

Aprendizado de Máquina

Aula 6

<http://www.ic.uff.br/~bianca/aa/>

Tópicos

1. Introdução – Cap. 1 (16/03)
2. Classificação Indutiva – Cap. 2 (23/03)
3. Árvores de Decisão – Cap. 3 (30/03)
4. Ensembles - Artigo (13/04)
5. Avaliação Experimental – Cap. 5 (20/04)
6. **Aprendizado de Regras – Cap. 10 (27/04)**
7. Redes Neurais – Cap. 4 (04/05)
8. Teoria do Aprendizado – Cap. 7 (11/05)
9. Máquinas de Vetor de Suporte – Artigo (18/05)
10. Aprendizado Bayesiano – Cap. 6 e novo cap. online (25/05)
11. Aprendizado Baseado em Instâncias – Cap. 8 (01/05)
12. Classificação de Textos – Artigo (08/06)
13. Aprendizado Não-Supervisionado – Artigo (15/06)

Aula 6 - 27/04/2010

2

Aprendizado de Regras

- Regras SE-ENTÃO são uma representação padrão de conhecimento que se mostrou útil em sistemas especialistas e outros sistemas de IA.
 - Em lógica proposicional, um conjunto de regras para um conceito é equivalente a DNF.
- Regras são inteligíveis e podem ajudar os usuários a descobrir padrões interessantes nos dados.
- Métodos para induzir regras automaticamente a partir dos dados se mostraram mais precisos do que os baseados em extração de conhecimento de especialistas, para algumas aplicações.
- Métodos de aprendizado de regras foram estendidos para lógica de primeira ordem de tal modo a lidar com representações relacionais.
 - Programação Lógica Indutiva (ILP) permite aprendizado de programas Prolog a partir de exemplos de entrada e saída.
 - Vai além da representação tradicional de vetores de atributos.

Aula 6 - 27/04/2010

3

Abordagens para o Aprendizado de Regras

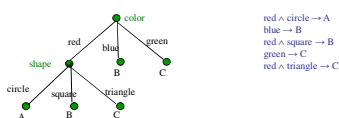
- Transformar árvores de decisão em regras (C4.5)
- Algoritmos de cobertura sequenciais
 - Geral-para-específico (top-down) (CN2, FOIL)
 - Específico-para-geral (bottom-up) (GOLEM, CIGOL)
 - Busca Híbrida (AQ, Chillin, Progol)
- Traduzir redes neurais em regras (TREPAN)

Aula 6 - 27/04/2010

4

Árvores de Decisão para Regras

- Para cada caminho em uma árvore de decisão a partir da raiz até uma folha, crie uma regra com a conjunção de testes ao longo do caminho como um antecedente e a etiqueta da folha como consequente.



Aula 6 - 27/04/2010

5

Pós-Processamento de Regras vindas de Árvores

- Regras resultantes podem conter antecedentes desnecessários que resultam em super-ajuste.
- Regras são podadas através da remoção gulosa de antecedentes ou regras até que o desempenho em dados de treinamento ou de validação não seja significativamente deteriorada.
- Regras resultantes podem levar a conclusões conflitantes para alguns exemplos.
- Ordenar regras pela acurácia de treinamento (ou validação) para criar uma lista de decisão ordenada. A primeira regra na lista que for aplicável é utilizada na classificação.

```

red ^ circle -> A (97% train accuracy)
red ^ big -> B (95% train accuracy)
:
:
Test case: <big, red, circle> assigned to class A
  
```

Aula 6 - 27/04/2010

6

Cobertura Sequencial

- Um conjunto de regras é aprendido uma de cada vez, sempre encontrando uma regra que cubra um grande número de exemplos positivos sem cobrir negativos, e removendo os positivos cobertos.

Let P be the set of positive examples

Until P is empty do:

Learn a rule R that covers a large number of elements of P

but

no negatives.

Add R to the list of rules.

