

Árvores de Decisão

Fabício Olivetti de França

Universidade Federal do ABC

1. Árvores de Decisão

Árvores de Decisão

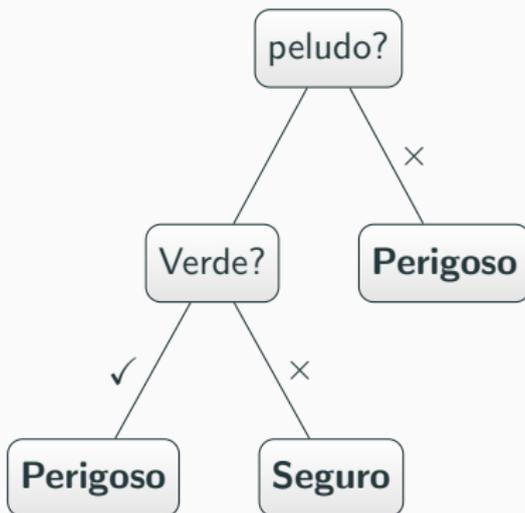
Relembrando nossa base de dados utilizada na aula passada:

Pele	Cor	Tamanho	Carne	Classe
peludo	marrom	grande	dura	seguro
peludo	verde	grande	dura	seguro
liso	vermelho	grande	macia	perigoso
peludo	verde	grande	macia	seguro
peludo	vermelho	pequeno	dura	seguro
liso	vermelho	pequeno	dura	seguro
liso	marrom	pequeno	dura	seguro
peludo	verde	pequeno	macia	perigoso
liso	verde	pequeno	dura	perigoso
peludo	vermelho	grande	dura	seguro
liso	marrom	grande	macia	seguro
liso	verde	pequeno	macia	perigoso
peludo	vermelho	pequeno	macia	seguro
liso	vermelho	grande	dura	perigoso
liso	vermelho	pequeno	dura	
peludo	verde	pequeno	dura	

Uma outra forma de pensar na solução para esse problema é através de decisões baseadas em uma sequência de perguntas e respostas:

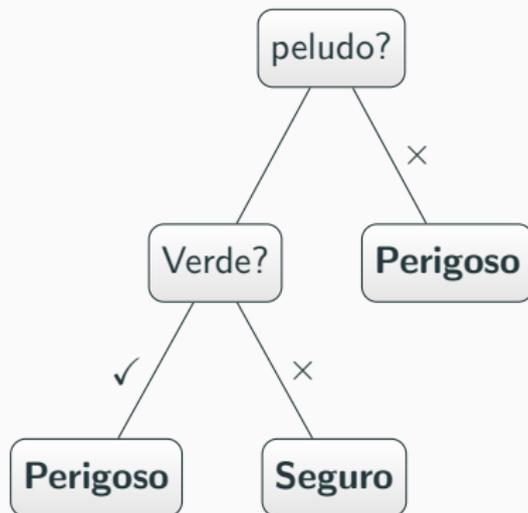


Cada nova pergunta aumenta nossa certeza em relação a classificação da amostra.



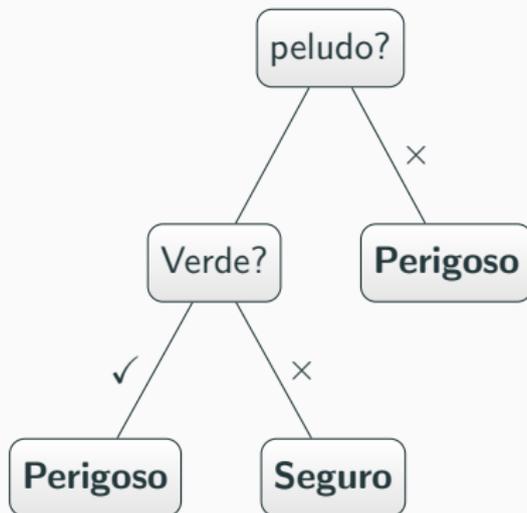
Árvore de Decisão

Essa técnica é conhecida como Árvore de Decisão.



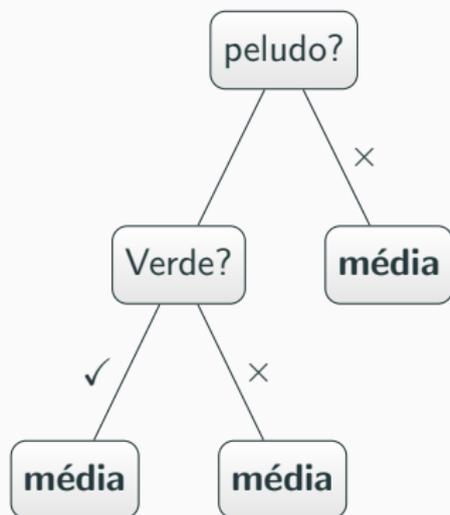
Árvore de Decisão

Cada nó da árvore divide os exemplos em dois grupos, cada divisão subsequente formam novos grupos.



