

Aprendizagem de Máquina

Prof. Júlio Cesar Nievola
PPGIA - PUCPR

Introdução

- Justificativa
 - Recente progresso em algoritmos e teoria
 - Disponibilidade crescente de dados online
 - Poder computacional disponível
 - Crescimento da indústria
- Alguns nichos - aprendizagem de máquina
 - Data mining
 - Softwares aplicativos que não podem ser programados manualmente
 - Programas auto-personalizados

Disciplinas Relevantes

- Inteligência Artificial
- Métodos Bayesianos
- Teoria da Complexidade Computacional
- Teoria de Controle
- Teoria da Informação
- Filosofia
- Psicologia e neurobiologia
- Estatística
- ...

Definição de Problemas de Aprendizagem Bem Especificados

Diz-se que um programa computacional aprende a partir da experiência E em relação a uma classe de tarefas T com medida de desempenho P se seu desempenho nas tarefas T , medida por P , melhora com a experiência E .

Exemplos

- Jogo de Damas
 - Tarefa T: jogar damas
 - Medida de Desempenho P: percentagem de jogos vencidos contra oponentes
 - Experiência de treinamento E: praticar jogos contra si mesmo
- Reconhecimento de Escrita Manual
 - Tarefa T: reconhecer e classificar palavras manuscritas contidas em imagens
 - Medida de Desempenho P: percentual de palavras corretamente classificadas
 - Experiência de treinamento E: base de dados com palavras manuscritas e respectivas classificações

Decisões

- Quais algoritmos (e quando) podem aproximar adequadamente funções?
- Como o número de exemplos de treinamento influencia a precisão?
- Como dados ruidosos influenciam a precisão?
- Quais são os limites teóricos da capacidade de aprendizagem?
- Como o conhecimento prévio pode auxiliar?

Hipótese da Aprendizagem Indutiva

Qualquer hipótese obtida que aproxima adequadamente a função objetivo sobre um conjunto de exemplos de treinamento suficientemente grande também aproximará adequadamente a função objetivo sobre outros exemplos não observados

Aprendizagem

- Aprendizagem Supervisionada
 - Problema de classificação
 - Necessita de um “professor”
- Aprendizagem Não-Supervisionada
 - Problema de agrupamento
 - Não há saída “correta”
- Aprendizagem com Reforço
 - Não se sabe qual a saída correta

Elementos da Aprendizagem

- Componentes da entrada:
 - Conceitos: tipos de coisas que podem ser aprendidas
 - Instâncias: exemplos individuais e independentes de um conceito
 - Atributos: aspectos mensuráveis de uma instância
- Tipos de atributos: binários, categóricos, numéricos

O que é um conceito?

- Tipos de aprendizagem:
 - Classificação: Previsão de uma classe discreta
 - Associação: Detectar associações entre atributos
 - Agrupamento: Agrupar instâncias similares
 - Previsão numérica: Prever uma quantidade numérica
- Conceito: coisa a ser aprendida
- Descrição do conceito: saída do esquema de aprendizagem

Características de Dados

- Corretos
- Faltantes
 - Desconhecido, não armazenado, irrelevante
- Imprecisos
 - Erro tipográfico, erro de obtenção ou erro deliberado
- Incorreto
 - Atípico (“outlier”)

Representação de Padrões Estruturais

- Diferentes formas de representar padrões estruturais (regras, árvores)
- É a representação do conhecimento
- Determina o método de inferência
- Conhecer a saída é o ponto chave para compreender o método de aprendizagem
- Diferentes tipos de saída para diferentes problemas de aprendizagem

Tabela de Decisão

- Usa o mesmo formato como entrada
- Problema principal: selecionar os atributos corretos

Outlook	Humidity	Play
Sunny	High	No
Sunny	Normal	Yes
Overcast	High	Yes
Overcast	Normal	Yes
Rainy	High	No
Rainy	Normal	No

