

Aprendizado de Máquina e Deep Learning

Árvore de Decisão

Prof. Dr. Thiago Meirelles Ventura

Exercício 2

- Jogue Akinator: <https://pt.akinator.com>
- Ao acertar um personagem, anote:
 - Quantas perguntas foram feitas
 - Quantas vezes o seu personagem já foi jogado











Discussão

- Como isso funciona?
- O que acontece quando ele erra?
- Por quê foram feitas aquelas primeiras perguntas?

Árvore de decisão











- Algoritmo de aprendizado de máquina
 - Utilizado principalmente para classificação
 - Há variações para ser utilizado com problemas de regressão
- Utiliza dos atributos disponíveis para tomar a decisão de classificação
- Deixa claro como a decisão de classificação foi tomada

Explicando por meio de exemplo

Pessoa		Comprimento do Cabelo	Peso	Idade	Classe: Sexo
	Homer	0	250	36	M
	Marge	10	150	34	F
	Bart	2	90	10	M
	Lisa	6	78	8	F
	Maggie	4	20	1	F
	Abe	1	170	70	M
	Selma	8	160	41	F
	Otto	10	180	38	M
	Krusty	6	200	45	M
	Comic	8	290	38	?

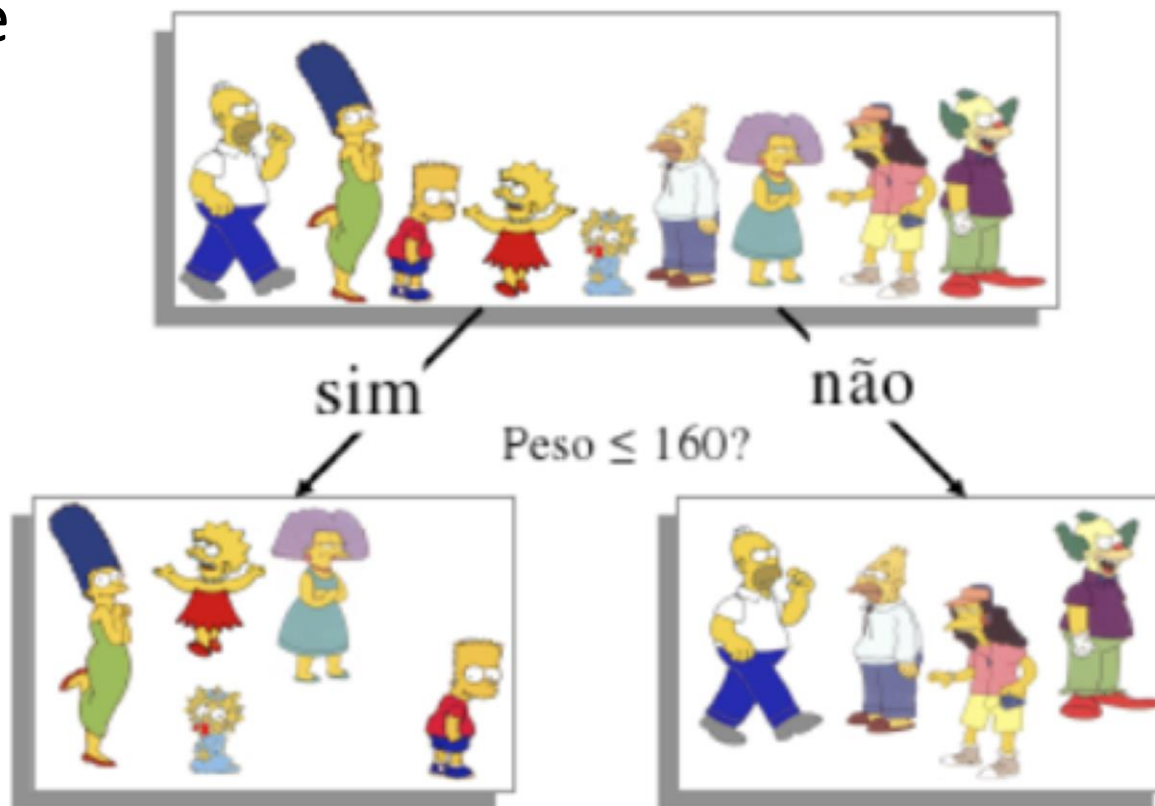
Explicando por meio de exemplo

- Nessa base temos 4 características e a classe de cada personagem
- Características: nome do personagem, comprimento do cabelo, peso e idade
- Classe: masculino ou feminino

Pessoa	Comprimento do Cabelo	Peso	Idade	Classe: Sexo
 Homer	0	250	36	M
 Marge	10	150	34	F
 Bart	2	90	10	M
 Lisa	6	78	8	F
 Maggie	4	20	1	F
 Abe	1	170	70	M
 Selma	8	160	41	F
 Otto	10	180	38	M
 Krusty	6	200	45	M
 Comic	8	290	38	?

