

Inteligência Artificial – ACH2016

Aula 18 – Árvores de Decisão e Florestas Aleatórias

Norton Trevisan Roman
(norton@usp.br)

13 de maio de 2019

Árvores de Decisão

- Uma das formas mais simples de aprendizado

Árvores de Decisão

- Uma das formas mais simples de aprendizado
- Entrada:

Árvores de Decisão

- Uma das formas mais simples de aprendizado
- Entrada:
 - Objeto ou situação descritos por um conjunto de atributos

Árvores de Decisão

- Uma das formas mais simples de aprendizado
- Entrada:
 - Objeto ou situação descritos por um conjunto de atributos
 - Atributos podem ser discretos ou contínuos

Árvores de Decisão

- Uma das formas mais simples de aprendizado
- Entrada:
 - Objeto ou situação descritos por um conjunto de atributos
 - Atributos podem ser discretos ou contínuos
- Saída:

Árvores de Decisão

- Uma das formas mais simples de aprendizado
- Entrada:
 - Objeto ou situação descritos por um conjunto de atributos
 - Atributos podem ser discretos ou contínuos
- Saída:
 - Uma decisão – a saída prevista para a entrada (discreta ou contínua)

Árvores de Decisão

- Uma das formas mais simples de aprendizado
- Entrada:
 - Objeto ou situação descritos por um conjunto de atributos
 - Atributos podem ser discretos ou contínuos
- Saída:
 - Uma decisão – a saída prevista para a entrada (discreta ou contínua)
 - Obtida por meio de uma sequência de testes

Árvores de Decisão

- Representação bastante natural

Árvores de Decisão

- Representação bastante natural
 - Sequência de regras do tipos “se ... então”

Árvores de Decisão

- Representação bastante natural
 - Sequência de regras do tipos “se ... então”
 - Fáceis de interpretar

Árvores de Decisão

- Representação bastante natural
 - Sequência de regras do tipos “se ... então”
 - Fáceis de interpretar

Tipos:

Árvores de Decisão

- Representação bastante natural
 - Sequência de regras do tipos “se ... então”
 - Fáceis de interpretar

Tipos:

- Quando usadas para aprender uma função de valores discretos:

Árvores de Decisão

- Representação bastante natural
 - Sequência de regras do tipos “se ... então”
 - Fáceis de interpretar

Tipos:

- Quando usadas para aprender uma função de valores discretos:
 - Árvore de Classificação

Árvores de Decisão

- Representação bastante natural
 - Sequência de regras do tipos “se ... então”
 - Fáceis de interpretar

Tipos:

- Quando usadas para aprender uma função de valores discretos:
 - Árvore de Classificação
- Quando usadas para aprender uma função de valores contínuos:

Árvores de Decisão

- Representação bastante natural
 - Sequência de regras do tipos “se ... então”
 - Fáceis de interpretar

Tipos:

- Quando usadas para aprender uma função de valores discretos:
 - Árvore de Classificação
- Quando usadas para aprender uma função de valores contínuos:
 - Árvore de Regressão

Árvores de Decisão

Características:

Árvores de Decisão

Características:

- Cada nó interno representa uma função teste $f_m(x)$ para o valor de um dos atributos de entrada x

Árvores de Decisão

Características:

- Cada nó interno representa uma função teste $f_m(x)$ para o valor de um dos atributos de entrada x
- Em vez de usar todo o conjunto de atributos, diferentes subconjuntos são usados em diferentes níveis da árvore

Árvores de Decisão

Características:

- Cada nó interno representa uma função teste $f_m(x)$ para o valor de um dos atributos de entrada x
 - Em vez de usar todo o conjunto de atributos, diferentes subconjuntos são usados em diferentes níveis da árvore
- Os galhos são rotulados com cada valor possível do teste feito no nó que os origina

Árvores de Decisão

Características:

- Cada nó interno representa uma função teste $f_m(x)$ para o valor de um dos atributos de entrada x
 - Em vez de usar todo o conjunto de atributos, diferentes subconjuntos são usados em diferentes níveis da árvore
- Os galhos são rotulados com cada valor possível do teste feito no nó que os origina
 - Os valores são mutuamente exclusivos e exaustivos

Árvores de Decisão

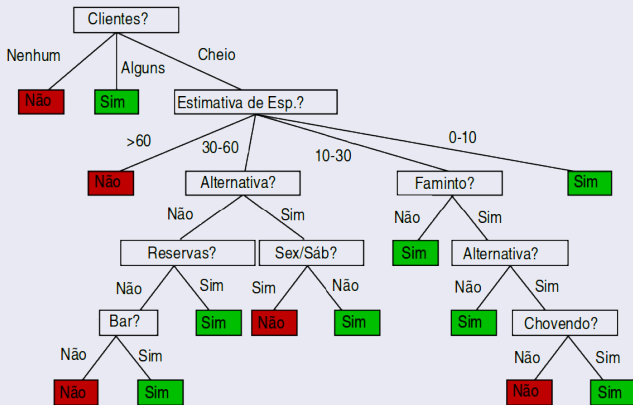
Características:

- Cada nó interno representa uma função teste $f_m(x)$ para o valor de um dos atributos de entrada x
 - Em vez de usar todo o conjunto de atributos, diferentes subconjuntos são usados em diferentes níveis da árvore
- Os galhos são rotulados com cada valor possível do teste feito no nó que os origina
 - Os valores são mutuamente exclusivos e exaustivos
- Cada folha especifica o valor que deve ser retornado se a folha for atingida

Árvores de Decisão

Exemplo (classificação booleana):

Esperaremos para comer no restaurante no sábado?

