

DECISION TREE

Prof. Nielsen Rechia

nielsen.machado@uniritter.edu.br

DECISION TREE

2

Paradigmas	Supervisionado		Não-supervisionado	
	Classificação		Análise associativa	
Tarefas	Regressão		Agrupamento (<i>clustering</i>)	
	Outros		Redução de dimensionalidade	
			Outros	

7 tarefas comuns de aprendizado de máquina:
<http://vitalflux.com/7-common-machine-learning-tasks-related-methods/>

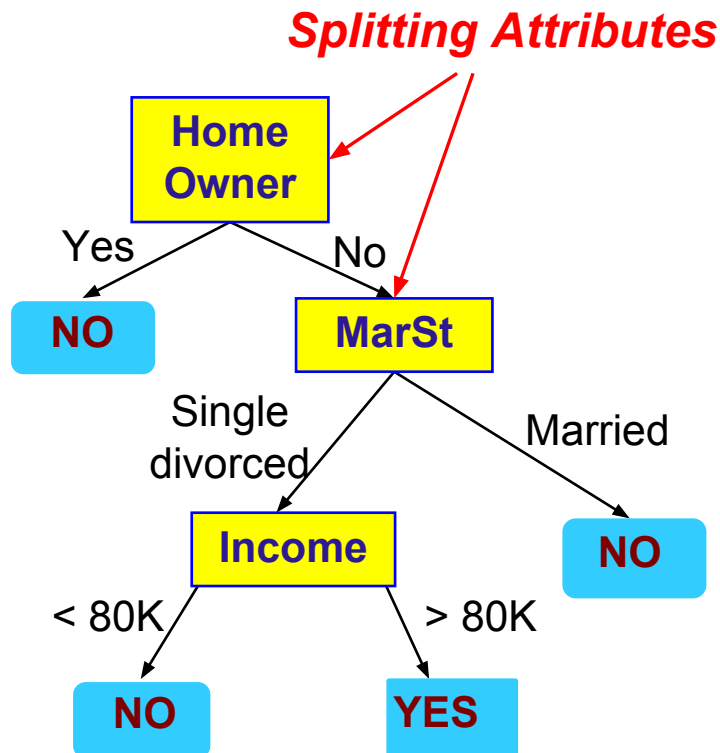
EXEMPLO DE UMA DECISION TREE

3

categorical categorical continuous class