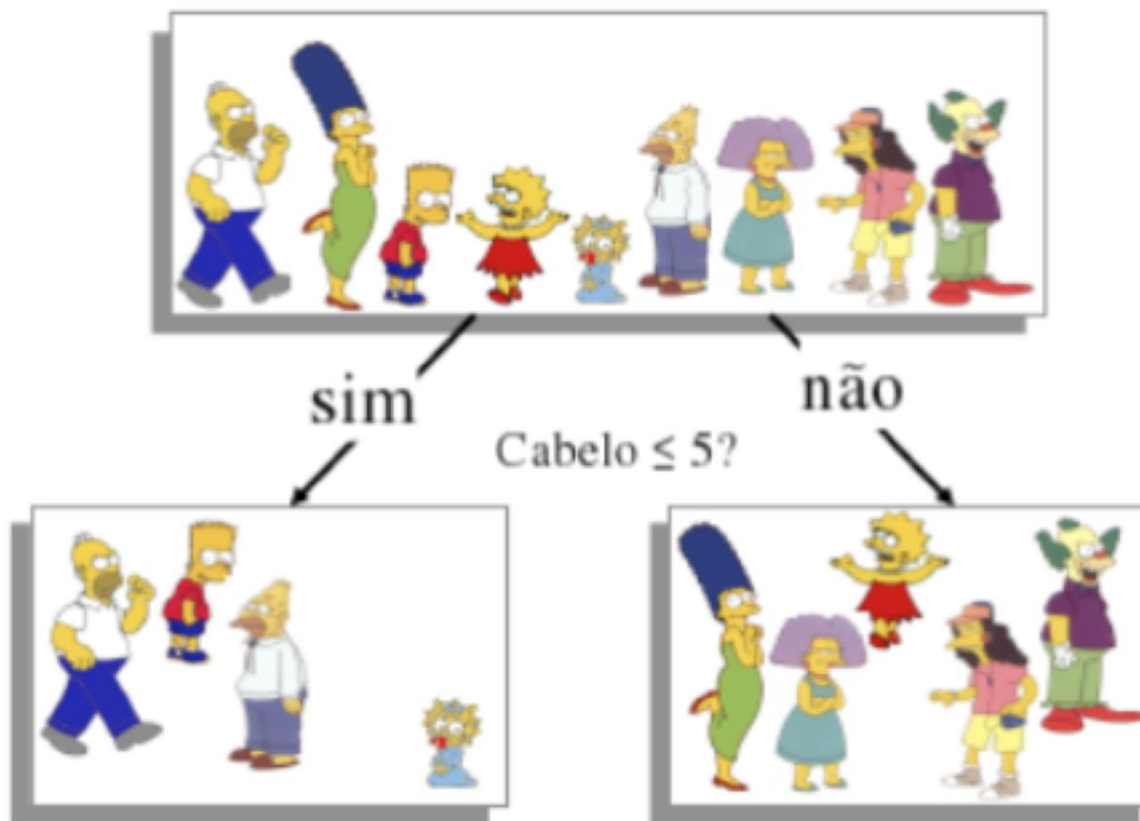
Explicando por meio de exemplo

- Análise das características, separando os dados de acordo com cada atributo existente



