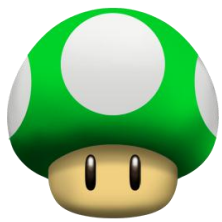
- ❏ Uma das formas de algoritmo de aprendizado mais **simples** e de **maior sucesso**.
- ❏ Uma árvore de decisão tem como entrada um objeto ou situação descritos por um **conjunto de atributos** e como saída uma “**decisão**” (previsão do valor de saída dada a entrada).
- ❏ Uma árvore de decisão toma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



# Árvores de Decisão

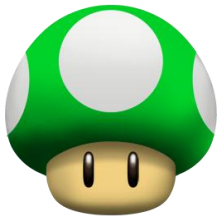
- ❏ Cada **nó** interno da árvore corresponde a um teste do valor de uma propriedade.
- ❏ Os **ramos** dos nós são rotulados com os resultados possíveis do teste.
- ❏ Cada **nó folha** da árvore especifica o valor a ser retornado se aquela folha for alcançada.
- ❏ A representação de uma árvore de decisão é bem natural para os seres humanos.





# Exemplo – Restaurante

- ❗ **Problema:** Esperar por uma mesa em um restaurante.
- ❗ O **objetivo** é aprender uma definição para o predicado “vai esperar”.
- ❗ Primeiramente é necessário definir quais **atributos** estão disponíveis para descrever alguns exemplos nesse domínio.