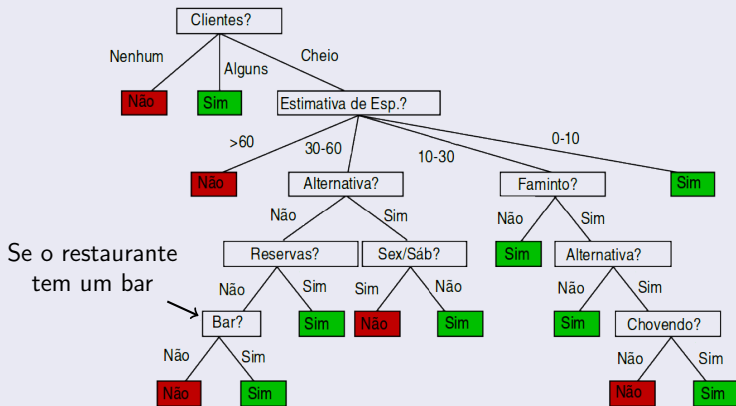


Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Árvores de Decisão

Exemplo (classificação booleana):

Esperaremos para comer no restaurante no sábado?

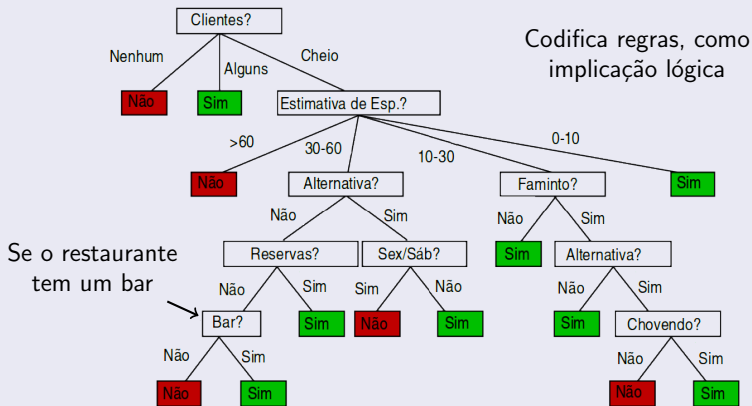


Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Árvores de Decisão

Exemplo (classificação booleana):

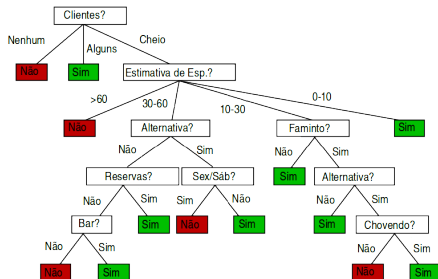
Esperaremos para comer no restaurante no sábado?



Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Árvores de Decisão

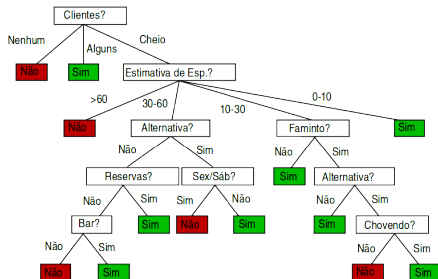
- Cada nó representa um atributo a ser testado



Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Árvores de Decisão

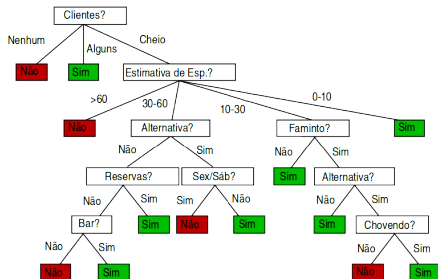
- Cada nó representa um atributo a ser testado
- Modelo univariado



Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Árvores de Decisão

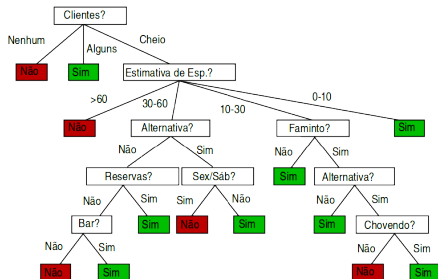
- Cada nó representa um atributo a ser testado
- Modelo univariado
- Os exemplos são processados a partir da raiz seguindo até uma folha



Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Árvores de Decisão

- Cada nó representa um atributo a ser testado
- Modelo univariado
- Os exemplos são processados a partir da raiz seguindo até uma folha



Fonte: Adaptado de slides de AIMA. Russell & Norvig.

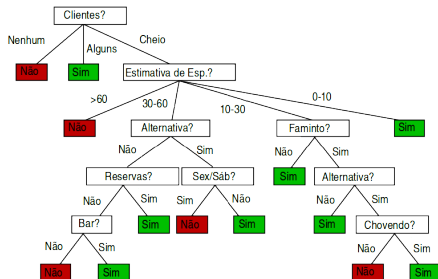
- Em geral, após o primeiro atributo dividir os exemplos, cada saída é um novo problema

Árvores de Decisão

- Cada nó representa um atributo a ser testado

- Modelo univariado

- Os exemplos são processados a partir da raiz seguindo até uma folha



Fonte: Adaptado de slides de AIMA. Russell & Norvig.

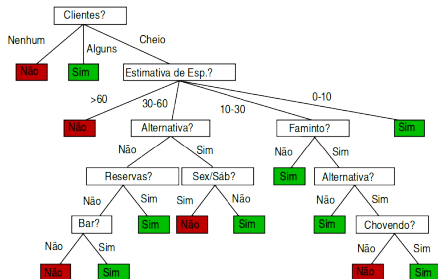
- Em geral, após o primeiro atributo dividir os exemplos, cada saída é um novo problema
 - Com menos exemplos

Árvores de Decisão

- Cada nó representa um atributo a ser testado

- Modelo univariado

- Os exemplos são processados a partir da raiz seguindo até uma folha



Fonte: Adaptado de slides de AIMA. Russell & Norvig.