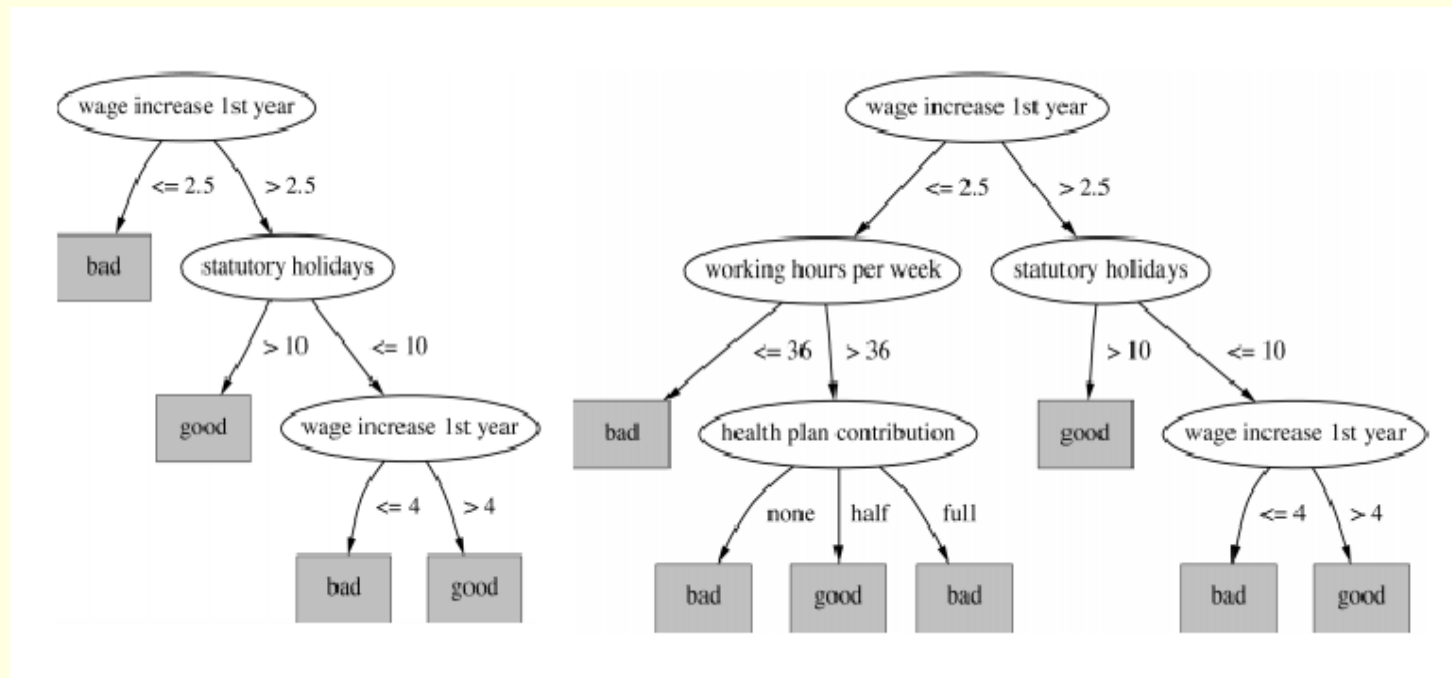
Árvore de Decisão

- Abordagem “dividir para conquistar”
- Nós representam o teste de um atributo
- Normalmente o valor do atributo é comparado a uma constante
- Folhas indicam a classificação ou distribuição de probabilidade de instâncias

Atributos nominais e numéricos

- Atributo nominal: número de filhos em geral igual ao número de valores (não é testado mais de uma vez)
 - Outra situação: divisão em dois subconjuntos
- Atributo numérico: testa se o valor é maior ou menor que uma constante (pode ser testado várias vezes)
 - Inteiro: maior que, menor que, igual
 - Real: abaixo, acima, entre

Árvores de Decisão e Poda



Regras de Classificação

- Antecedente (pré-condição): série de testes
- Testes são geralmente ligados de forma conjuntiva
- Conseqüente (conclusão): classe, conjunto de classes ou distribuição de probabilidades assinalada pela regra
- Regras individuais são geralmente ligadas de forma disjuntiva
 - Conflitos surgem se houverem conclusões diferentes

Extração de Regras Completas e Corretas

A complete and correct rule set

```
If tear production rate = reduced then recommendation = none
If age = young and astigmatic = no and tear production rate = normal
  then recommendation = soft
If age = pre-presbyopic and astigmatic = no and
  tear production rate = normal then recommendation = soft
If age = presbyopic and spectacle prescription = myope and
  astigmatic = no then recommendation = none
If spectacle prescription = hypermetrope and astigmatic = no and
  tear production rate = normal then recommendation = soft
If spectacle prescription = myope and astigmatic = yes and
  tear production rate = normal then recommendation = hard
If age young and astigmatic = yes and tear production rate = normal
  then recommendation = hard
If age = pre-presbyopic and spectacle prescription = hypermetrope
  and astigmatic = yes then recommendation = none
If age = presbyopic and spectacle prescription = hypermetrope and
  astigmatic = yes then recommendation = none
```

Conversão árvores => regras

- Uma regra para cada folha:
 - Antecedente contém uma condição para cada nó no caminho entre a raiz e a folha
 - Conseqüente é a classe indicada pela folha
- Produz regras que não são ambíguas (independente da ordem de execução)
- Regras são desnecessariamente complexas

Conjuntos de Regras

- Duas formas de executar um conjunto de regras:
 - Lista de Decisão: conjunto ordenado de regras (a ordem é importante)
 - Conjunto não ordenado de regras: regras podem se sobrepor e levar a conclusões diferentes para uma mesma instância
- O que fazer quando há conflito de regras?
- O que fazer quando nenhuma regra é aplicável?