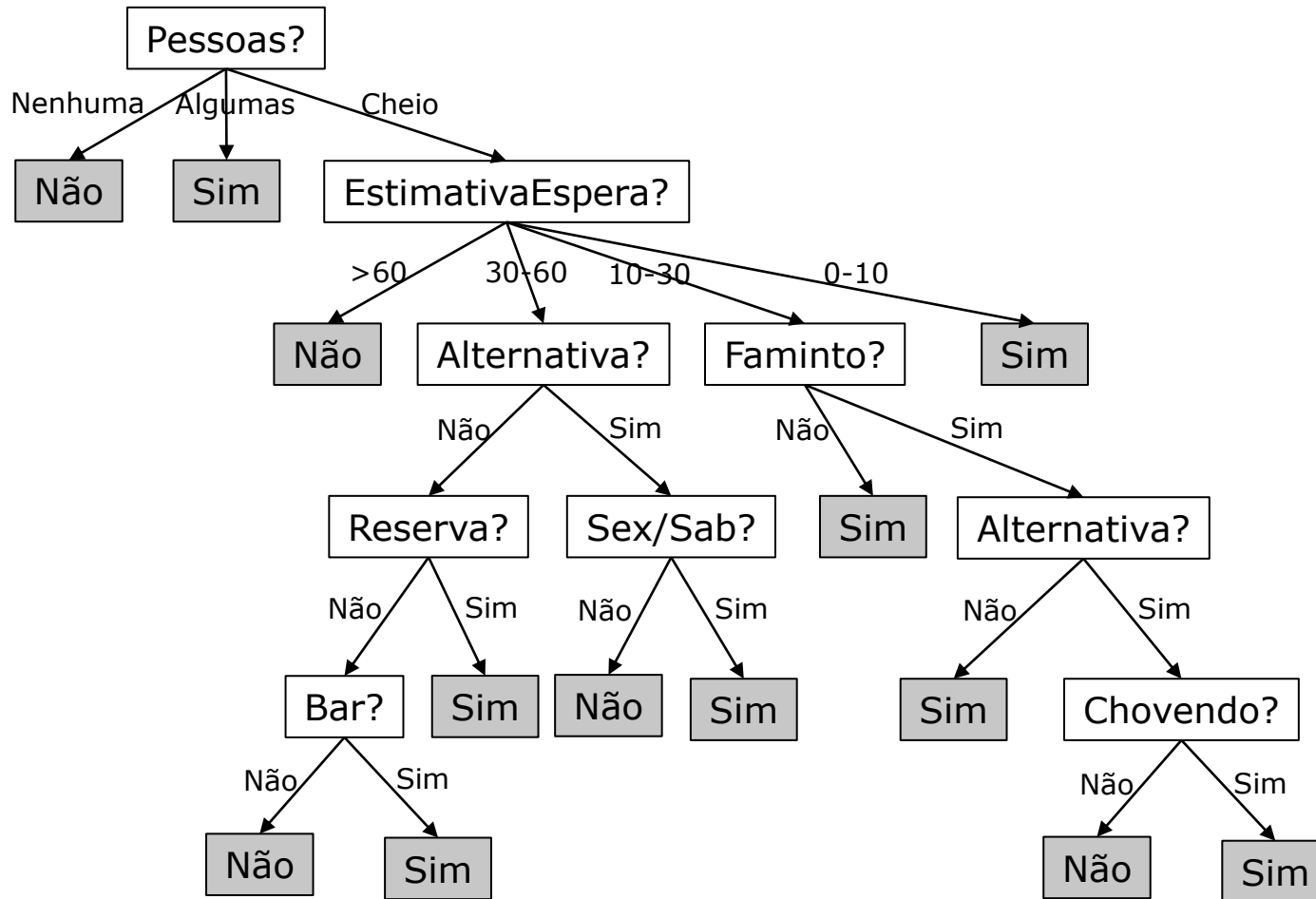
# Exemplo – Restaurante

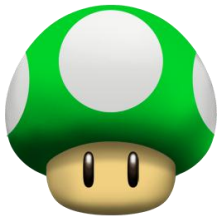
## 📌 Atributos:

- 📌 **Alternativa:** Verdadeiro se existe um restaurante alternativo adequado nas proximidades.
- 📌 **Bar:** Verdadeiro se o restaurante tem uma área de bar confortável para ficar esperando.
- 📌 **Sex/Sab:** Verdadeiro se o dia da semana for sexta ou sábado.
- 📌 **Faminto:** Verdadeiro se estamos com fome.
- 📌 **Pessoas:** Quantas pessoas estão no restaurante (os valores são Nenhuma, Algumas e Cheio).
- 📌 **Preço:** Preço do restaurante de (\$, \$ \$, \$\$\$).
- 📌 **Chuva:** Verdadeiro se está chovendo lá fora.
- 📌 **Reserva:** Verdadeiro se nós fizemos uma reserva.
- 📌 **Tipo:** Tipo de restaurante (Francês, Italiano, Tailandês, Hambúrguer).
- 📌 **EstimativaEspera:** Tempo de espera estimado (00-10, 10-30, 30-60, > 60 minutos).



# Exemplo - Restaurante





## Gerando Árvores de Decisão a partir de Exemplos

- ❏ É possível gerar uma árvore de decisão a partir de um **conjunto de exemplos**.
- ❏ **Exemplos positivos** são aqueles que levam a uma resposta positiva.  
Exemplo: “vai esperar” = Sim.
- ❏ **Exemplos negativos** são aqueles que levam a uma resposta negativa.  
Exemplo: “vai esperar” = Não.



# Conjunto de Treinamento

	Atributos										Obj.
Exemplo	Alt.	Bar	S/S	Fam.	Pes.	Pre.	Chov.	Res.	Tipo	Est.	Esp.
$X_1$	Sim	Não	Não	Sim	Algumas	\$\$\$	Não	Sim	Fran.	0-10	Sim
$X_2$	Sim	Não	Não	Sim	Cheio	\$	Não	Não	Tai.	30-60	Não
$X_3$	Não	Sim	Não	Não	Algumas	\$	Não	Não	Ham.	0-10	Sim
$X_4$	Sim	Não	Sim	Sim	Cheio	\$	Sim	Não	Tai.	10-30	Sim
$X_5$	Sim	Não	Sim	Não	Cheio	\$\$\$	Não	Sim	Fran.	>60	Não
$X_6$	Não	Sim	Não	Sim	Algumas	\$\$	Sim	Sim	Ital.	0-10	Sim
$X_7$	Não	Sim	Não	Não	Nenhuma	\$	Sim	Não	Ham.	0-10	Não
$X_8$	Não	Não	Não	Sim	Algumas	\$\$	Sim	Sim	Tai.	0-10	Sim
$X_9$	Não	Sim	Sim	Não	Cheio	\$	Sim	Não	Ham.	>60	Não
$X_{10}$	Sim	Sim	Sim	Sim	Cheio	\$\$\$	Não	Sim	Ital.	10-30	Não
$X_{11}$	Não	Não	Não	Não	Nenhuma	\$	Não	Não	Tai.	0-10	Não
$X_{12}$	Sim	Sim	Sim	Sim	Cheio	\$	Não	Não	Ham.	30-60	Sim