ID	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Training Data



Model: Decision Tree

OUTRO EXEMPLO

categorical

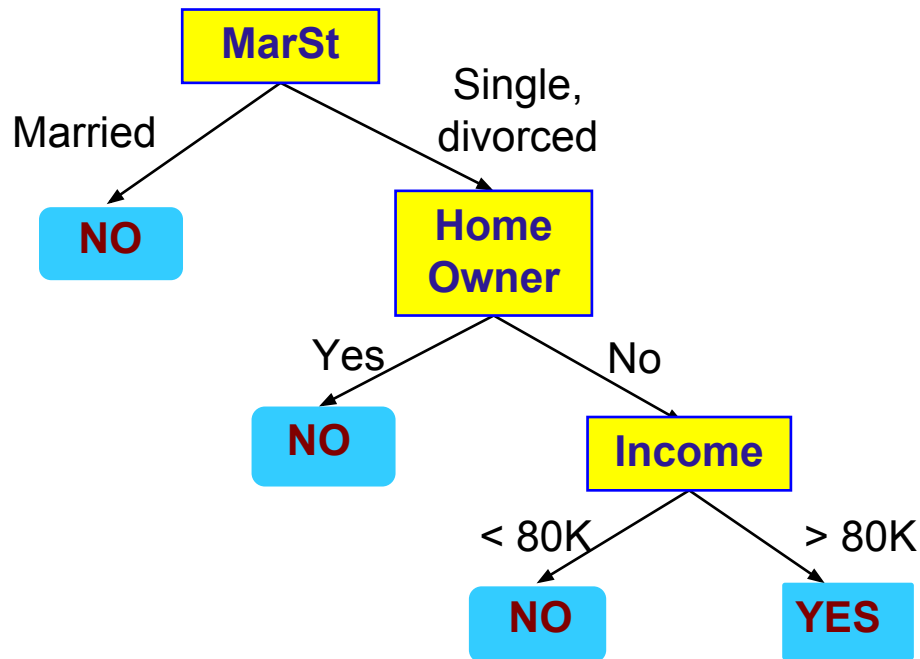
categorical

continuous

class

ID	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Training Data

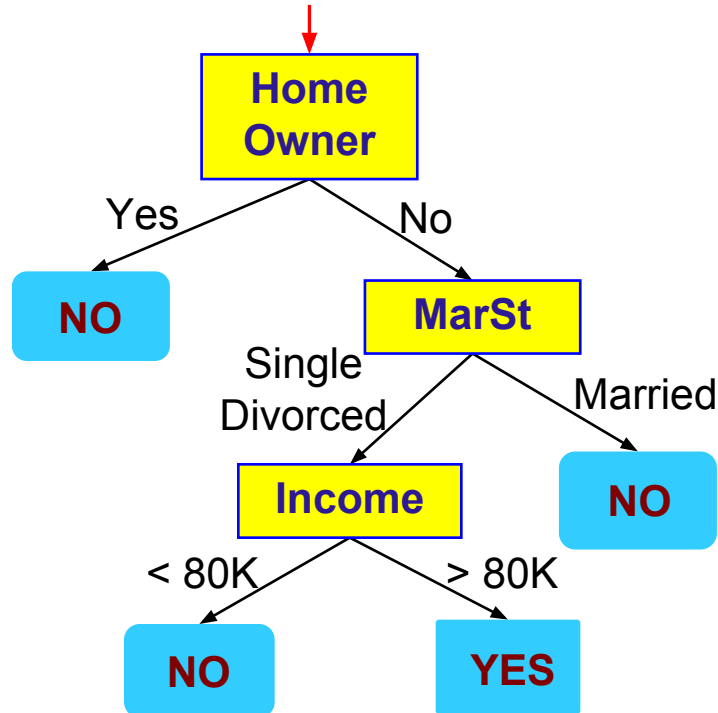


There could be more than one tree that fits the same data!

DECISION TREE

5

Start from the root of tree.

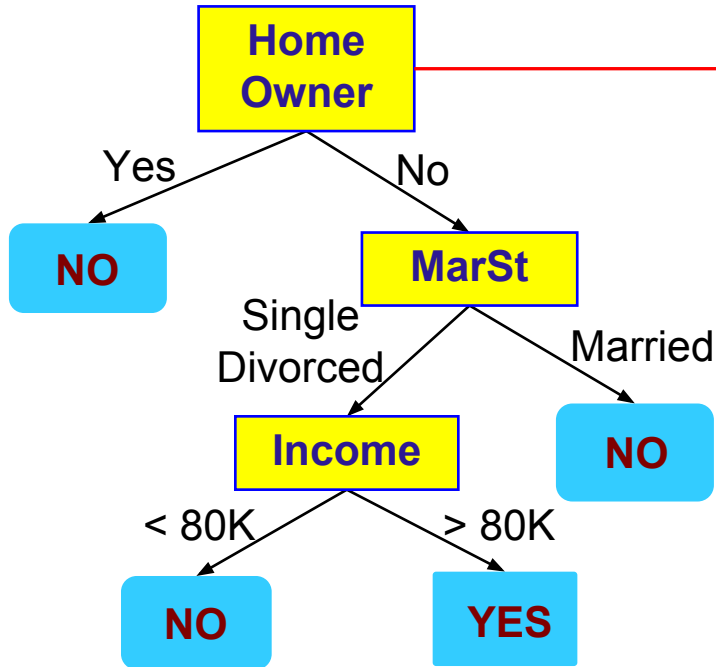


Home Owner	Marital Status	Annual Income	Defaulted Borrower
No	Married	80K	?