- Em geral, após o primeiro atributo dividir os exemplos, cada saída é um novo problema
 - Com menos exemplos
 - Com um atributo a menos

Árvores de Decisão

Expressividade:

- Podem expressar qualquer função booleana dos atributos de entrada

Árvores de Decisão

Expressividade:

- Podem expressar qualquer função booleana dos atributos de entrada
- São totalmente expressivas dentro da classe de linguagens proposicionais

Árvores de Decisão

Expressividade:

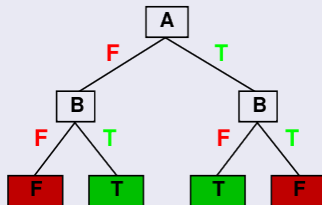
- Podem expressar qualquer função booleana dos atributos de entrada
- São totalmente expressivas dentro da classe de linguagens proposicionais
- Para funções booleanas, uma linha da tabela verdade é um caminho até uma folha

Árvores de Decisão

Expressividade:

- Podem expressar qualquer função booleana dos atributos de entrada
- São totalmente expressivas dentro da classe de linguagens proposicionais
- Para funções booleanas, uma linha da tabela verdade é um caminho até uma folha

| A | B | A xor B |
|---|---|---------|
| F | F | F |
| F | T | T |
| T | F | T |
| T | T | F |



Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Árvores de Decisão

Variações:

Árvores de Decisão

Variações:

- Os testes podem ser:

Árvores de Decisão

Variações:

- Os testes podem ser:
 - Univariados: Os nós testam um único atributo

Variações:

- Os testes podem ser:
 - Univariados: Os nós testam um único atributo
 - Multivariados: Múltiplos atributos testados

Variações:

- Os testes podem ser:
 - Univariados: Os nós testam um único atributo
 - Multivariados: Múltiplos atributos testados
- Testes podem ter:

Variações:

- Os testes podem ser:
 - Univariados: Os nós testam um único atributo
 - Multivariados: Múltiplos atributos testados
- Testes podem ter:
 - Duas saídas possíveis: Binárias

Variações:

- Os testes podem ser:
 - Univariados: Os nós testam um único atributo
 - Multivariados: Múltiplos atributos testados
- Testes podem ter:
 - Duas saídas possíveis: Binárias
 - Mais que duas saídas

Árvores de Decisão

Variações:

- Os testes podem ser:
 - Univariados: Os nós testam um único atributo
 - Multivariados: Múltiplos atributos testados
- Testes podem ter:
 - Duas saídas possíveis: Binárias
 - Mais que duas saídas
- Atributos podem ser categóricos ou numéricos

Construindo Árvores de Decisão

Árvores Booleanas

Construindo Árvores de Decisão

Árvores Booleanas

- Um exemplo de treino para uma árvore de decisão booleana consiste de um arranjo de atributos de entrada, e um único valor booleano de saída

Construindo Árvores de Decisão

Árvores Booleanas

- Um exemplo de treino para uma árvore de decisão booleana consiste de um arranjo de atributos de entrada, e um único valor booleano de saída
- Cada exemplo é classificados como positivo (verdadeiro) ou negativo (falso)

Construindo Árvore de Decisão

Árvores Booleanas

- Um exemplo de treino para uma árvore de decisão booleana consiste de um arranjo de atributos de entrada, e um único valor booleano de saída
- Cada exemplo é classificados como positivo (verdadeiro) ou negativo (falso)
- O conjunto completo de exemplos é o conjunto de treino

Construindo Árvores de Decisão

Considere o exemplo (restaurante):

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x1 | S | N | N | S | Alg | \$\$\$ | N | S | Fr | 0-10 | S |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x3 | N | S | N | N | Alg | \$ | N | N | Bu | 0-10 | S |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x6 | N | S | N | S | Alg | \$\$ | S | S | It | 0-10 | S |
| x7 | N | S | N | N | Nen | \$ | S | N | Bu | 0-10 | N |
| x8 | N | N | N | S | Alg | \$\$ | S | S | Ta | 0-10 | S |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x11 | N | N | N | N | Nen | \$ | N | N | Ta | 0-10 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |

Ex – exemplo

Alt – há alternativa

Bar – tem bar

Sex – abre sex/sab

Fam – estou

faminto

Cli – nº de clientes

Pre – preço

Chu – chovendo

Res – reserva

Tip – tipo

Est – estimativa de
espera

Construindo Árvores de Decisão

Considere o exemplo (restaurante):

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x1 | S | N | N | S | Alg | \$\$\$ | N | S | Fr | 0-10 | S |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x3 | N | S | N | N | Alg | \$ | N | N | Bu | 0-10 | S |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x6 | N | S | N | S | Alg | \$\$ | S | S | It | 0-10 | S |
| x7 | N | S | N | N | Nen | \$ | S | N | Bu | 0-10 | N |
| x8 | N | N | N | S | Alg | \$\$ | S | S | Ta | 0-10 | S |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x11 | N | N | N | N | Nen | \$ | N | N | Ta | 0-10 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |

Ex – exemplo

Alt – há alternativa

Bar – tem bar

Sex – abre sex/sab

Fam – estou

faminto

Cli – nº de clientes

Pre – preço

Chu – chovendo

Res – reserva

Tip – tipo

Est – estimativa de
espera

Existem muitas árvores que codificam esses exemplos sem erro.

Construindo Árvores de Decisão

- Tentamos encontrar a menor árvore consistente com os exemplos

Construindo Árvores de Decisão

- Tentamos encontrar a menor árvore consistente com os exemplos
 - Navalha de Ockham (escolha a hipótese mais simples)

Construindo Árvores de Decisão

- Tentamos encontrar a menor árvore consistente com os exemplos
 - Navalha de Ockham (escolha a hipótese mais simples)
 - NP-completo (Quinlan 1986, Hyafil and Rivest 1976)

Construindo Árvores de Decisão

- Tentamos encontrar a menor árvore consistente com os exemplos
 - Navalha de Ockham (escolha a hipótese mais simples)
 - NP-completo (Quinlan 1986, Hyafil and Rivest 1976)

Heurística:

- (Recursivamente) Escolha o atributo mais significativo como raiz da (sub)árvore

Construindo Árvores de Decisão

- Tentamos encontrar a menor árvore consistente com os exemplos
 - Navalha de Ockham (escolha a hipótese mais simples)
 - NP-completo (Quinlan 1986, Hyafil and Rivest 1976)

Heurística:

- (Recursivamente) Escolha o atributo mais significativo como raiz da (sub)árvore
 - Aquele que faz a maior diferença na classificação de um exemplo

Construindo Árvores de Decisão

- Tentamos encontrar a menor árvore consistente com os exemplos
 - Navalha de Ockham (escolha a hipótese mais simples)
 - NP-completo (Quinlan 1986, Hyafil and Rivest 1976)

Heurística:

- (Recursivamente) Escolha o atributo mais significativo como raiz da (sub)árvore
 - Aquele que faz a maior diferença na classificação de um exemplo
 - Tentativa de classificar com um pequeno número de testes

Construindo Árvores de Decisão

Algoritmos

- Existem muitos algoritmos

Construindo Árvores de Decisão

Algoritmos

- Existem muitos algoritmos
 - Ex: ID3 (Quinlan 1986) e sua variação C4.5 (Quinlan 1993); CART (Breiman et al. 1984)

Construindo Árvores de Decisão

Algoritmos

- Existem muitos algoritmos
 - Ex: ID3 (Quinlan 1986) e sua variação C4.5 (Quinlan 1993); CART (Breiman et al. 1984)

ID3

- Emprega uma estratégia top-down gulosa de busca no espaço de possíveis árvores de decisão

Construindo Árvores de Decisão

Algoritmos

- Existem muitos algoritmos
 - Ex: ID3 (Quinlan 1986) e sua variação C4.5 (Quinlan 1993); CART (Breiman et al. 1984)

ID3

- Emprega uma estratégia top-down gulosa de busca no espaço de possíveis árvores de decisão
 - Começa com a pergunta “que atributo deve ser testado na raiz”?

Construindo Árvores de Decisão

Algoritmos

- Existem muitos algoritmos
 - Ex: ID3 (Quinlan 1986) e sua variação C4.5 (Quinlan 1993); CART (Breiman et al. 1984)

ID3

- Emprega uma estratégia top-down gulosa de busca no espaço de possíveis árvores de decisão
 - Começa com a pergunta “que atributo deve ser testado na raiz”?
 - Avalia cada atributo para determinar quão bem ele classifica os exemplos de treino

Construindo Árvores de Decisão

ID3

- O melhor atributo é então selecionado para a raiz

Construindo Árvores de Decisão

ID3

- O melhor atributo é então selecionado para a raiz
- Um descendente da raiz é criado para cada valor possível desse atributo

Construindo Árvores de Decisão

ID3

- O melhor atributo é então selecionado para a raiz
- Um descendente da raiz é criado para cada valor possível desse atributo
- Os exemplos de treino são rearranjados conforme cada nó descendente

Construindo Árvores de Decisão

ID3

- O melhor atributo é então selecionado para a raiz
- Um descendente da raiz é criado para cada valor possível desse atributo
- Os exemplos de treino são rearranjados conforme cada nó descendente
 - São filtrados conforme o valor daquele atributo

Construindo Árvores de Decisão

ID3

- O melhor atributo é então selecionado para a raiz
- Um descendente da raiz é criado para cada valor possível desse atributo
- Os exemplos de treino são rearranjados conforme cada nó descendente
 - São filtrados conforme o valor daquele atributo
- O processo é repetido para cada descendente, usando os exemplos associados com o nó desse descendente

Construindo Árvores de Decisão

Há 4 casos a considerar:

Construindo Árvores de Decisão

Há 4 casos a considerar:

- Se houver exemplos positivos e negativos, escolha o melhor atributo para separá-los

Construindo Árvores de Decisão

Há 4 casos a considerar:

- Se houver exemplos positivos e negativos, escolha o melhor atributo para separá-los
- Se todos os exemplos forem positivos (ou negativos), terminamos: podemos dizer sim ou não

Construindo Árvores de Decisão

Há 4 casos a considerar:

- Se houver exemplos positivos e negativos, escolha o melhor atributo para separá-los
- Se todos os exemplos forem positivos (ou negativos), terminamos: podemos dizer sim ou não
- Se não sobrarem exemplos, ou seja, nada parecido com a entrada foi observado

Construindo Árvores de Decisão

Há 4 casos a considerar:

- Se houver exemplos positivos e negativos, escolha o melhor atributo para separá-los
- Se todos os exemplos forem positivos (ou negativos), terminamos: podemos dizer sim ou não
- Se não sobrarem exemplos, ou seja, nada parecido com a entrada foi observado
 - Retornamos um valor padrão calculado com base na classificação da maioria no nó pai

Construindo Árvores de Decisão

Há 4 casos a considerar:

- Se houver exemplos positivos e negativos, escolha o melhor atributo para separá-los
- Se todos os exemplos forem positivos (ou negativos), terminamos: podemos dizer sim ou não
- Se não sobrarem exemplos, ou seja, nada parecido com a entrada foi observado
 - Retornamos um valor padrão calculado com base na classificação da maioria no nó pai
 - Ou então na classificação mais comum (moda), se não for binário

Construindo Árvores de Decisão

Há 4 casos a considerar (cont.):

- Se acabarem os atributos, e restarem tanto exemplos positivos quanto negativos, temos um problema

Construindo Árvores de Decisão

Há 4 casos a considerar (cont.):

- Se acabarem os atributos, e restarem tanto exemplos positivos quanto negativos, temos um problema
- Os exemplos têm a mesma descrição (conforme os atributos), mas classificações diferentes

Construindo Árvores de Decisão

Há 4 casos a considerar (cont.):

- Se acabarem os atributos, e restarem tanto exemplos positivos quanto negativos, temos um problema
 - Os exemplos têm a mesma descrição (conforme os atributos), mas classificações diferentes
- E o que isso significa?