# Gerando Árvores de Decisão a partir de Exemplos

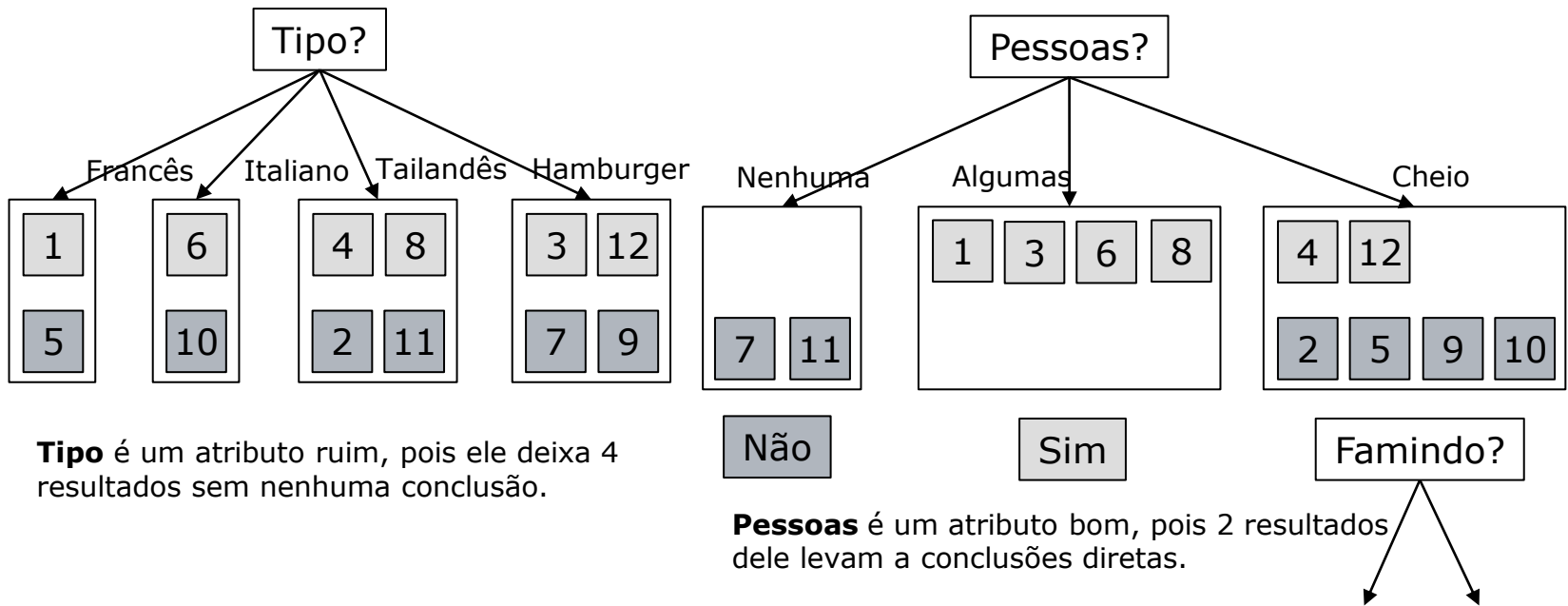
- ❏ Seguindo o **princípio de Ockham**, devemos encontrar a menor árvore de decisão que seja consistente com os exemplos de treinamento.
  - ❏ “Qualquer fenómeno deve assumir apenas as premissas estritamente necessárias à explicação do fenómeno e eliminar todas as que não causariam qualquer diferença aparente nas predições da hipótese ou teoria.”
- ❏ A idéia 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