Regras de Associação

- Podem prever qualquer atributo e combinações de atributos
- Não são projetadas para ser usadas juntas como um conjunto
- Problema: número muito elevado de associações possíveis
- A saída deve ser limitada para mostrar somente as associações com melhor previsão, ou seja, aquelas com elevado suporte e confiança

Suporte e Confidência de uma Regra

- Suporte: número de instâncias previstas corretamente
- Confiança: número de previsões corretas em relação ao total de instâncias às quais a regra se aplica
- Exemplo:
 - IF Sunny THEN Yes
 - Suporte: 1/6
 - Confiança: 1/2

Outlook	Humidity	Play
Sunny	High	No
Sunny	Normal	Yes
Overcast	High	Yes
Overcast	Normal	Yes
Rainy	High	No
Rainy	Normal	No

Regras com Exceções

```
default: Iris-setosa
except if petal-length  $\geq$  2.45 and petal-length  $<$  5.355
      and petal-width  $<$  1.75
  then Iris-versicolor
      except if petal-length  $\geq$  4.95 and petal-width  $<$  1.55
        then Iris-virginica
        else if sepal-length  $<$  4.95 and sepal-width  $\geq$  2.45
          then Iris-virginica
  else if petal-length  $\geq$  3.35
    then Iris-virginica
        except if petal-length  $<$  4.85 and sepal-length  $<$  5.95
          then Iris-versicolor
```

Vantagens das Regras com Exceções

- Pessoas normalmente pensam em termos de exceções
- Regras podem ser atualizadas de forma incremental
- Equivalente a IF...THEN...ELSE...
- Vantagem psicológica:
 - Default e testes iniciais ocorrem com mais frequência
 - Exceções refletem casos especiais

Árvores para Previsão Numérica

- *Regressão*: calcula uma expressão que prevê a quantidade numérica
- *Árvore de regressão*: cada folha prevê uma quantidade numérica (valor previsto é o valor médio das instâncias de treinamento que alcançam a folha)
- *Árvore de Modelo*: árvore de regressão com modelos de regressão linear nos nós folha

Base CPU

Previsão do Desempenho de CPU

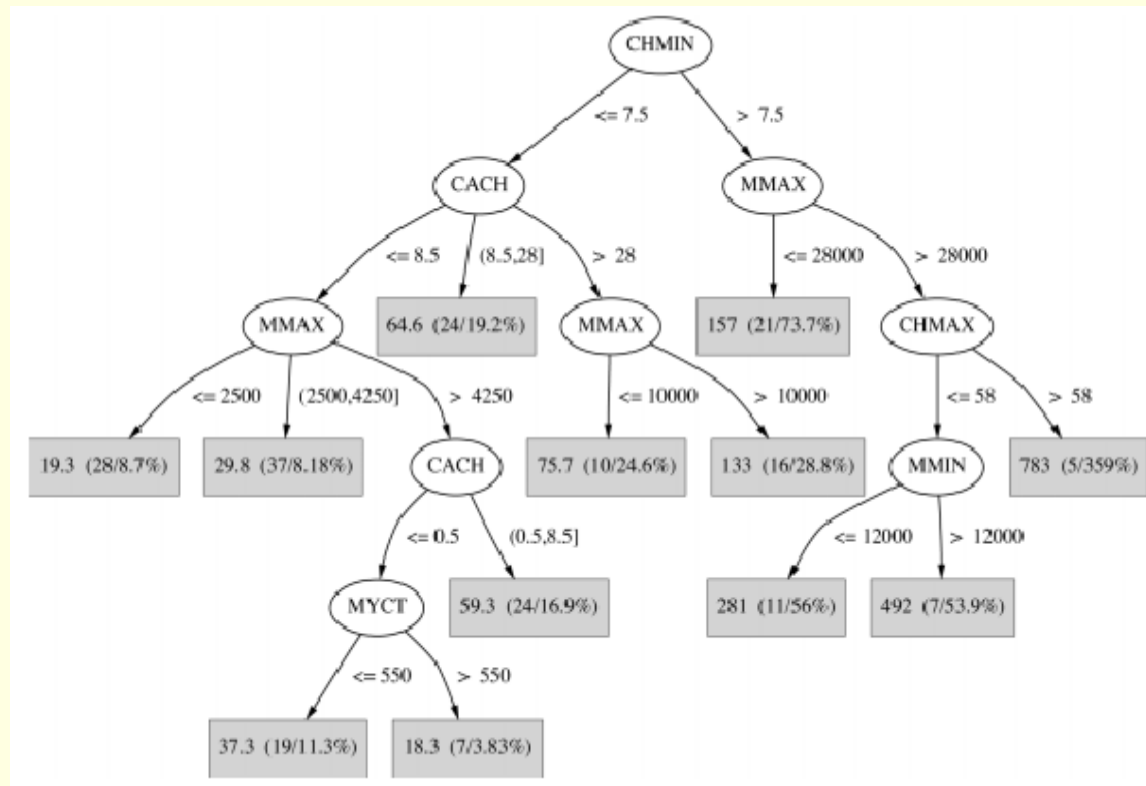
Examples: 209 different computer configurations