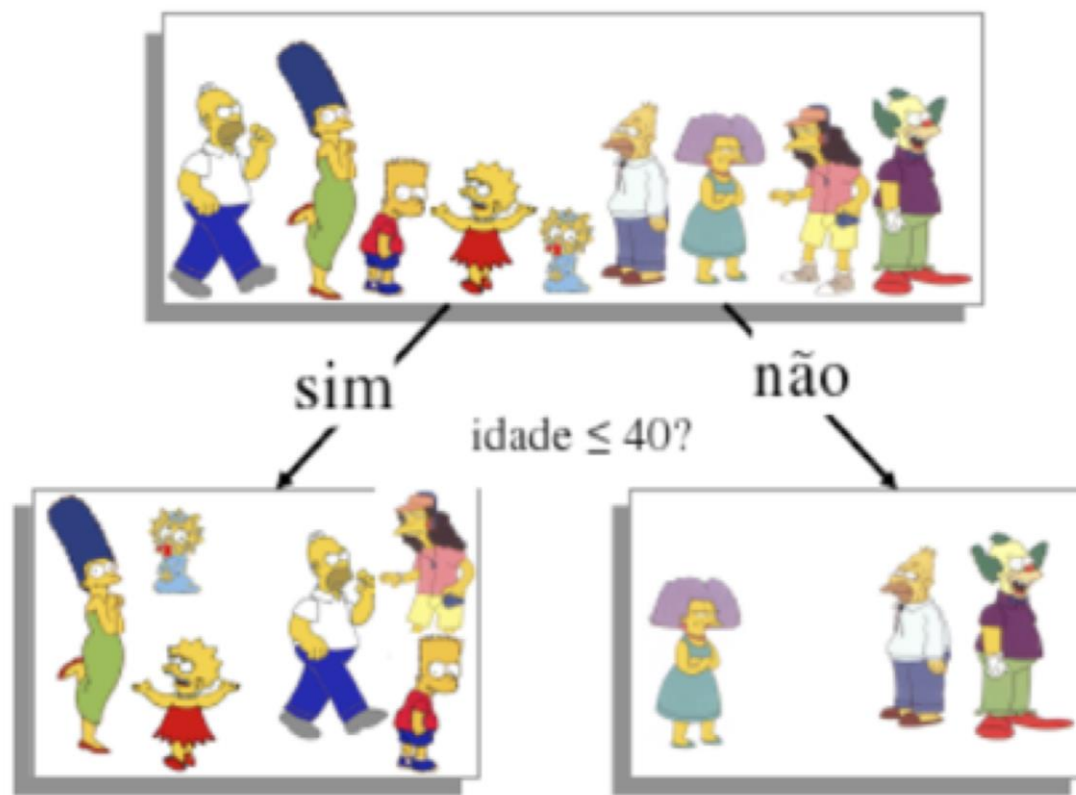
Explicando por meio de exemplo

- Análise das características, separando os dados de acordo com cada atributo existente