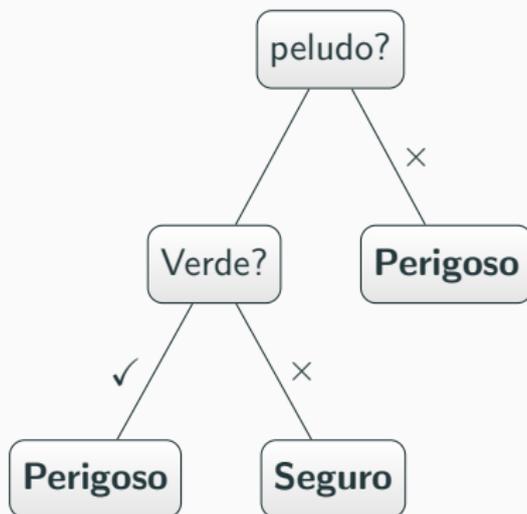
Árvore de Decisão

Para Regressão, basta fazer com que os nós folhas sejam a média dos valores-alvo do grupo referente aquele nó.



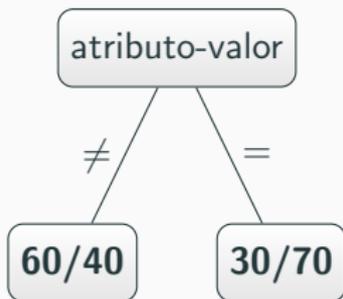
Árvore de Decisão

A grande questão é como construir essa árvore partindo dos exemplos de entrada.



Construindo a Árvore de Decisão

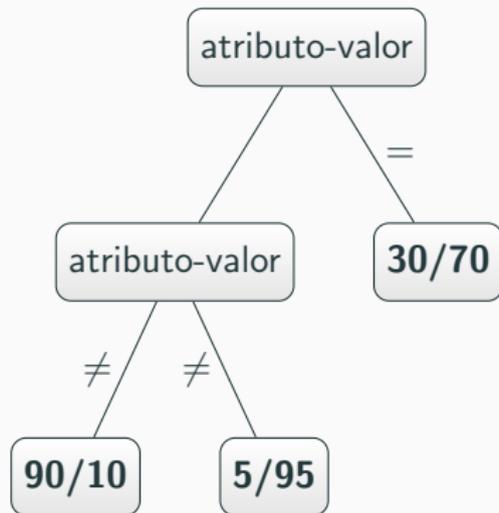
Assumindo duas classes, escolha um atributo-valor para ser o nó inicial e formule a questão para dividir os dados:



Os valores nos nós é a proporção de exemplos da classe1/classe2, respectivamente.

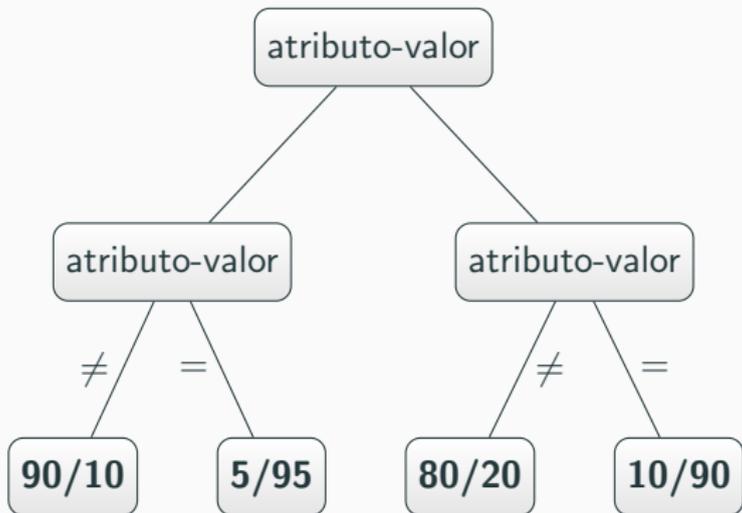
Construindo a Árvore de Decisão

Continue inserindo novos nós com os atributos restantes:



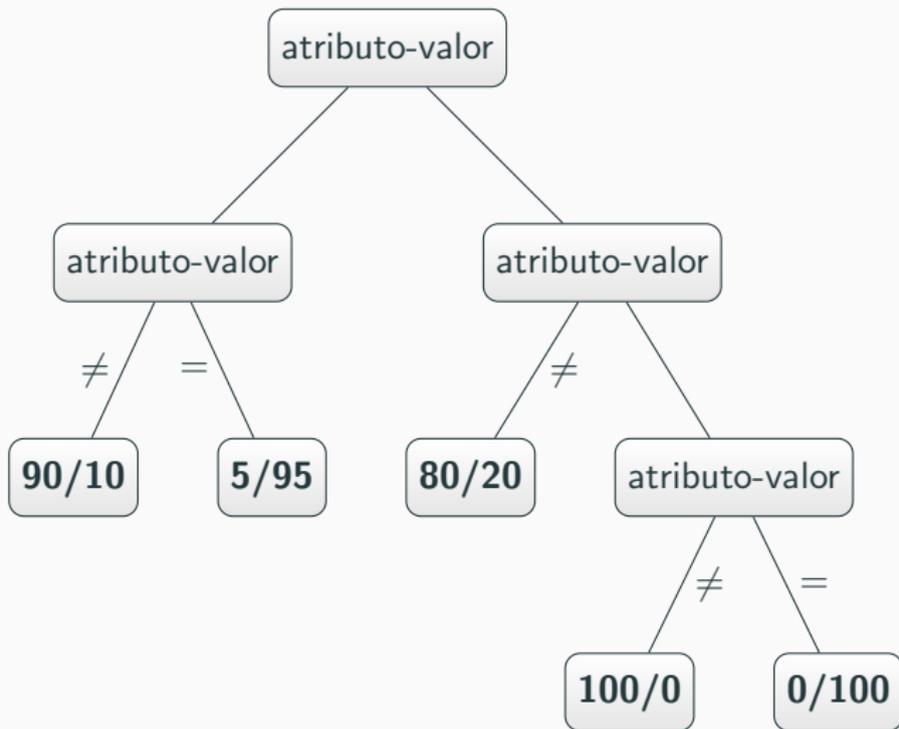
Construindo a Árvore de Decisão

Continue inserindo novos nós com os atributos restantes:



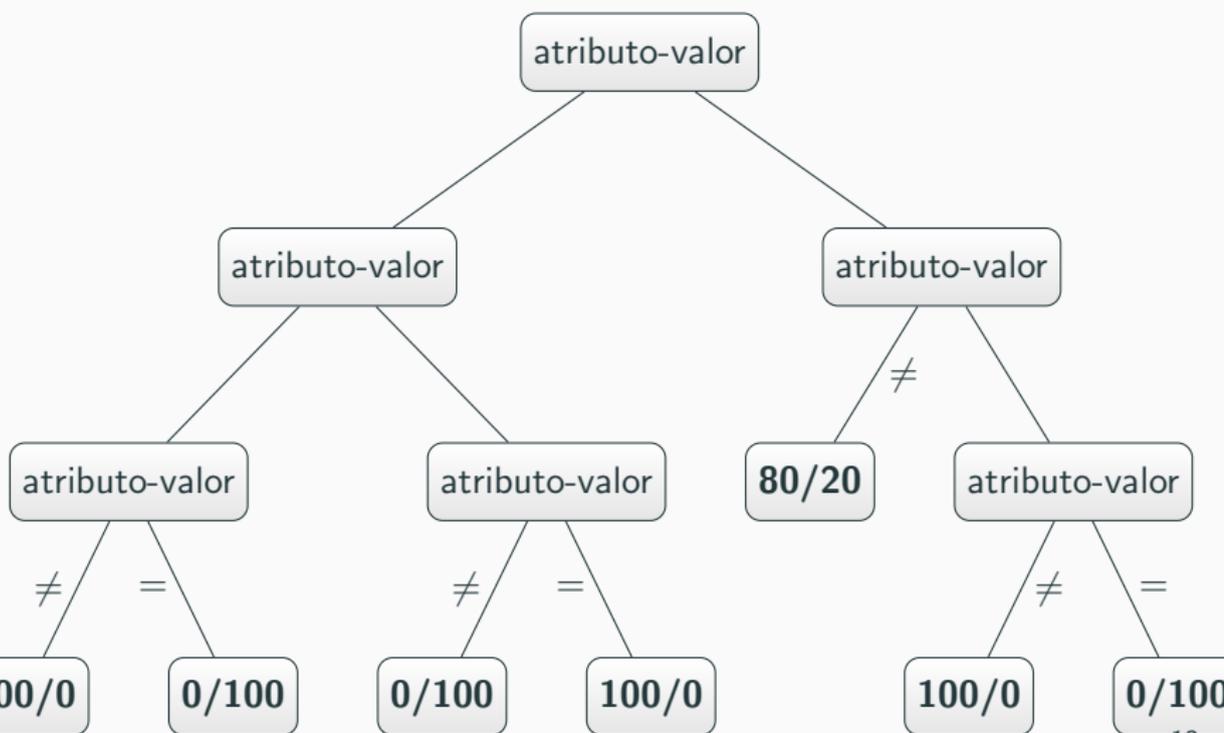
Construindo a Árvore de Decisão

Até onde devemos dividir?