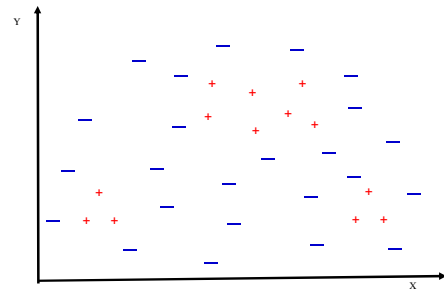
Remove positives covered by R from P

- É um algoritmo guloso que não garante o número mínimo de regras.
- Cobertura mínima é um problema NP-difícil e o algoritmo guloso é um algoritmo de aproximação padrão.
- Métodos para aprender regras individuais variam.

Aula 6 - 27/04/2010

7

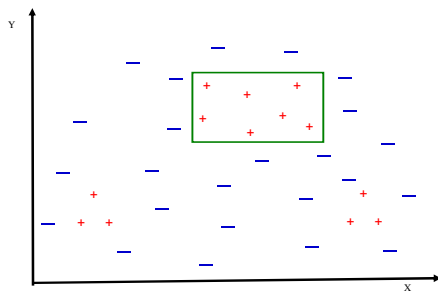
Cobertura Sequencial Gulosa Exemplo



Aula 6 - 27/04/2010

8

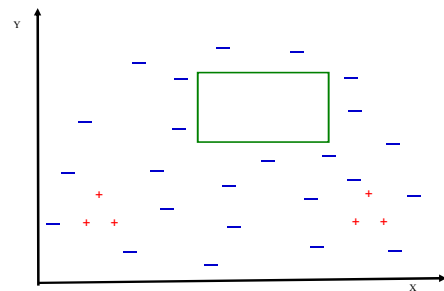
Cobertura Sequencial Gulosa Exemplo



Aula 6 - 27/04/2010

9

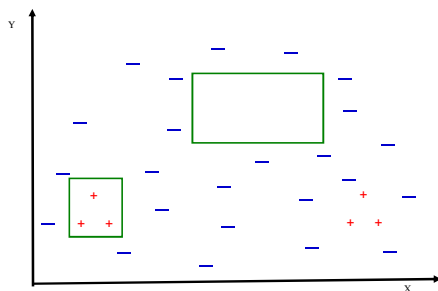
Cobertura Sequencial Gulosa Exemplo



Aula 6 - 27/04/2010

10

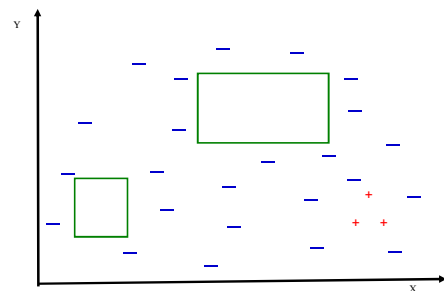
Cobertura Sequencial Gulosa Exemplo



Aula 6 - 27/04/2010

11

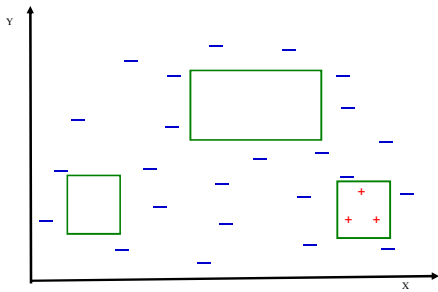
Cobertura Sequencial Gulosa Exemplo



Aula 6 - 27/04/2010

12

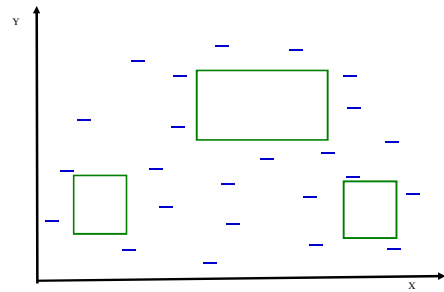
Cobertura Sequencial Gulosa Exemplo



Aula 6 - 27/04/2010

13

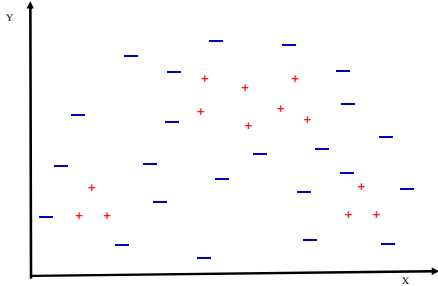