Explicando por meio de exemplo

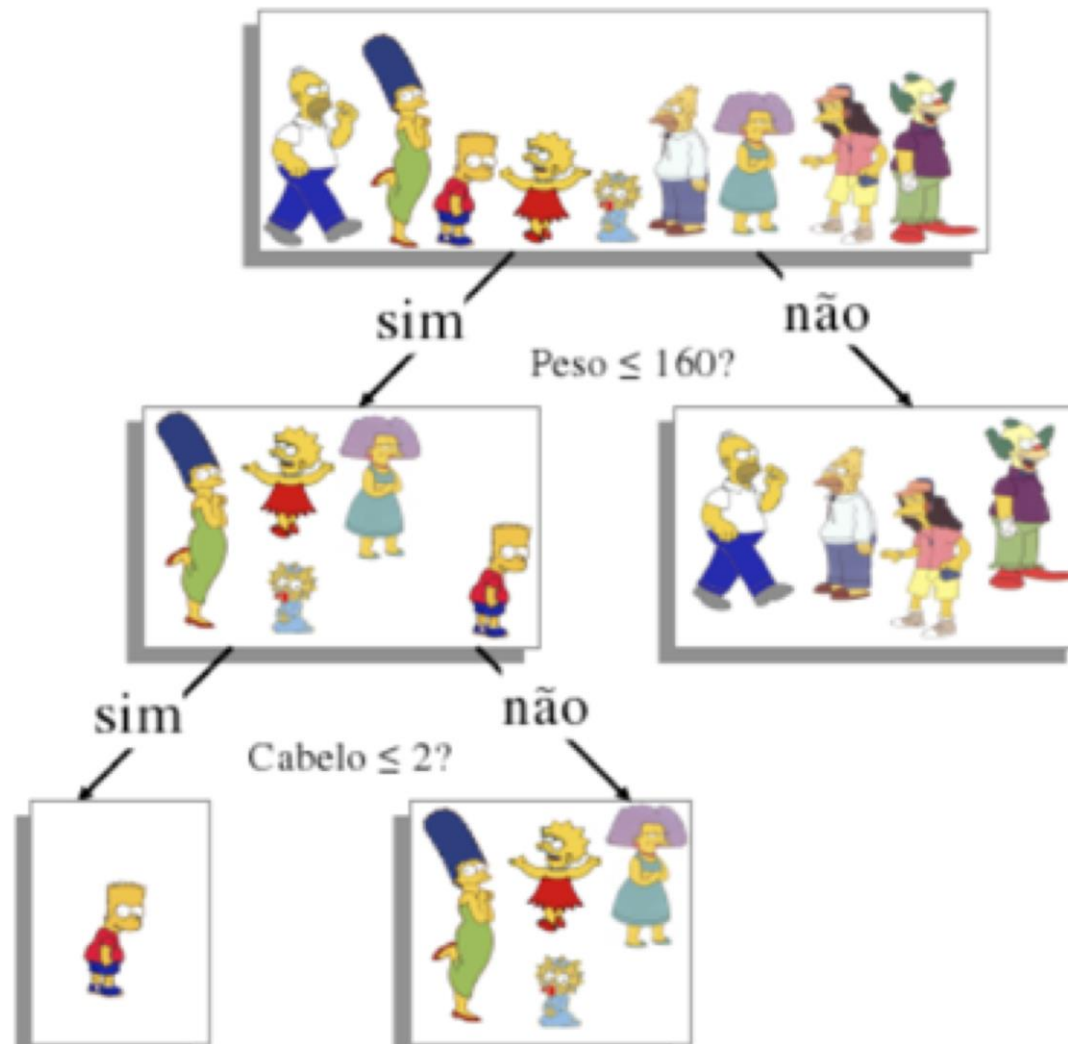
- Análise das características, separando os dados de acordo com cada atributo existente