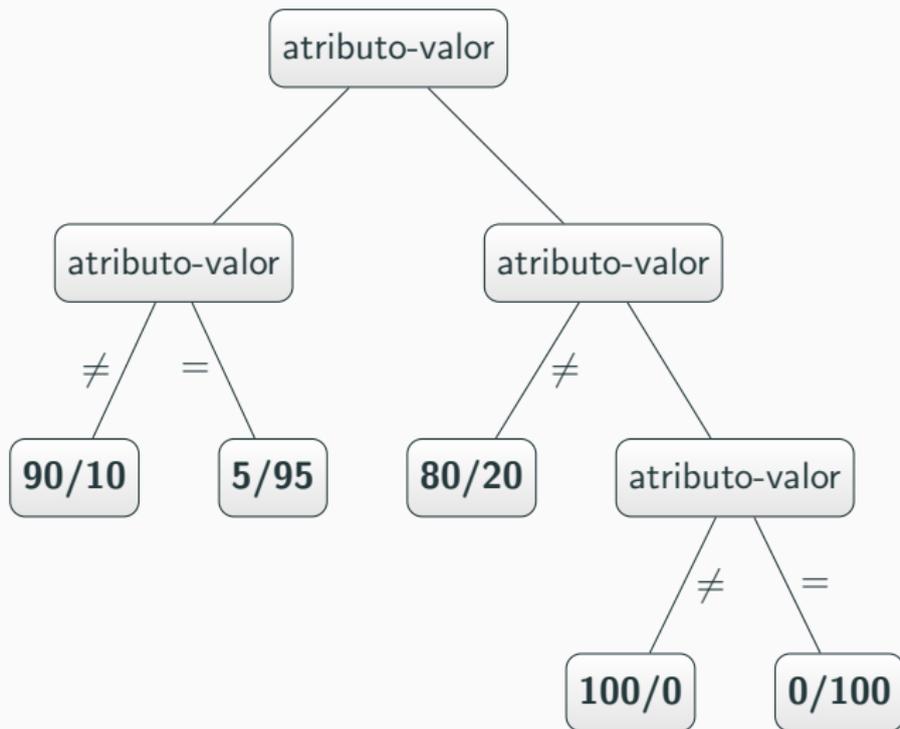
Construindo a Árvore de Decisão

Todos os nós folhas são puros...



