Exemplo: Problema *Weather*

(Witten & Frank, 2005)

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Algoritmo Rudimentar (1 Rule – 1R)

- 1R: Aprende uma árvore de decisão de um nível
  - Todas as regras usam somente um atributo
    - Atributo deve ser (ou ser transformado em) categórico
      - **Paradigma simbólico**
- Versão Básica:
  - Um ramo para cada valor do atributo
  - Para cada ramo, atribuir a classe mais freqüente
  - Para cada ramo, calcular a taxa de erro de classificação:
    - proporção de exemplos que não pertencem à classe mais freqüente
  - Escolher o atributo com a menor taxa de erro de classificação

Pseudo-Código para o 1R:

Para cada atributo:
  Para cada valor do atributo gerar uma regra como segue:
    Contar a freqüência de cada classe;
    Encontrar a classe mais freqüente*;
    Formar uma regra que atribui a classe mais freqüente a este atributo-valor;
  Calcular a taxa de erro de classificação das regras;
Escolher as regras com a menor taxa de erro de classificação.

*** Empates na classe mais freqüente podem ser decididos aleatoriamente.**

Exemplo: Problema *Weather*

(Witten & Frank, 2005)

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Attribute	Rules	Errors	Total Errors
Outlook	Sunny → No	2/5	4/14
	Overcast → Yes	0/4	
	Rainy → Yes	2/5	
Temp	Hot → No*	2/4	5/14
	Mild → Yes	2/6	
	Cool → Yes	1/4	
Humidity	High → No	3/7	4/14
	Normal → Yes	1/7	
Windy	False → Yes	2/8	5/14
	True → No*	3/6	

1R seria composto ou das 3 regras para Outlook ou das 2 Regras para Humidity: decisão poderia ser feita, por ex., de acordo com o desempenho em um outro conjunto de dados (dados de teste)

Discussão para o 1R:

- 1R foi descrito por Holte (1993)
  - Contém uma avaliação experimental em 16 bases de dados;
  - Regras simples do 1R não são muito piores do que árvores de decisão mais complexas !
- Interessado em resolver um problema de classificação ou em propor um novo classificador?
  - Experimente o 1R primeiro!
- Implementado no software Weka

Holte, Robert C., Very Simple Classification Rules Perform Well on Most Commonly Used Datasets, Machine Learning 11 (1), pp. 63-90, 1993.

Nota

- 1R pode ser visto como um método de seleção de atributos do tipo **embarcado**
- Mas pode também ser utilizado como **filtro** para outros classificadores que não dispõem de seleção de atributos embarcada
  - p. ex. K-NN

Exercício

- Obter um classificador 1R para os dados:

Febre	Enjôo	Mancha	Dor	Diagnóstico
Sim	Sim	Não	Sim	Não
Não	Sim	Não	Não	Sim
Sim	Sim	Sim	Não	Sim
Sim	Não	Não	Sim	Não
Sim	Não	Sim	Sim	Sim
Não	Não	Sim	Sim	Não

11

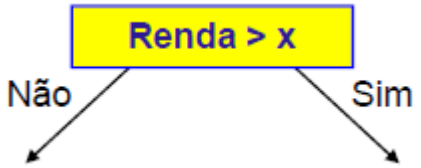
Naïve Bayes

- Naive Bayes é um dos mais simples e bem difundidos classificadores baseados no **Teorema de Bayes**
  - **Paradigma probabilístico**
- Para compreender esse classificador, devemos relembrar alguns conceitos elementares da teoria de probabilidade:
  - Probabilidade Conjunta
  - Probabilidade Condicional
  - Independência Condicional

12

Divisão de Atributos Contínuos

Id	Crédito	Estado Civil	Renda	Deve
1	Sim	Solteiro	125K	Não
2	Não	Casado	100K	Não
3	Não	Solteiro	70K	Não
4	Sim	Casado	120K	Não
5	Não	Divorced	95K	Sim
6	Não	Casado	60K	Não
7	Sim	Divorced	220K	Não
8	Não	Solteiro	85K	Sim
9	Não	Casado	75K	Não
10	Não	Solteiro	90K	Sim



Exemplo

Deve	Não		Não		Não		Sim		Sim		Sim		Não		Não		Não		Não	
Renda																				
Valores ordenados →	60		70		75		85		90		95		100		120		125		220	
Candidatos a pto. de ref. →	55		65		72		80		87		92		97		110		122		172		230	
	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>
Sim	0	3	0	3	0	3	0	3	1	2	2	1	3	0	3	0	3	0	3	0	3	0
Não	0	7	1	6	2	5	3	4	3	4	3	4	3	4	4	3	5	2	6	1	7	0
$Gini_d$	0.420		0.400		0.375		0.343		0.417		0.400		*0.300*		0.343		0.375		0.400		0.420	

* Nota: O exemplo acima assume o uso de desigualdades estritas (< e >) no teste, por isso toma valores candidatos intermediários aos valores do atributo, ao invés desses próprios