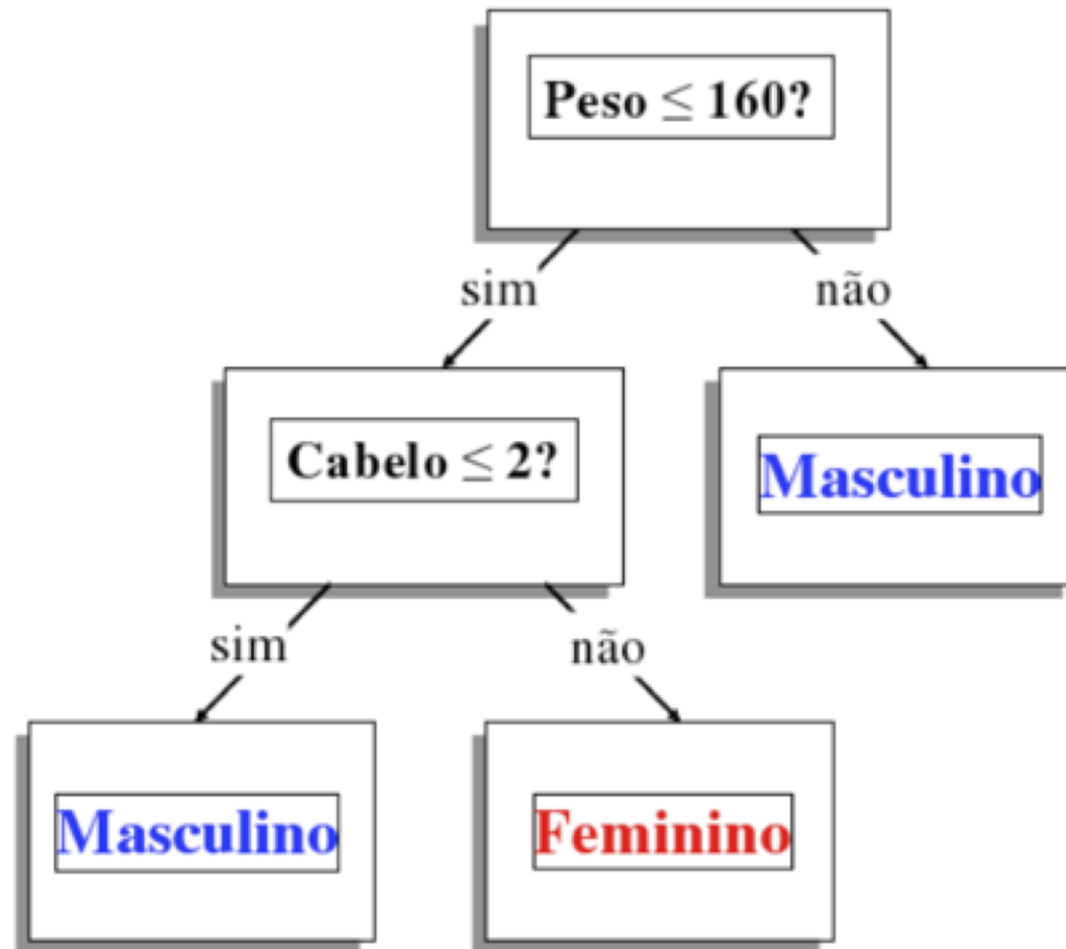
Explicando por meio de exemplo

- Considerando a classe, nenhum atributo conseguiu separar corretamente os registros
- Pelo menos isso não foi possível individualmente
- Mas e se usarmos mais de um atributo?

Explicando por meio de exemplo



Explicando por meio de exemplo




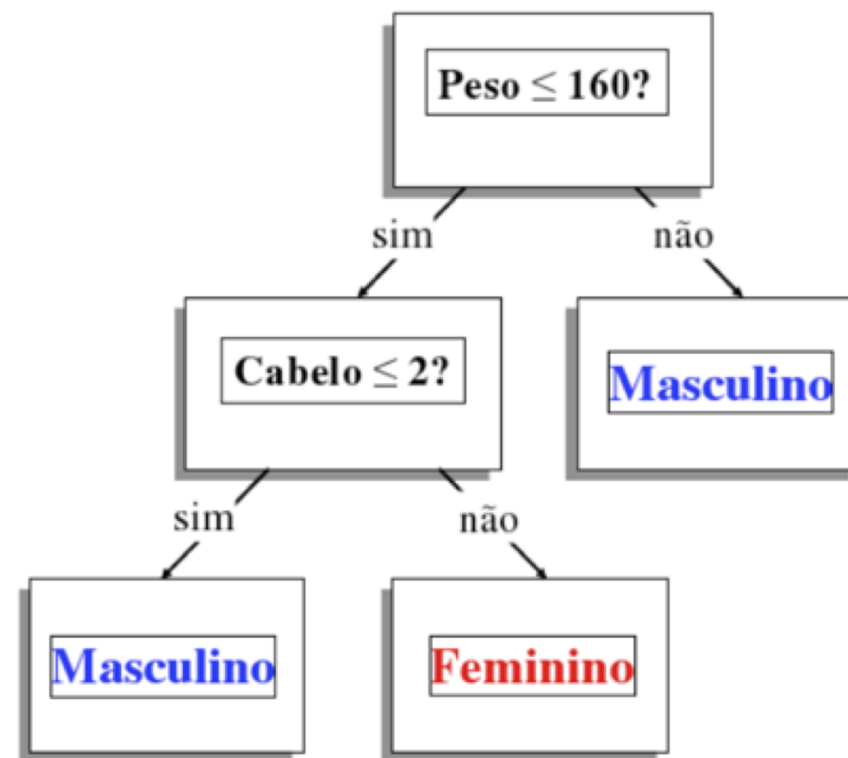
Funcionamento

- O método de árvore de decisão realiza classificações executando uma sequência de testes
- A cada nó interno da árvore corresponde a um teste do valor de uma das características da base de treinamento
- As ramificações a partir do nó são identificadas com os possíveis valores do teste
- Cada nó folha (nó final) especifica o valor a ser retornado se aquela folha for alcançada
 - Representa a classe do exemplo que está sendo verificado

Exercício 3

- Avalia como o personagem Comic será classificado de acordo com a árvore de decisão montada
- E uma personagem que tivesse um peso de 90, cabelo de comprimento 5 e idade 27?

Pessoa	Comprimento do Cabelo	Peso	Idade	Classe: Sexo
 Comic	8	290	38	?



Construção de uma árvore de decisão

- Como podemos montar uma árvore de decisão?
- Quais atributos devemos utilizar?
- Qual deve ser o primeiro atributo analisado? E o segundo?