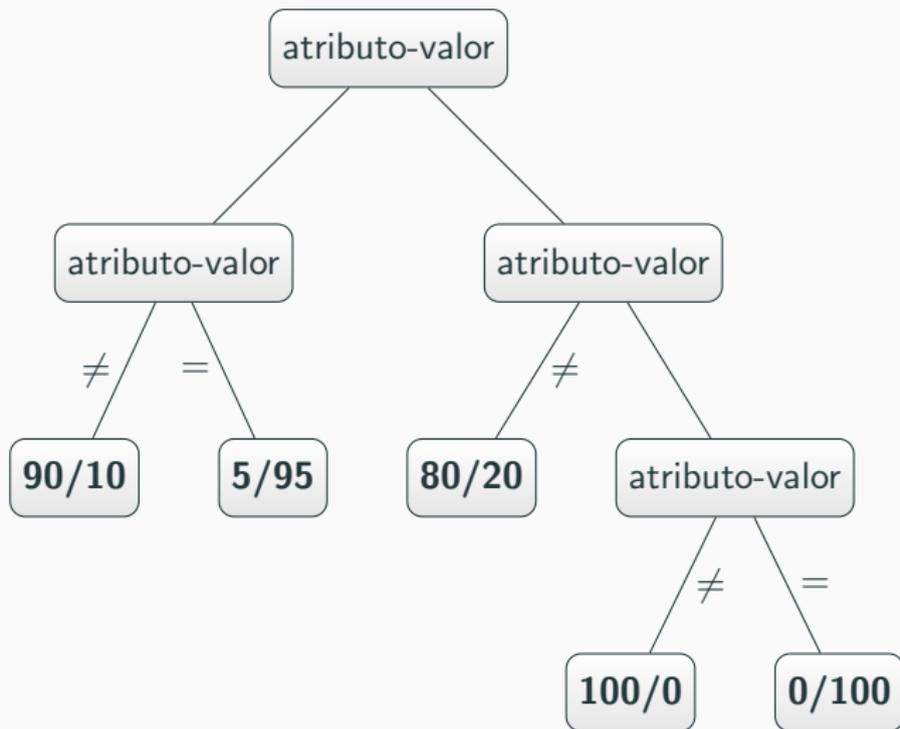
Construindo a Árvore de Decisão

...ou uma máxima altura é atingida...



Construindo a Árvore de Decisão

...ou um critério de desempenho é atingido.



Na prática o primeiro critério causa *overfitting* pois tende a criar questões específicas até que os nós folhas contenham apenas um exemplo da base. Os dois outros critérios costumam serem utilizados em conjunto.

Construindo a Árvore de Decisão

Como escolher o melhor atributo-valor para cada divisão?

Pele	Cor	Tamanho	Carne	Classe
peludo	marrom	grande	dura	seguro
peludo	verde	grande	dura	seguro
liso	vermelho	grande	macia	perigoso
peludo	verde	grande	macia	seguro
peludo	vermelho	pequeno	dura	seguro
liso	vermelho	pequeno	dura	seguro
liso	marrom	pequeno	dura	seguro
peludo	verde	pequeno	macia	perigoso
liso	verde	pequeno	dura	perigoso
peludo	vermelho	grande	dura	seguro
liso	marrom	grande	macia	seguro
liso	verde	pequeno	macia	perigoso
peludo	vermelho	pequeno	macia	seguro
liso	vermelho	grande	dura	perigoso

Construindo a Árvore de Decisão

Queremos que cada ramificação nos traga o máximo de informação!

Pele	Cor	Tamanho	Carne	Classe
peludo	marrom	grande	dura	seguro
peludo	verde	grande	dura	seguro
liso	vermelho	grande	macia	perigoso
peludo	verde	grande	macia	seguro
peludo	vermelho	pequeno	dura	seguro
liso	vermelho	pequeno	dura	seguro
liso	marrom	pequeno	dura	seguro
peludo	verde	pequeno	macia	perigoso
liso	verde	pequeno	dura	perigoso
peludo	vermelho	grande	dura	seguro
liso	marrom	grande	macia	seguro
liso	verde	pequeno	macia	perigoso
peludo	vermelho	pequeno	macia	seguro
liso	vermelho	grande	dura	perigoso

Construindo a Árvore de Decisão

Vamos utilizar o erro de classificação: $e = 1 - \max p(y)$

Pele	Cor	Tamanho	Carne	Classe
peludo	marrom	grande	dura	seguro
peludo	verde	grande	dura	seguro
liso	vermelho	grande	macia	perigoso
peludo	verde	grande	macia	seguro
peludo	vermelho	pequeno	dura	seguro
liso	vermelho	pequeno	dura	seguro
liso	marrom	pequeno	dura	seguro
peludo	verde	pequeno	macia	perigoso
liso	verde	pequeno	dura	perigoso
peludo	vermelho	grande	dura	seguro
liso	marrom	grande	macia	seguro
liso	verde	pequeno	macia	perigoso
peludo	vermelho	pequeno	macia	seguro
liso	vermelho	grande	dura	perigoso

Construindo a Árvore de Decisão

O erro inicial é $1 - 9/14 = 0.36$

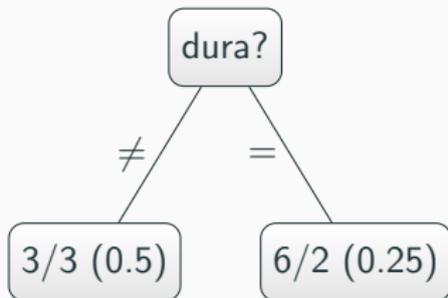
Pele	Cor	Tamanho	Carne	Classe
peludo	marrom	grande	dura	seguro
peludo	verde	grande	dura	seguro
liso	vermelho	grande	macia	perigoso
peludo	verde	grande	macia	seguro
peludo	vermelho	pequeno	dura	seguro
liso	vermelho	pequeno	dura	seguro
liso	marrom	pequeno	dura	seguro
peludo	verde	pequeno	macia	perigoso
liso	verde	pequeno	dura	perigoso
peludo	vermelho	grande	dura	seguro
liso	marrom	grande	macia	seguro
liso	verde	pequeno	macia	perigoso
peludo	vermelho	pequeno	macia	seguro
liso	vermelho	grande	dura	perigoso