Conjunto de Treinamento

1	3	4	6	8	12
2	5	7	9	10	11

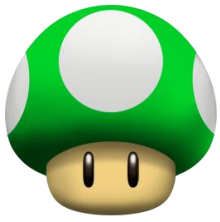




# Gerando Árvores de Decisão a partir de Exemplos

## 💡 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.

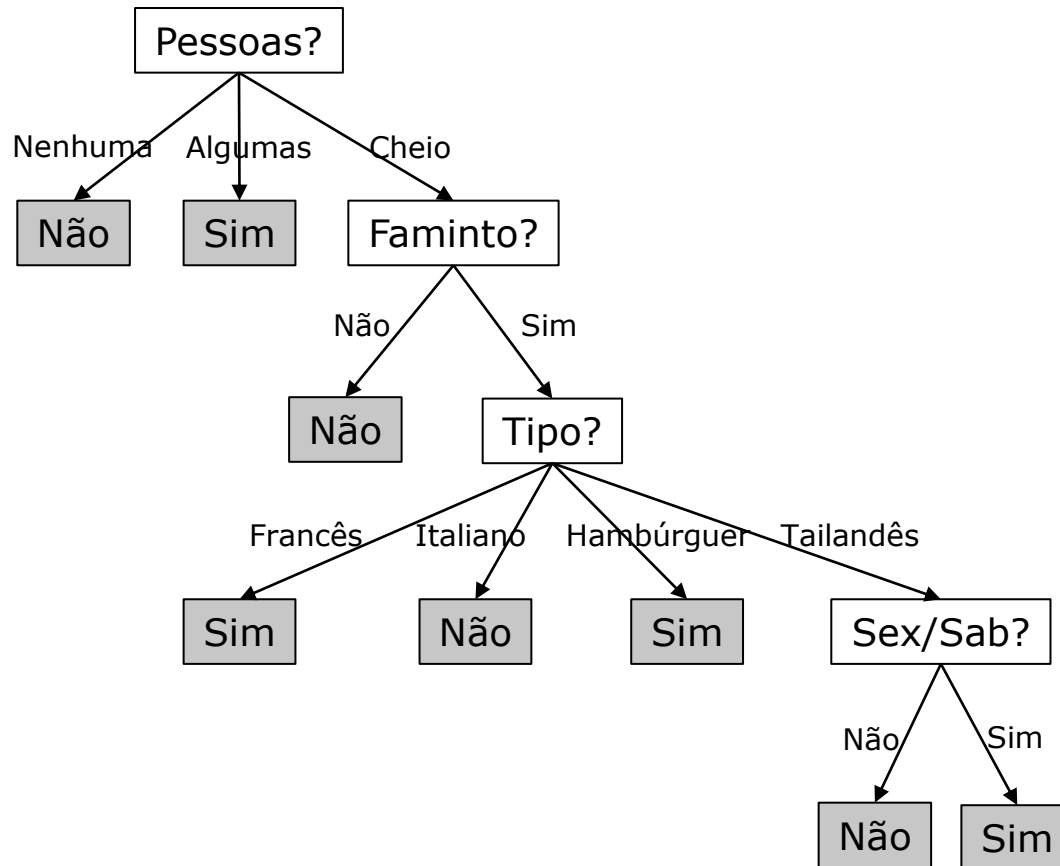


# Gerando Árvores de Decisão a partir de Exemplos

- ❏ Quando não existem atributos restantes, mas ainda existem exemplos positivos e negativos significa que:
  - ❏ Esses exemplos têm exatamente a **mesma descrição**, mas **classificações diferentes**. Isso acontece quando alguns dos dados estão incorretos, ou seja há **ruído nos dados**.
  - ❏ Também acontece quando os atributos **não dão informação suficiente** para descrever a situação completamente, ou quando o domínio é realmente **não-determinístico**.
  - ❏ Uma saída simples do problema é a utilização de uma **votação majoritária**.



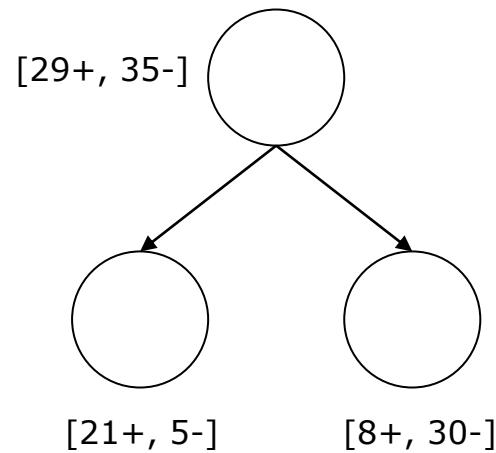
# Gerando Árvores de Decisão a partir de Exemplos





# Escolhendo os Melhores Atributos

💡 Qual é o melhor atributo?





# 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  $S$ .



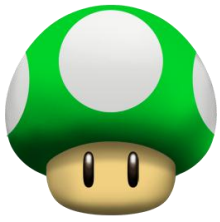
# Escolhendo os Melhores Atributos

- ❏ **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 se 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 número maior de exemplos) e repete-se novamente o processo.



# Medindo Desempenho