Cobertura Sequencial Gulosa Exemplo



Aula 6 - 27/04/2010

14

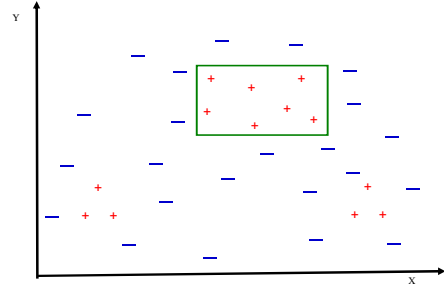
Cobertura Não-Ótima Exemplo



Aula 6 - 27/04/2010

15

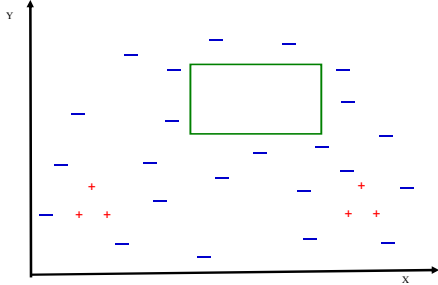
Cobertura Não-Ótima Exemplo



Aula 6 - 27/04/2010

16

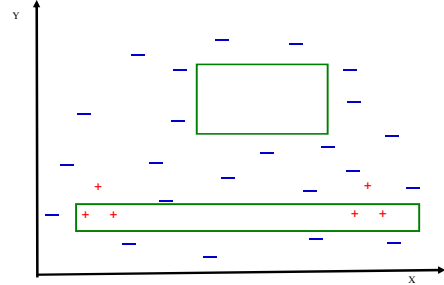
Cobertura Não-Ótima Exemplo



Aula 6 - 27/04/2010

17

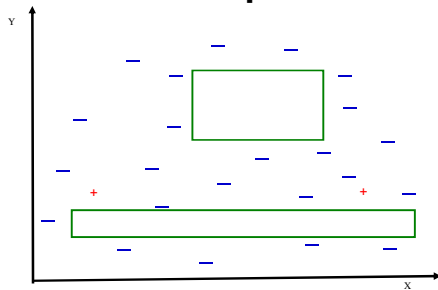
Cobertura Não-Ótima Exemplo



Aula 6 - 27/04/2010

18

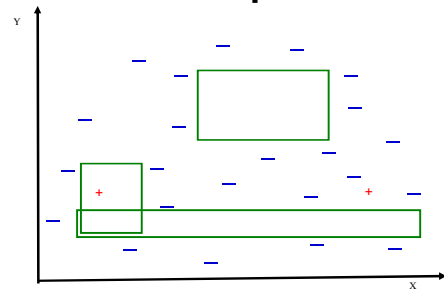
Cobertura Não-Ótima Exemplo



Aula 6 - 27/04/2010

19

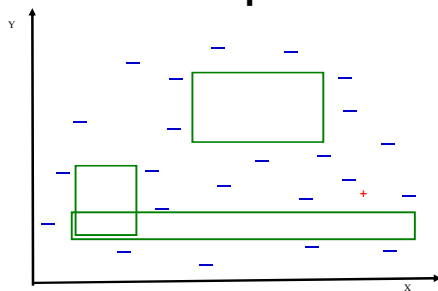
Cobertura Não-Ótima Exemplo



Aula 6 - 27/04/2010

20

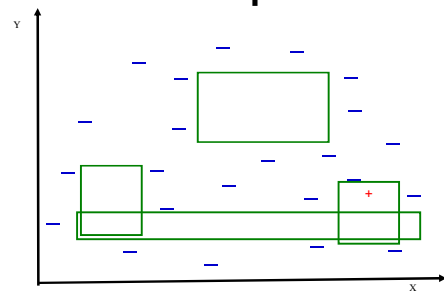
Cobertura Não-Ótima Exemplo



Aula 6 - 27/04/2010

21

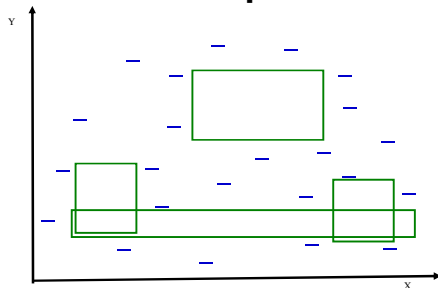
Cobertura Não-Ótima Exemplo



Aula 6 - 27/04/2010

22

Cobertura Não-Ótima Exemplo



Aula 6 - 27/04/2010

23

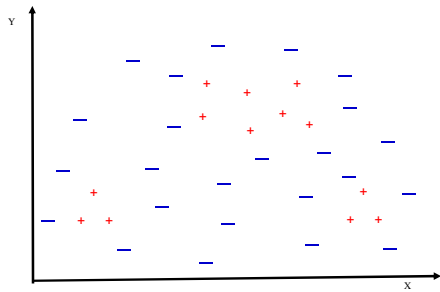
Estratégias para Aprender uma Regra