Construção de uma árvore de decisão

- Existem vários algoritmos para montar uma árvore de decisão
- O ID3 é um bom algoritmo para aprender o funcionamento geral

Construção de uma árvore de decisão

- Algoritmo ID3
 - A árvore de decisão é construída de forma recursiva
 - No início, todos os exemplos de treinamento estão na raiz
 - Depois, é escolhido a característica que melhor divide os exemplos
 - Para a característica escolhida, é criado um nó filho para cada valor possível da característica
 - O processo é realizado novamente para cada nó que contenha diferentes exemplos

Construção de uma árvore de decisão

- Para esse algoritmo funcionar as características devem ser categóricas
 - os valores contínuos devem ser discretizados
- É necessário ter alguma forma para avaliar as características, para decidir qual a melhor a ser utilizada para dividir os exemplos
 - Para isso é usado o ganho de informação

Construção de uma árvore de decisão

- Para cada nó a ser avaliado na árvore é calculado o ganho de informação de cada característica
- A característica selecionada é a que apresentar o maior ganho de informação
- Para calcular o ganho de informação é necessário calcular a entropia

Entropia

- A entropia serve para medir o grau de pureza de um conjunto
- É dada por:

$$\text{entropia}(\text{nó}) = - \sum_{i=1}^c p(i / \text{nó}) \cdot \log_2[p(i / \text{nó})]$$

- c é o número de classes
- $p(i/\text{nó})$ é a fração dos exemplos pertencentes à classe i no respectivo nó

Ganho de informação

- Ganho de informação é dada por:

$$\text{ganho} = \text{entropia}(\text{pai}) - \sum_{j=1}^n \left[\frac{N(v_j)}{N} \text{entropia}(v_j) \right]$$

- n é número de características
- N é o número de exemplo no nó pai
- $N(v_j)$ é o número de exemplos associados ao nó filho v_j

Escolha dos nós até os nós folhas

- Os cálculos devem ser refeitos para cada novo nó da árvore
- Esses passos se repetem até que se encontre um critério de parada
 - Todos os exemplos para um dado nó pertencem à mesma classe
 - Não existe atributos remanescentes para particionamento (a classe majoritária é escolhida para classificar a folha)

Construção - exemplo

- Vamos utilizar como exemplo nossa base dos Simpsons

Pessoa	Comprimento do Cabelo	Peso	Idade	Classe: Sexo
 Homer	0	250	36	M
 Marge	10	150	34	F
 Bart	2	90	10	M
 Lisa	6	78	8	F
 Maggie	4	20	1	F
 Abe	1	170	70	M
 Selma	8	160	41	F
 Otto	10	180	38	M
 Krusty	6	200	45	M

Construção - exemplo

- Calcular o ganho de informação das características existentes
- Avaliar qual delas seria a melhor para ser escolhida como a raiz da árvore
- Análise das classes
 - 5 exemplos da classe Masculino
 - 4 exemplos da classe Feminino
 - Total de 9 exemplos

Construção - exemplo

- Entropia do nó pai

$$\text{entropia}(\text{nó}) = - \sum_{i=1}^c p(i / \text{nó}) \cdot \log_2[p(i / \text{nó})]$$

$$\text{Entropia} = -5/9 \cdot \log_2(5/9) - 4/9 \cdot \log_2(4/9) = 0,99$$

- Agora podemos calcular a entropia de cada característica

Construção - exemplo

- Cabelo

- Por ser valores contínuos, é necessário uma definição de intervalos
 - cabelo curto: ≤ 5
 - cabelo longo: > 5
- Para cabelos curtos temos 3 masculinos e 1 feminino
- Para cabelos longos temos 2 masculinos e 3 femininos

$$\text{entropia}(\text{nó}) = - \sum_{i=1}^c p(i / \text{nó}) \cdot \log_2[p(i / \text{nó})]$$

$$\text{entropia}(\text{cabelo curto}) = -3/4 \cdot \log_2(3/4) - 1/4 \cdot \log_2(1/4) = 0,81$$

$$\text{entropia}(\text{cabelo longo}) = -2/5 \cdot \log_2(2/5) - 3/5 \cdot \log_2(3/5) = 0,97$$