Test Data

DECISION TREE

6

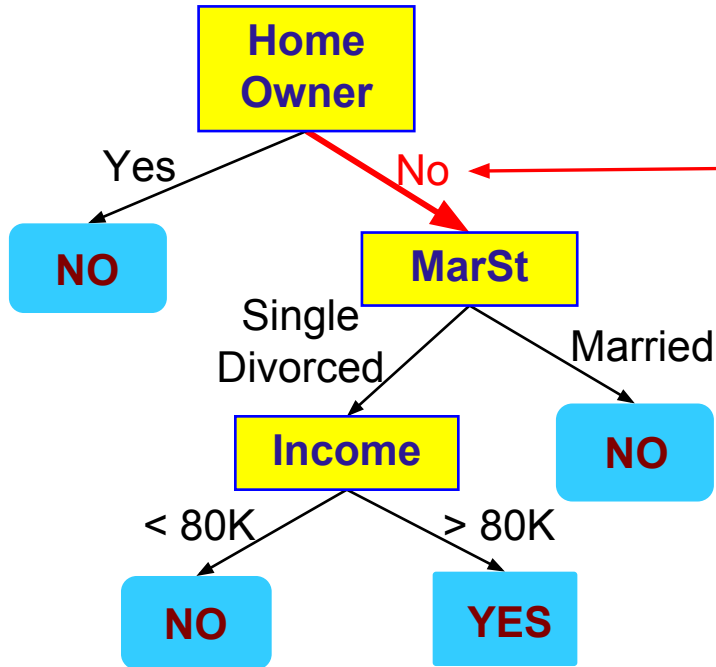


Home Owner	Marital Status	Annual Income	Defaulted Borrower
No	Married	80K	?

Test Data

DECISION TREE

7

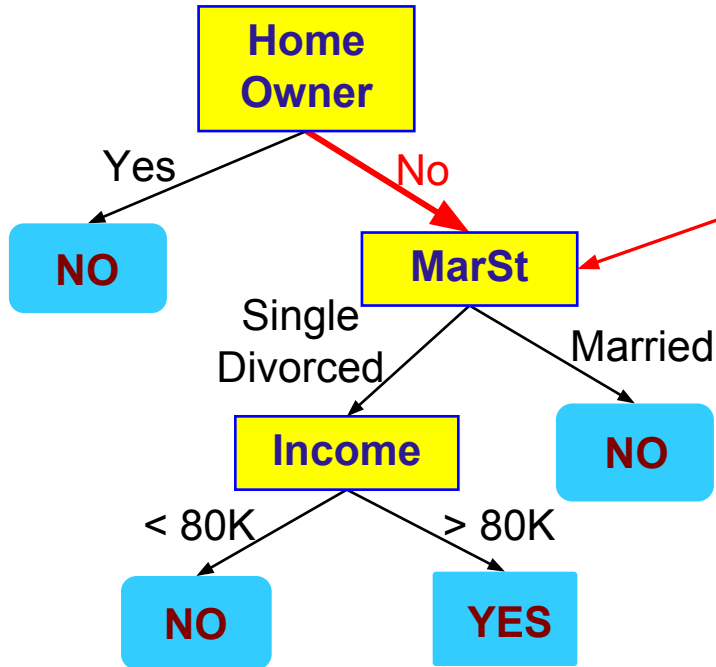


Home Owner	Marital Status	Annual Income	Defaulted Borrower
No	Married	80K	?