- Top Down (Geral para Específico):
 - Começar com a regra mais geral (vazia).
 - Repetidamente adicionar restrições ao antecedente usando características que eliminem o máximo de negativos enquanto mantém o maior número de positivos.
 - Parar quando somente positivos estiverem cobertos.
- Bottom Up (Específico para Geral)
 - Começar com a regra mais específica (por exemplo, uma instância aleatória).
 - Repetidamente remover restrições do antecedente para cobrir mais positivos.
 - Parar quando generalização começar a cobrir negativos.

Aula 6 - 27/04/2010

24

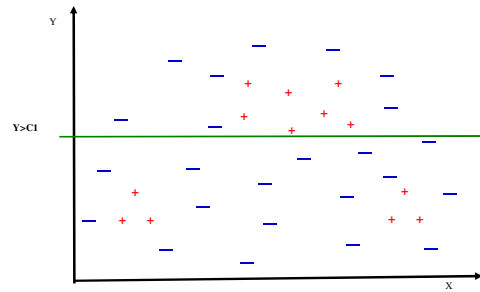
Exemplo de Aprendizado de Regra Top-Down



Aula 6 - 27/04/2010

25

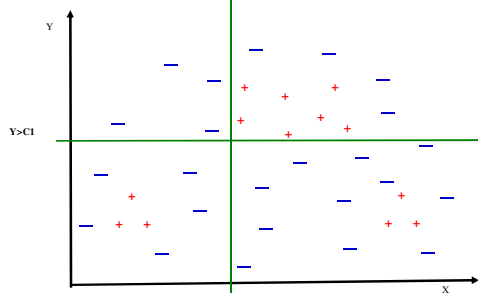
Exemplo de Aprendizado de Regra Top-Down



Aula 6 - 27/04/2010

26

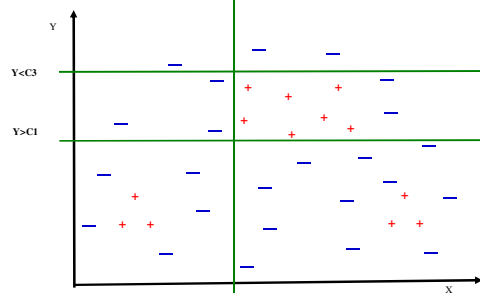
Exemplo de Aprendizado de Regra Top-Down



Aula 6 - 27/04/2010

27

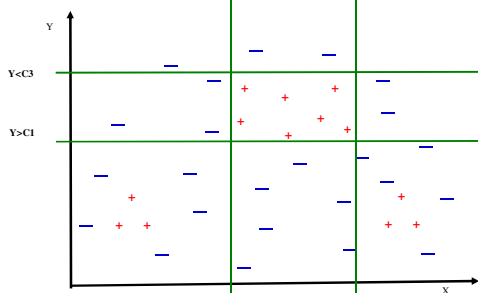
Exemplo de Aprendizado de Regra Top-Down