Construção - exemplo

- Cabelo

- Entropia do nó pai e de cada valor de cabelo

$$\text{entropia}(\text{pai}) = 0,99$$

$$\text{entropia}(\text{cabelo curto}) = 0,81$$

$$\text{entropia}(\text{cabelo longo}) = 0,97$$

- Agora é possível calcular o ganho de informação da característica “cabelo”

$$\text{ganho} = \text{entropia}(\text{pai}) - \sum_{j=1}^n \left[\frac{N(v_j)}{N} \text{entropia}(v_j) \right]$$

$$\text{ganho}(\text{cabelo}) = 0,99 - 4/9 * 0,81 - 5/9 * 0,97 = 0,09$$

Construção - exemplo

- Peso

- O mesmo procedimento pode ser feito para avaliar a característica de peso
- Intervalos e exemplos na base de treinamento
 - ≤ 160 : 1 masculino e 4 femininos
 - > 160 : 4 masculino e 0 feminino

$$\text{entropia}(\leq 160) = -1/5 \cdot \log_2(1/5) - 4/5 \cdot \log_2(4/5) = 0,72$$

$$\text{entropia}(> 160) = -4/4 \cdot \log_2(4/4) - 0/4 \cdot \log_2(0/4) = 0$$

$$\text{ganho}(\text{peso}) = 0,99 - 5/9 \cdot 0,72 - 4/9 \cdot 0 = 0,59$$

Construção - exemplo

- Idade

- O mesmo procedimento pode ser feito para avaliar a característica de idade
- Intervalos e exemplos na base de treinamento
 - ≤ 40 : 3 masculino e 3 femininos
 - > 40 : 2 masculino e 1 feminino

$$\text{entropia}(\leq 40) = -3/6 \cdot \log_2(3/6) - 3/6 \cdot \log_2(3/6) = 1$$

$$\text{entropia}(> 40) = -2/3 \cdot \log_2(2/3) - 1/3 \cdot \log_2(1/3) = 0,92$$

$$\text{ganho(idade)} = 0,99 - 6/9 * 1 - 3/9 * 0,92 = 0,02$$

Construção - exemplo

- Temos:

$\text{ganho}(\text{cabelo}) = 0,09$

$\text{ganho}(\text{peso}) = 0,59$

$\text{ganho}(\text{idade}) = 0,02$

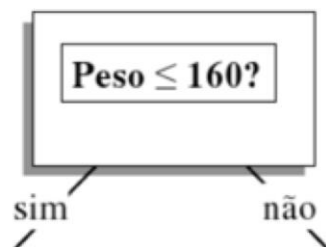
- Qual atributo deve ser selecionado para compor o primeiro nó da árvore de decisão?
 - O atributo de **peso**

Construção - exemplo

- Escolhendo o atributo peso seriam criados dois nós filhos
 - Exemplos com peso ≤ 160
 - Exemplos com pesos > 160
- Para cada um desses nós seria feito o mesmo procedimento para selecionar a característica a ser utilizada como próximo atributo
 - mas desta vez considerando apenas os exemplos que já foram filtrados pelos nós anteriores

Exercício 4

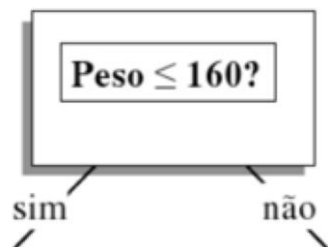
- Quais exemplos seriam considerados para cada partição (menor igual a 160 e maior que 160)?



Pessoa	Comprimento do Cabelo	Peso	Idade	Classe: Sexo
 Homer	0	250	36	M
 Marge	10	150	34	F
 Bart	2	90	10	M
 Lisa	6	78	8	F
 Maggie	4	20	1	F
 Abe	1	170	70	M
 Selma	8	160	41	F
 Otto	10	180	38	M
 Krusty	6	200	45	M

Exercício 5

- Mostre como definir os nós restantes da árvore de decisão



Pessoa	Comprimento do Cabelo	Peso	Idade	Classe: Sexo
 Homer	0	250	36	M
 Marge	10	150	34	F
 Bart	2	90	10	M
 Lisa	6	78	8	F
 Maggie	4	20	1	F
 Abe	1	170	70	M
 Selma	8	160	41	F
 Otto	10	180	38	M
 Krusty	6	200	45	M