Construindo a Árvore de Decisão

9/5 (0.36)

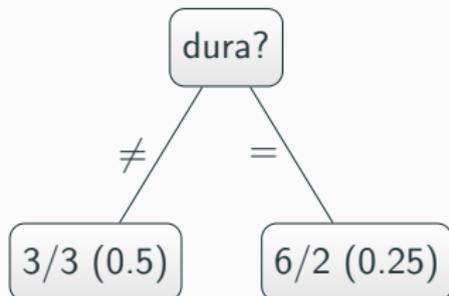
Construindo a Árvore de Decisão

Escolhendo se ele tem pele dura ou não, temos:



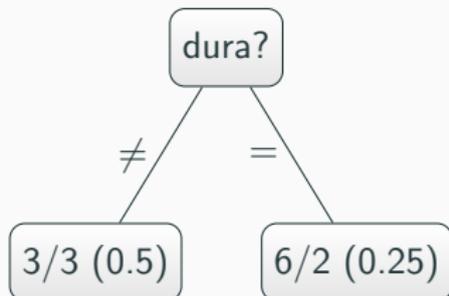
Construindo a Árvore de Decisão

A diferença do erro foi: $0.36 - (6/14) \cdot 0.5 - (8/14) \cdot 0.25 = 0.00$. Não houve ganho, a busca para por aqui!



Construindo a Árvore de Decisão

A divisão por erro de predição não leva em conta uma distribuição uniforme dos exemplos nas divisões.



Construindo a Árvore de Decisão

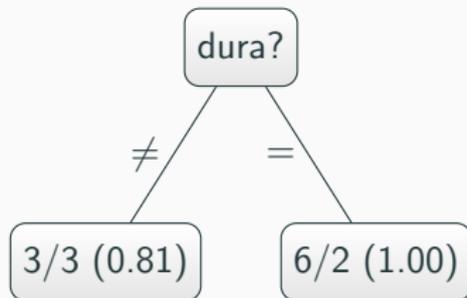
Vamos utilizar a entropia: $H = -\sum_{i=1}^n p(i) \log_2 p(i)$. Inicialmente temos:

$$-(9/14) \log (9/14) - (5/14) \log (5/14) = 0.94$$

$$9/5 (0.36)$$

Construindo a Árvore de Decisão

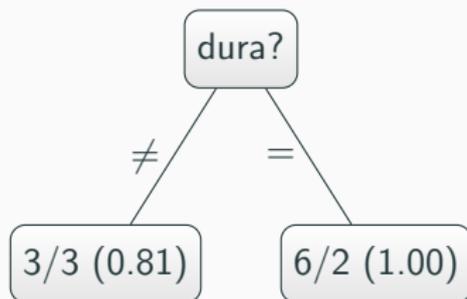
Escolhendo se ele tem pele dura ou não, temos:



Construindo a Árvore de Decisão

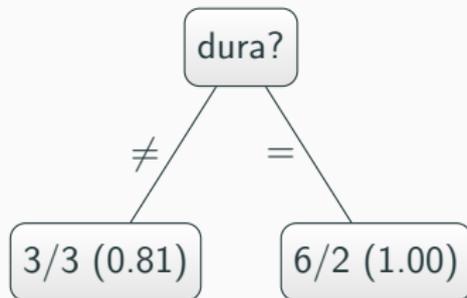
A diferença de entropia fica:

$$0.94 - (8/14) \cdot 0.81 - (6/14) \cdot 1.0 = 0.05$$



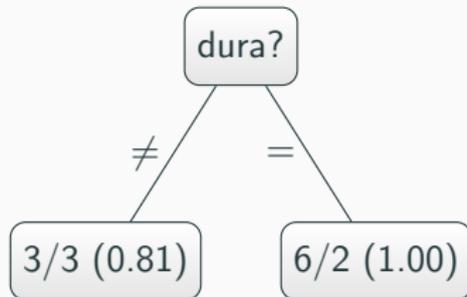
Construindo a Árvore de Decisão

Agora temos um ganho!