Aula 6 - 27/04/2010

28

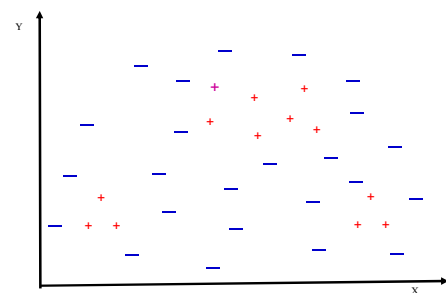
Exemplo de Aprendizado de Regra Top-Down



Aula 6 - 27/04/2010

29

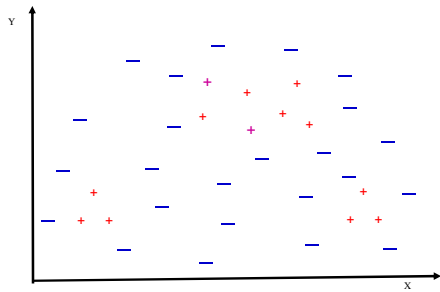
Exemplo de Aprendizado de Regra Bottom-Up



Aula 6 - 27/04/2010

30

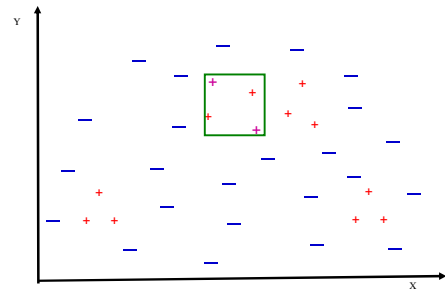
Exemplo de Aprendizado de Regra Bottom-Up



Aula 6 - 27/04/2010

31

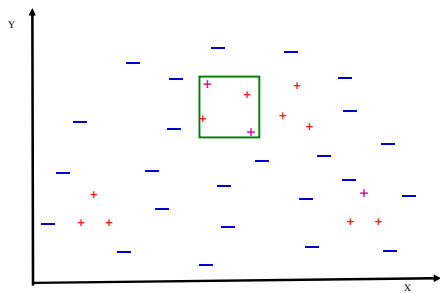
Exemplo de Aprendizado de Regra Bottom-Up



Aula 6 - 27/04/2010

32

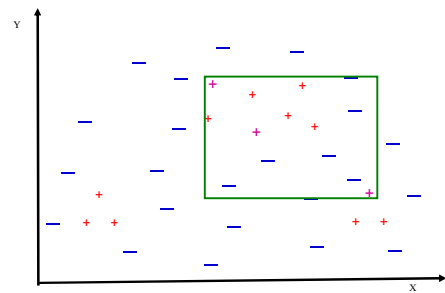
Exemplo de Aprendizado de Regra Bottom-Up



Aula 6 - 27/04/2010

33

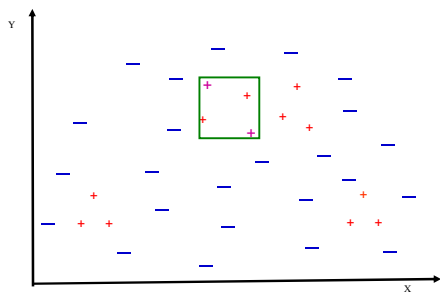
Exemplo de Aprendizado de Regra Bottom-Up



Aula 6 - 27/04/2010

34

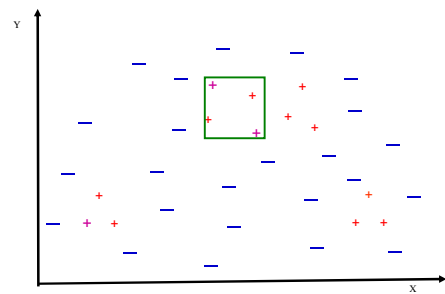
Exemplo de Aprendizado de Regra Bottom-Up



Aula 6 - 27/04/2010

35

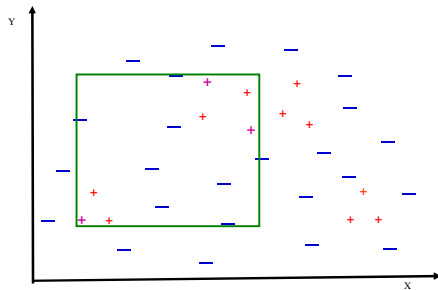
Exemplo de Aprendizado de Regra Bottom-Up