Test Data

DECISION TREE

8

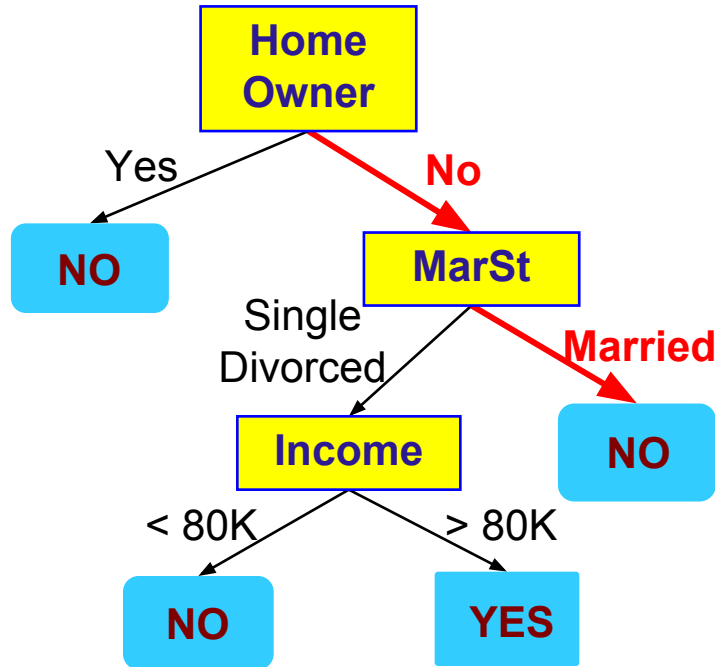


Home Owner	Marital Status	Annual Income	Defaulted Borrower
No	Married	80K	?

Test Data

DECISION TREE

9

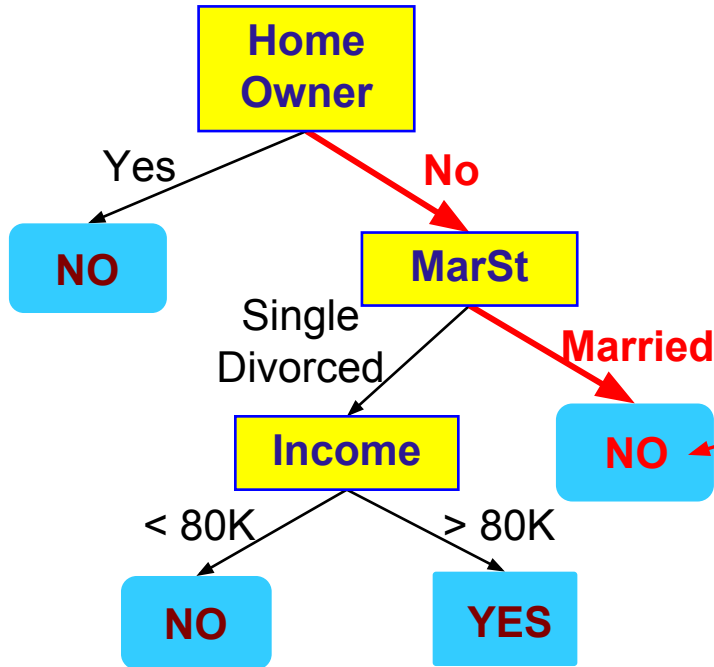


Home Owner	Marital Status	Annual Income	Defaulted Borrower
No	Married	80K	?

Test Data

DECISION TREE

10



Home Owner	Marital Status	Annual Income	Defaulted Borrower
No	Married	80K	?

Test Data

Assign Defaulted to "No"

Exercício prático

11

Para o dataset iris:

Realizar uma classificação com árvore de decisão;

Varie o tamanho dos dados para teste;

Anote os resultados de acurácia;

Compare com resultado do KNN

Exercício para entregar

Use o conjunto de dados censo:

O atributo indica se um indivíduo ganha mais que 50,000 dólares por ano ou não

Separe o dataset em treino e teste (use seed=0), com 75% para treino e 25% para teste

Adicione valores faltantes, trate os atributos categóricos

Treine uma árvore de decisão neste conjunto

Reporte a acurácia

Bonus Visualize a árvore de decisão (tutorial: [link](#))

Conclusão

13

Leitura recomendada:

Capítulo 4 e 5 de Introduction to Data Mining