Construindo Árvores de Decisão

Há 4 casos a considerar (cont.):

- Se acabarem os atributos, e restarem tanto exemplos positivos quanto negativos, temos um problema
 - Os exemplos têm a mesma descrição (conforme os atributos), mas classificações diferentes
- E o que isso significa?
 - Alguns dos dados estão incorretos – há ruído; ou

Construindo Árvores de Decisão

Há 4 casos a considerar (cont.):

- Se acabarem os atributos, e restarem tanto exemplos positivos quanto negativos, temos um problema
 - Os exemplos têm a mesma descrição (conforme os atributos), mas classificações diferentes
- E o que isso significa?
 - Alguns dos dados estão incorretos – há ruído; ou
 - Os atributos não dão informação suficiente; ou

Construindo Árvores de Decisão

Há 4 casos a considerar (cont.):

- Se acabarem os atributos, e restarem tanto exemplos positivos quanto negativos, temos um problema
 - Os exemplos têm a mesma descrição (conforme os atributos), mas classificações diferentes
- E o que isso significa?
 - Alguns dos dados estão incorretos – há ruído; ou
 - Os atributos não dão informação suficiente; ou
 - O domínio é verdadeiramente não-determinístico.

Construindo Árvore de Decisão

Há 4 casos a considerar (cont.):

- Se acabarem os atributos, e restarem tanto exemplos positivos quanto negativos, temos um problema
 - Os exemplos têm a mesma descrição (conforme os atributos), mas classificações diferentes
- E o que isso significa?
 - Alguns dos dados estão incorretos – há ruído; ou
 - Os atributos não dão informação suficiente; ou
 - O domínio é verdadeiramente não-determinístico.
 - Também aqui usamos a classificação mais comum.

Construindo Árvores de Decisão

Algoritmo

Função *ID3*(*Exemplos*, *Atributos*, *Padrão*): **árvore de decisão**

se *Exemplos* estiver vazio **então**

└ **retorna** nó com rótulo = *Padrão*

se todos os *Exemplos* têm a mesma classificação **então**

└ **retorna** nó com rótulo = classificação

se *Atributos* estiver vazia **então**

└ **retorna** nó com rótulo = ValorMaioria(*exemplos*)

At ← EscolhaMelhorAtributo(*Atributos*, *Exemplos*)

Árvore ← nova árvore de decisão com *At* como raiz

M ← ValorMaioria(*Exemplos*)

para cada valor possível, v_i , de *At* **faça**

└ *Exemplos* _{v_i} ← subconjunto de *Exemplos* com *At* = v_i

└ *Subárvore* ← *ID3*(*Exemplos* _{v_i} , *Atributos* − *At*, *M*)

└ Adicione um galho a *Árvore* com o rótulo v_i e subárvore *Subárvore*

retorna *Árvore*

Construindo Árvores de Decisão

Algoritmo

Função *ID3*(*Exemplos*, *Atributos*, *Padrão*): *árvore de decisão*

se *Exemplos* estiver vazio **então**

└ **retorna** nó com rótulo = *Padrão*

se todos os *Exemplos* têm a mesma classificação **então**

└ **retorna** nó com rótulo = classificação

se *Atributos* estiver vazia **então**

└ **retorna** nó com rótulo = ValorMaioria(*exemplos*)

At ← EscolhaMelhorAtributo(*Atributos*, *Exemplos*)

Árvore ← nova árvore de decisão com *At* como raiz

M ← ValorMaioria(*Exemplos*)

para cada valor possível, v_i , de *At* **faça**

└ *Exemplos* _{v_i} ← subconjunto de *Exemplos* com *At* = v_i

└ *Subárvore* ← *ID3*(*Exemplos* _{v_i} , *Atributos* − *At*, *M*)

└ Adicione um galho a *Árvore* com o rótulo v_i e subárvore *Subárvore*

retorna *Árvore*

Conjunto de exemplos de treino

Construindo Árvores de Decisão

Algoritmo

Função *ID3*(*Exemplos*, *Atributos*, *Padrão*): *árvore de decisão*

se *Exemplos* estiver vazio **então**

└ **retorna** nó com rótulo = *Padrão*

se todos os *Exemplos* têm a mesma classificação **então**

└ **retorna** nó com rótulo = classificação

se *Atributos* estiver vazia **então**

└ **retorna** nó com rótulo = *ValorMaioria*(*exemplos*)

At ← *EscolhaMelhorAtributo*(*Atributos*, *Exemplos*)

Árvore ← nova árvore de decisão com *At* como raiz

M ← *ValorMaioria*(*Exemplos*)

para cada *valor possível*, v_i , de *At* **faça**

└ *Exemplos* _{v_i} ← subconjunto de *Exemplos* com *At* = v_i

└ *Subárvore* ← *ID3*(*Exemplos* _{v_i} , *Atributos* − *At*, *M*)

└ Adicione um galho a *Árvore* com o rótulo v_i e subárvore *Subárvore*

retorna *Árvore*

Lista de atributos

Construindo Árvores de Decisão

Algoritmo

Função *ID3*(*Exemplos*, *Atributos*, *Padrão*): árvore de decisão

se *Exemplos* estiver vazio **então**

└ **retorna** nó com rótulo = *Padrão*

se todos os *Exemplos* têm a mesma classificação **então**

└ **retorna** nó com rótulo = classificação

se *Atributos* estiver vazia **então**

└ **retorna** nó com rótulo = ValorMaioria(*exemplos*)

At ← EscolhaMelhorAtributo(*Atributos*, *Exemplos*)

Árvore ← nova árvore de decisão com *At* como raiz

M ← ValorMaioria(*Exemplos*)

para cada valor possível, v_i , de *At* **faça**

└ *Exemplos* _{v_i} ← subconjunto de *Exemplos* com *At* = v_i

└ *Subárvore* ← *ID3*(*Exemplos* _{v_i} , *Atributos* − *At*, *M*)

└ Adicione um galho a *Árvore* com o rótulo v_i e subárvore *Subárvore*

retorna *Árvore*

Valor padrão para a
instância classificada

Construindo Árvores de Decisão

Algoritmo

Função *ID3*(*Exemplos*, *Atributos*, *Padrão*): **árvore de decisão**

se *Exemplos* estiver vazio **então**

└ **retorna** nó com rótulo = *Padrão*

se todos os *Exemplos* têm a mesma classificação **então**

└ **retorna** nó com rótulo = classificação

se *Atributos* estiver vazia **então**

└ **retorna** nó com rótulo = ValorMaioria(*exemplos*)