Aula 6 - 27/04/2010

36

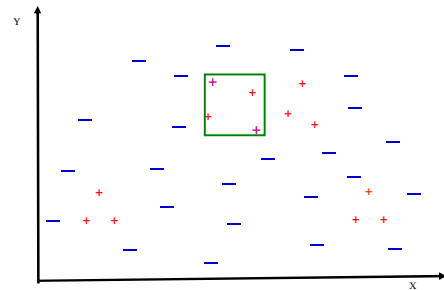
Exemplo de Aprendizado de Regra Bottom-Up



Aula 6 - 27/04/2010

37

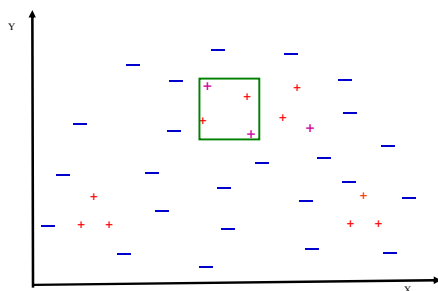
Exemplo de Aprendizado de Regra Bottom-Up



Aula 6 - 27/04/2010

38

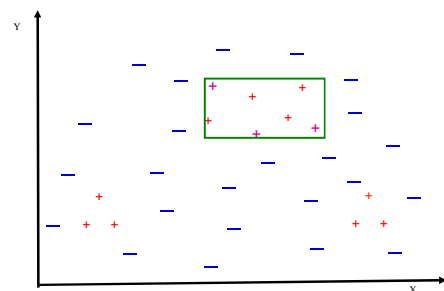
Exemplo de Aprendizado de Regra Bottom-Up



Aula 6 - 27/04/2010

39

Exemplo de Aprendizado de Regra Bottom-Up



Aula 6 - 27/04/2010

40

Aprendendo uma Regra com FOIL

- Abordagem top-down originalmente aplicada a lógica de primeira ordem (Quinlan, 1990).
- Algoritmo básico para exemplos com atributos discretos:
 - Let $A = \{\}$ (set of rule antecedents)
 - Let N be the set of negative examples
 - Let P be the current set of uncovered positive examples
 - Until N is empty do
 - For every feature-value pair (literal) $(F_i = V_{ij})$ calculate $\text{Gain}(F_i = V_{ij}, P, N)$
 - Pick literal, L , with highest gain.
 - Add L to A .
 - Remove from N any examples that do not satisfy L .
 - Remove from P any examples that do not satisfy L .
 - Return the rule: $A_1 \wedge A_2 \wedge \dots \wedge A_n \rightarrow \text{Positive}$

Aula 6 - 27/04/2010

41

Métrica de Ganho do FOIL

- Queremos atingir dois objetivos:
 - Diminuir a cobertura de exemplos negativos
 - Medir o aumento da porcentagem de positivos cobertos quando o literal é adicionado à regra..
 - Manter cobertura do maior número de positivos possível.
 - Contar número de positivos cobertos.

Define $\text{Gain}(L, P, N)$

Let p be the subset of examples in P that satisfy L .
 Let n be the subset of examples in N that satisfy L .
 Return: $|p| * [\log_2(|p|/(|p|+|n|))] - \log_2(|P|/(|P|+|N|))$

Aula 6 - 27/04/2010

42

Exemplo - FOIL

Example	Size	Color	Shape	Category
1	small	red	circle	positive
2	big	red	circle	positive
3	small	red	triangle	negative
4	big	blue	circle	negative
5	medium	red	circle	negative

Aula 6 - 27/04/2010

43

Propositional FOIL Trace

New Disjunct:
 SIZE=BIG Gain: 0.322
 SIZE=MEDIUM Gain: 0.000
 SIZE=SMALL Gain: 0.322
 COLOR=BLUE Gain: 0.000
 COLOR=RED Gain: 0.644
 COLOR=GREEN Gain: 0.000
 SHAPE=SQUARE Gain: 0.000
 SHAPE=TRIANGLE Gain: 0.000
 SHAPE=CIRCLE Gain: 0.644
 Best feature: COLOR=RED