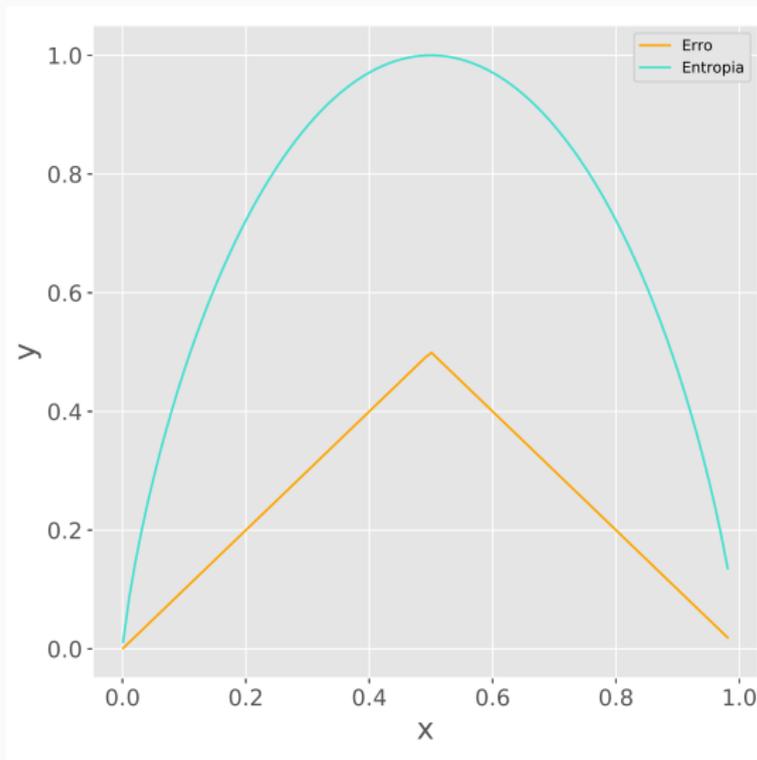
Construindo a Árvore de Decisão

Por que o erro de classificação nos enganou?



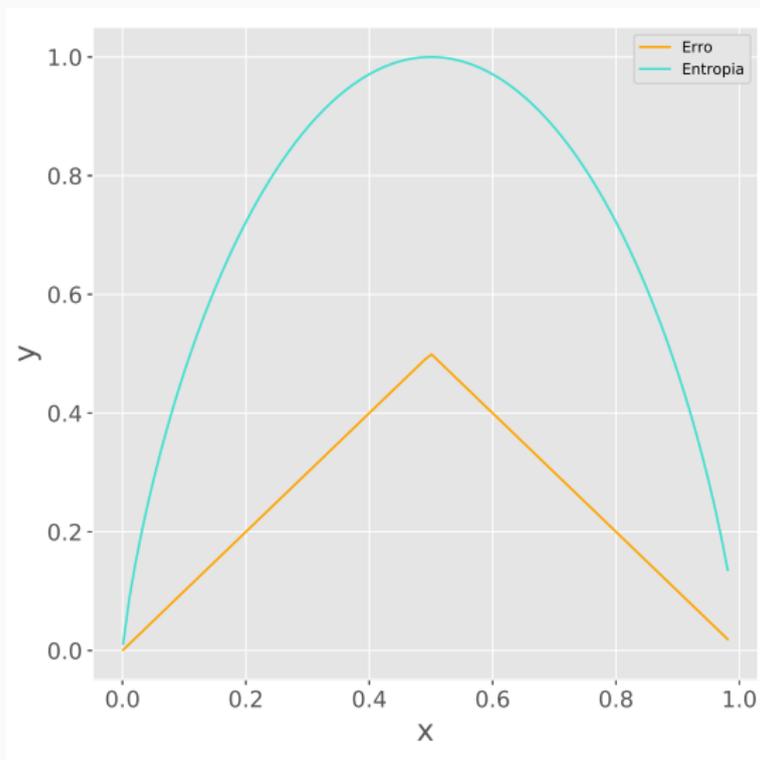
Erro vs Entropia

A função de erro forma um *triângulo* com pico no centro, a entropia tem uma curva mais suave com pico também no centro:



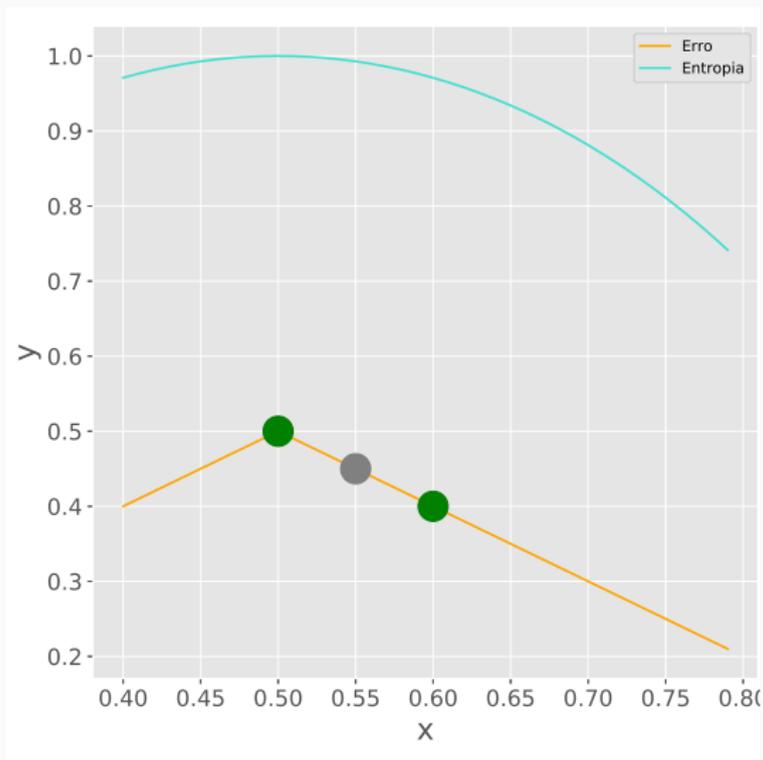
Erro vs Entropia

Mas qual a diferença?



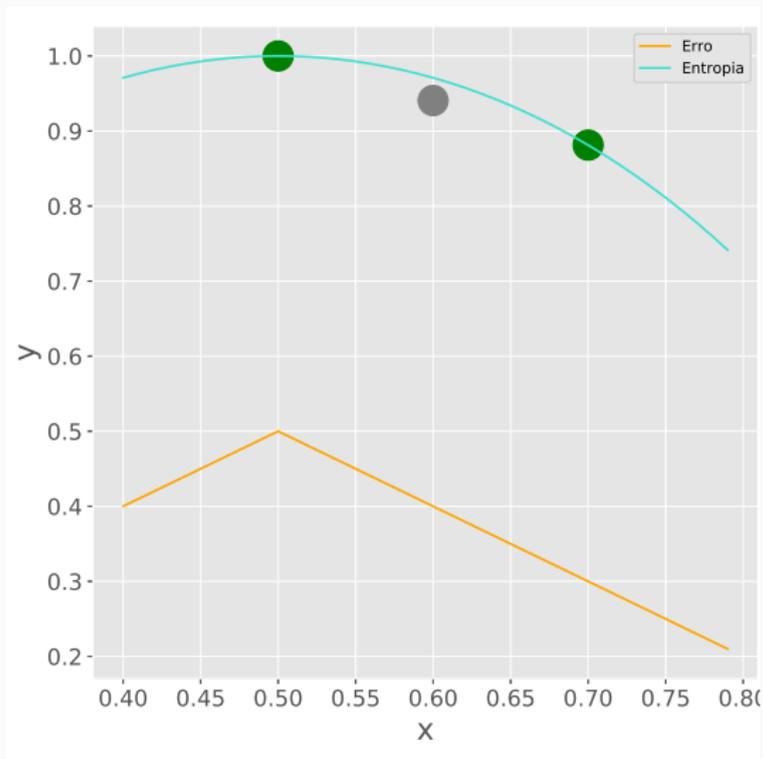
Erro vs Entropia

No erro de predição, a média entre o erro do nó da esquerda e nó da direita tem chances de coincidir com o erro do nó central.



Erro vs Entropia

Na entropia, isso não ocorre:



Na prática a métrica utilizada é o Índice de Gini:

$$E = 1 - \sum_{i=1} np(i)^2.$$

As Árvores de Decisão costumam causar overfitting, pois é capaz de se especializar a nível de amostra única.

Pequenas mudanças nos exemplos podem causar uma grande mudança na predição, dependendo do que for alterado.

Solução: poda de árvores

Para a poda das árvores utilizar o erro de predição como critério é suficiente para generalizar a predição.

A ideia é "corte as folhas com erro de predição menor do que x , para evitar especialização".

É fácil interpretar e converter para uma lógica *se-então-senão*.

Funciona para qualquer tipo de variável sem necessidade de transformação.

Não é necessário escalonar ou normalizar variáveis.

```
from sklearn.tree import DecisionTreeClassifier

clf = DecisionTreeClassifier(criterion='gini',
                             max_features=10, max_depth=5)
clf = clf.fit(X_train, y_train)
yi = clf.predict(xi)
```

Alterar para *DecisionTreeRegressor* para regressão.

Na próxima aula aprenderemos sobre:

- Redes Neurais Artificiais.

Complete os Laboratórios:

Decision_Trees_Exercises.ipynb