At ← EscolhaMelhorAtributo(*Atributos*, *Exemplos*)

Árvore ← nova árvore de decisão com *At* como raiz

M ← ValorMaioria(*Exemplos*)

para cada valor possível, v_i , de *At* **faça**

└ *Exemplos* _{v_i} ← subconjunto de *Exemplos* com *At* = v_i

└ *Subárvore* ← *ID3*(*Exemplos* _{v_i} , *Atributos* - *At*, *M*)

└ Adicione um galho a *Árvore* com o rótulo v_i e subárvore *Subárvore*

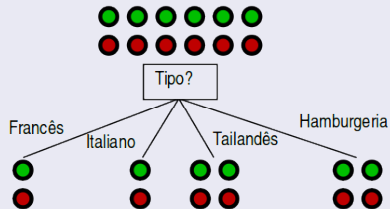
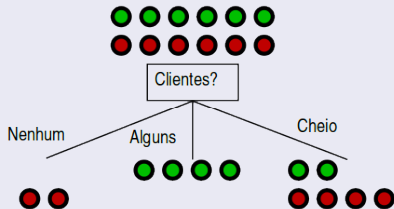
retorna *Árvore*

Novo padrão calculado
com base na classi-
ficação da maioria dos
exemplos (usada nos
filhos do nó com *At*)

Escolha do Atributo

O que significa ser o melhor atributo?

Considere os atributos:

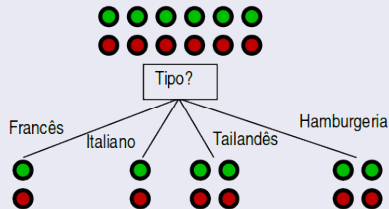
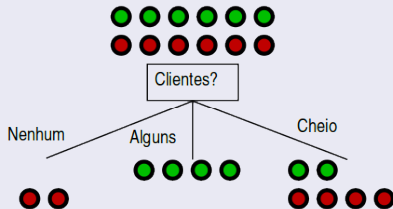


Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Escolha do Atributo

O que significa ser o melhor atributo?

Considere os atributos:



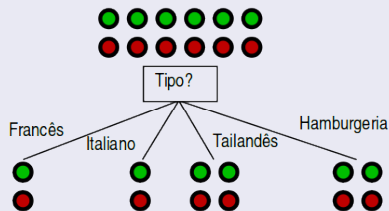
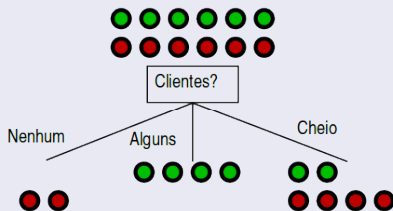
Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Um bom atributo divide os exemplos em subgrupos que são (idealmente) “todos positivos” ou “todos negativos”

Escolha do Atributo

O que significa ser o melhor atributo?

Considere os atributos:



Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Um bom atributo divide os exemplos em subgrupos que são (idealmente) “todos positivos” ou “todos negativos”

- Clientes é uma escolha melhor que Tipo

Escolha do Atributo

Como medimos quão bom é um atributo?

Escolha do Atributo

Como medimos quão bom é um atributo?

- Terá valor máximo quando o atributo é perfeito

Escolha do Atributo

Como medimos quão bom é um atributo?

- Terá valor máximo quando o atributo é perfeito
 - Divide os exemplos em conjuntos ou totalmente positivos ou negativos

Escolha do Atributo

Como medimos quão bom é um atributo?

- Terá valor máximo quando o atributo é perfeito
 - Divide os exemplos em conjuntos ou totalmente positivos ou negativos
- Valor mínimo quando o atributo é inútil

Escolha do Atributo

Como medimos quão bom é um atributo?

- Terá valor máximo quando o atributo é perfeito
 - Divide os exemplos em conjuntos ou totalmente positivos ou negativos
- Valor mínimo quando o atributo é inútil
 - Torna a escolha algo aleatório – igual número de exemplos para cada valor

Escolha do Atributo

Como medimos quão bom é um atributo?

- Terá valor máximo quando o atributo é perfeito
 - Divide os exemplos em conjuntos ou totalmente positivos ou negativos
- Valor mínimo quando o atributo é inútil
 - Torna a escolha algo aleatório – igual número de exemplos para cada valor
- Uma medida possível é a quantidade esperada de informação fornecida pelo atributos

Escolha do Atributo

Como medimos quão bom é um atributo?

- Terá valor máximo quando o atributo é perfeito
 - Divide os exemplos em conjuntos ou totalmente positivos ou negativos
- Valor mínimo quando o atributo é inútil
 - Torna a escolha algo aleatório – igual número de exemplos para cada valor
- Uma medida possível é a quantidade esperada de informação fornecida pelo atributos
 - Medida pela entropia – conceito emprestado da física e teoria da informação (Shannon and Weaver, 1949)

Escolha do Atributo

Como medimos quão bom é um atributo?

- Também conhecida como Ganho de Informação

Escolha do Atributo

Como medimos quão bom é um atributo?

- Também conhecida como Ganho de Informação
 - Quão bem um dado atributo separa os exemplos de treino de acordo com sua classificação

Escolha do Atributo

Como medimos quão bom é um atributo?

- Também conhecida como Ganho de Informação
 - Quão bem um dado atributo separa os exemplos de treino de acordo com sua classificação

Caso binário:

Escolha do Atributo

Como medimos quão bom é um atributo?

- Também conhecida como Ganho de Informação
 - Quão bem um dado atributo separa os exemplos de treino de acordo com sua classificação

Caso binário:

- Dada uma coleção S

Escolha do Atributo

Como medimos quão bom é um atributo?

- Também conhecida como Ganho de Informação
 - Quão bem um dado atributo separa os exemplos de treino de acordo com sua classificação

Caso binário:

- Dada uma coleção S
 - Seja p a proporção de exemplos positivos em S

Escolha do Atributo

Como medimos quão bom é um atributo?