SIZE=BIG Gain: 1.000
 SIZE=MEDIUM Gain: 0.000
 SIZE=SMALL Gain: 0.000
 SHAPE=SQUARE Gain: 0.000
 SHAPE=TRIANGLE Gain: 0.000
 SHAPE=CIRCLE Gain: 0.830
 Best feature: SIZE=BIG
 Learned Disjunct: COLOR=RED & SIZE=BIG

Aula 6 - 27/04/2010

44

Propositional FOIL Trace

New Disjunct:
 SIZE=BIG Gain: 0.000
 SIZE=MEDIUM Gain: 0.000
 SIZE=SMALL Gain: 1.000
 COLOR=BLUE Gain: 0.000
 COLOR=RED Gain: 0.415
 COLOR=GREEN Gain: 0.000
 SHAPE=SQUARE Gain: 0.000
 SHAPE=TRIANGLE Gain: 0.000
 SHAPE=CIRCLE Gain: 0.415
 Best feature: SIZE=SMALL

COLOR=BLUE Gain: 0.000
 COLOR=RED Gain: 0.000
 COLOR=GREEN Gain: 0.000
 SHAPE=SQUARE Gain: 0.000
 SHAPE=TRIANGLE Gain: 0.000
 SHAPE=CIRCLE Gain: 1.000
 Best feature: SHAPE=CIRCLE
 Learned Disjunct: SIZE=SMALL & SHAPE=CIRCLE
 Final Definition: COLOR=RED & SIZE=BIG ∨ SIZE=SMALL & SHAPE=CIRCLE

Aula 6 - 27/04/2010

45

Poda no FOIL

- Pré-poda baseada em MDL.
- Pós-poda para eliminar complexidade devida ao fato do algoritmo ser guloso.
 - For each rule, R
 - For each antecedent, A , of rule
 - If deleting A from R does not cause negatives to become covered then delete A
 - For each rule, R
 - If deleting R does not uncover any positives (since they are redundantly covered by other rules) then delete R

Aula 6 - 27/04/2010

46

Questões do Aprendizado de Regras

- O que é melhor, árvores ou regras?
 - Árvores compartilham estrutura entre disjunções.
 - Regras permitem atributos completamente diferentes para as disjunções.
 - Mapear regras em árvores pode criar árvores exponencialmente maiores.

 $A \wedge B \rightarrow P$
 $C \wedge D \rightarrow P$

E se adicionarmos a regra:
 $E \wedge F \rightarrow P$
 ??

Aula 6 - 27/04/2010

47

Questões do Aprendizado de Regras

- O que é melhor: aprendizado top-down ou bottom-up?
 - O bottom-up é mais sujeito a ruído, pela escolha das sementes.
 - O top-down perde tempo quando existem muitas características que não ocorrem para exemplos positivos (exemplo: classificação de textos).

Aula 6 - 27/04/2010

48

Aprendizado de Regras vs. Aquisição de Conhecimento

- Um experimento com um método de aprendizado de regras (AQ) feito por Michalski (1980) comparou os resultados com aquisição de conhecimento (aquisição de regras através de entrevistas com especialistas).
- Regras adquiridas dos especialistas
 - Pesos associados com cada atributo da regra
 - Método para somar evidências
 - Não usavam disjunções explícitas
- Dados para indução:
 - Exemplos de 15 doenças da soja descritas usando 35 atributos discretos, 630 exemplos no total.
 - 290 "melhores" (diversos) exemplos selecionados para treinamento. Resto usado para teste.
 - O que há de errado com essa metodologia?

Aula 6 - 27/04/2010

49

Resultados Experimentais

- Tempo de construção das regras:
 - Especialistas: 45 horas de consulta
 - AQ11: 4.5 minutos treinamento em um IBM 360/75
 - O que isso não leva em conta?
- Acurácia de teste:

	1 st choice correct	Some choice correct	Number of diagnoses
AQ11	97.6%	100.0%	2.64
Manual KE	71.8%	96.9%	2.90

Aula 6 - 27/04/2010

50