	Cycle time (ns)	Main memory (Kb)		Cache (Kb)	Channels		Performance
	MYCT	MMIN	MMAX	CACH	CHMIN	CHMAX	PRP
1	125	256	6000	256	16	128	198
2	29	8000	32000	32	8	32	269
...							
208	480	512	8000	32	0	0	67
209	480	1000	4000	0	0	0	45

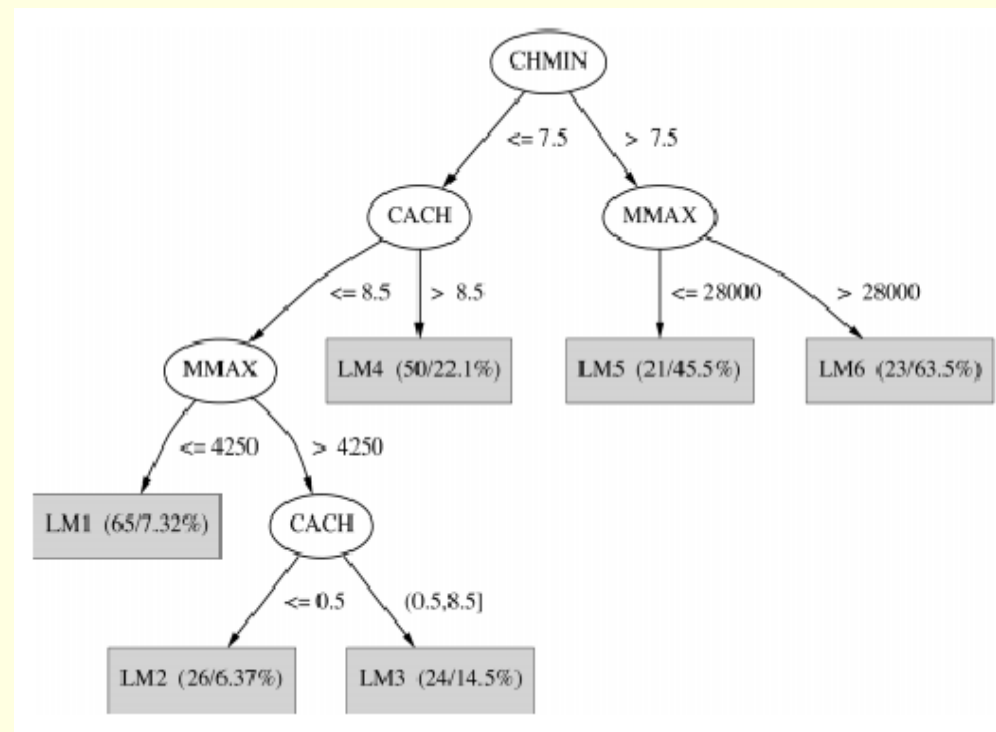
Regressão Linear para Dados da Base CPU

```
PRP =  
  - 56.1  
    + 0.049 MYCT  
    + 0.015 MMIN  
    + 0.006 MMAX  
    + 0.630 CACH  
    - 0.270 CHMIN  
    + 1.46  CHMAX
```

Árvore de Regressão para a Base CPU



Árvore de Modelo para a Base CPU



Representação Baseada em Instâncias

- Buscam-se entre as instâncias de treinamento aquela que esteja mais próxima da nova instância
- As instâncias em si representam o conhecimento
- Uma função de similaridade define o que será “aprendido”
- Uso do método k-vizinho mais próximo