- Também conhecida como Ganho de Informação
 - Quão bem um dado atributo separa os exemplos de treino de acordo com sua classificação

Caso binário:

- Dada uma coleção S
 - Seja p a proporção de exemplos positivos em S
 - Seja n a proporção de exemplos negativos em S

Escolha do Atributo

Como medimos quão bom é um atributo?

- Também conhecida como Ganho de Informação
 - Quão bem um dado atributo separa os exemplos de treino de acordo com sua classificação

Caso binário:

- Dada uma coleção S
 - Seja p a proporção de exemplos positivos em S
 - Seja n a proporção de exemplos negativos em S
- A entropia de S em relação a essa classificação será:

$$H(S) = -p \cdot \log_2 p - n \cdot \log_2 n$$

Escolha do Atributo

Interpretação da entropia (teoria da informação):

Escolha do Atributo

Interpretação da entropia (teoria da informação):

- Especifica o número mínimo de bits de informação necessária para codificar a classificação de um membro arbitrário de S

Interpretação da entropia (teoria da informação):

- Especifica o número mínimo de bits de informação necessária para codificar a classificação de um membro arbitrário de S
- Um membro de S escolhido aleatoriamente, em uma distribuição uniforme

Escolha do Atributo

Interpretação da entropia (teoria da informação):

- Especifica o número mínimo de bits de informação necessária para codificar a classificação de um membro arbitrário de S
- Um membro de S escolhido aleatoriamente, em uma distribuição uniforme
- 1 bit é suficiente para uma pergunta do tipo sim/não, sobre a qual nada sabemos

Escolha do Atributo

Interpretação da entropia (teoria da informação):

- Especifica o número mínimo de bits de informação necessária para codificar a classificação de um membro arbitrário de S
- Um membro de S escolhido aleatoriamente, em uma distribuição uniforme
- 1 bit é suficiente para uma pergunta do tipo sim/não, sobre a qual nada sabemos
- Quanto menos soubermos sobre a resposta, mais informação será provida nela – Conteúdo de informação da resposta

Escolha do Atributo

Interpretação da entropia (teoria da informação):

- Especifica o número mínimo de bits de informação necessária para codificar a classificação de um membro arbitrário de S
- Um membro de S escolhido aleatoriamente, em uma distribuição uniforme
- 1 bit é suficiente para uma pergunta do tipo sim/não, sobre a qual nada sabemos
- Quanto menos soubermos sobre a resposta, mais informação será provida nela – Conteúdo de informação da resposta
 - Mais bits ela terá

Escolha do Atributo

Interpretação da entropia (teoria da informação):

- Ex:

Interpretação da entropia (teoria da informação):

- Ex:
 - $p = 1 \Rightarrow$ o receptor sabe que o exemplo sorteado é positivo

Interpretação da entropia (teoria da informação):

- Ex:
 - $p = 1 \Rightarrow$ o receptor sabe que o exemplo sorteado é positivo
 - Nenhuma mensagem precisa ser enviada – entropia 0

Interpretação da entropia (teoria da informação):

- Ex:
 - $p = 1 \Rightarrow$ o receptor sabe que o exemplo sorteado é positivo
 - Nenhuma mensagem precisa ser enviada – entropia 0
 - $p = 0.5 \Rightarrow$ é necessário 1 bit para indicar se o exemplo sorteado é positivo ou negativo

Interpretação da entropia (teoria da informação):

- Ex:
 - $p = 1 \Rightarrow$ o receptor sabe que o exemplo sorteado é positivo
 - Nenhuma mensagem precisa ser enviada – entropia 0
 - $p = 0.5 \Rightarrow$ é necessário 1 bit para indicar se o exemplo sorteado é positivo ou negativo
 - $p = 0.8 \Rightarrow$ um conjunto de mensagens pode ser codificadas usando em média menos de 1 bit/mensagem

Interpretação da entropia (teoria da informação):

- Ex:
 - $p = 1 \Rightarrow$ o receptor sabe que o exemplo sorteado é positivo
 - Nenhuma mensagem precisa ser enviada – entropia 0
 - $p = 0.5 \Rightarrow$ é necessário 1 bit para indicar se o exemplo sorteado é positivo ou negativo
 - $p = 0.8 \Rightarrow$ um conjunto de mensagens pode ser codificadas usando em média menos de 1 bit/mensagem
 - Associando-se códigos mais curtos a conjuntos de exemplos positivos e maiores aos menos prováveis exemplos negativos

Escolha do Atributo

Em geral, se o atributo alvo pode assumir n valores diferentes, então a entropia de S relativa a essa classificação é definida como

$$H(S) \equiv - \sum_{i=1}^n P_i \cdot \log_2 P_i$$

onde P_i é a proporção de S pertencente à classe i .

Escolha do Atributo

Em geral, se o atributo alvo pode assumir n valores diferentes, então a entropia de S relativa a essa classificação é definida como

$$H(S) \equiv - \sum_{i=1}^n P_i \cdot \log_2 P_i$$

onde P_i é a proporção de S pertencente à classe i .

Alternativamente...

Em geral, se as saídas (classes) possíveis v_i em S tiverem probabilidades $P(v_i|S)$, o conteúdo de informação I da resposta (ou sua impureza), no conjunto S relativo a um determinado nó, será:

$$I(S) = I(P(v_1), \dots, P(v_n)) = - \sum_{i=1}^n P(v_i) \log_2 P(v_i)$$

Usamos $P(v_i)$ em vez de $P(v_i|S)$ para simplificar.

Escolha do Atributo

Exemplo

Escolha do Atributo

Exemplo

- Suponha que temos p exemplos positivos e n negativos a partir da escolha de um atributo

Escolha do Atributo

Exemplo

- Suponha que temos p exemplos positivos e n negativos a partir da escolha de um atributo
- Qual o grau de desordem desse conjunto, sabendo que $n_p = n_n = 6$?

Escolha do Atributo

Exemplo

- Suponha que temos p exemplos positivos e n negativos a partir da escolha de um atributo
- Qual o grau de desordem desse conjunto, sabendo que $n_p = n_n = 6$?

$$I(P(v_1), \dots, P(v_n)) = - \sum_{i=1}^n P(v_i) \log_2 P(v_i)$$

Escolha do Atributo

Exemplo

- Suponha que temos p exemplos positivos e n negativos a partir da escolha de um atributo
- Qual o grau de desordem desse conjunto, sabendo que $n_p = n_n = 6$?

$$I(P(v_1), \dots, P(v_n)) = - \sum_{i=1}^n P(v_i) \log_2 P(v_i)$$

$$I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = -\frac{n_p}{n_p+n_n} \log_2 \left(\frac{n_p}{n_p+n_n}\right) - \frac{n_n}{n_p+n_n} \log_2 \left(\frac{n_n}{n_p+n_n}\right)$$

Escolha do Atributo

Exemplo

- Suponha que temos p exemplos positivos e n negativos a partir da escolha de um atributo
- Qual o grau de desordem desse conjunto, sabendo que $n_p = n_n = 6$?

$$I(P(v_1), \dots, P(v_n)) = - \sum_{i=1}^n P(v_i) \log_2 P(v_i)$$

$$\begin{aligned} I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) &= -\frac{n_p}{n_p+n_n} \log_2 \left(\frac{n_p}{n_p+n_n}\right) - \frac{n_n}{n_p+n_n} \log_2 \left(\frac{n_n}{n_p+n_n}\right) \\ &= -\left(\frac{1}{2} \log_2 \frac{1}{2}\right) - \left(\frac{1}{2} \log_2 \frac{1}{2}\right) \end{aligned}$$

Escolha do Atributo

Exemplo

- Suponha que temos p exemplos positivos e n negativos a partir da escolha de um atributo
- Qual o grau de desordem desse conjunto, sabendo que $n_p = n_n = 6$?

$$I(P(v_1), \dots, P(v_n)) = - \sum_{i=1}^n P(v_i) \log_2 P(v_i)$$

$$\begin{aligned} I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) &= -\frac{n_p}{n_p+n_n} \log_2 \left(\frac{n_p}{n_p+n_n}\right) - \frac{n_n}{n_p+n_n} \log_2 \left(\frac{n_n}{n_p+n_n}\right) \\ &= -\left(\frac{1}{2} \log_2 \frac{1}{2}\right) - \left(\frac{1}{2} \log_2 \frac{1}{2}\right) \\ &= 1 \end{aligned}$$

Escolha do Atributo

E como afinal escolhemos o atributo?

- Verificamos quanta informação ainda precisaremos após o teste de cada atributo

Escolha do Atributo

E como afinal escolhemos o atributo?

- Verificamos quanta informação ainda precisaremos após o teste de cada atributo
- Cada atributo A separa o conjunto de treino S em subconjuntos S_1, \dots, S_v , de acordo com cada valor distinto v_i de A

Escolha do Atributo

E como afinal escolhemos o atributo?

- Verificamos quanta informação ainda precisaremos após o teste de cada atributo
- Cada atributo A separa o conjunto de treino S em subconjuntos S_1, \dots, S_v , de acordo com cada valor distinto v_i de A
 - Assim, S_i é o subconjunto de S em que $A = v_i$

Escolha do Atributo

E como afinal escolhemos o atributo?

- Verificamos quanta informação ainda precisaremos após o teste de cada atributo
- Cada atributo A separa o conjunto de treino S em subconjuntos S_1, \dots, S_v , de acordo com cada valor distinto v_i de A
 - Assim, S_i é o subconjunto de S em que $A = v_i$
- Cada um (espera-se) necessitando de menos informação para completar a classificação

Escolha do Atributo

E como afinal escolhemos o atributo?

- Verificamos quanta informação ainda precisaremos após o teste de cada atributo
 - Cada atributo A separa o conjunto de treino S em subconjuntos S_1, \dots, S_v , de acordo com cada valor distinto v_i de A
 - Assim, S_i é o subconjunto de S em que $A = v_i$
 - Cada um (espera-se) necessitando de menos informação para completar a classificação
 - Suponha que S_i tenha p_i exemplos positivos e n_i exemplos negativos:

Escolha do Atributo

E como afinal escolhemos o atributo?

- Verificamos quanta informação ainda precisaremos após o teste de cada atributo
- Cada atributo A separa o conjunto de treino S em subconjuntos S_1, \dots, S_v , de acordo com cada valor distinto v_i de A
 - Assim, S_i é o subconjunto de S em que $A = v_i$
- Cada um (espera-se) necessitando de menos informação para completar a classificação
- Suponha que S_i tenha p_i exemplos positivos e n_i exemplos negativos:
 - Para classificar um novo exemplo, a partir desse ponto, serão necessários $I\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right)$ bits

Escolha do Atributo

E como afinal escolhemos o atributo?

- Um exemplo aleatório de S terá o valor v_i para A , no nó em que A é testado, com probabilidade

$$\frac{|S_i|}{|S|} = \frac{p_i + n_i}{p + n}$$

Escolha do Atributo

E como afinal escolhemos o atributo?

- Um exemplo aleatório de S terá o valor v_i para A , no nó em que A é testado, com probabilidade $\frac{|S_i|}{|S|} = \frac{p_i + n_i}{p + n}$
- Em média, após testarmos A , precisaremos de

$$\text{Restante}(S, A) = \sum_{i=1}^v \frac{p_i + n_i}{p + n} I \left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i} \right)$$

bits de informação para classificar o exemplo, onde v é o número de valores distintos do atributo A

Escolha do Atributo

E como afinal escolhemos o atributo?

- Um exemplo aleatório de S terá o valor v_i para A , no nó em que A é testado, com probabilidade

$$\frac{|S_i|}{|S|} = \frac{p_i + n_i}{p + n}$$

- Em média, após testarmos A , precisaremos de

Média ponderada da informação necessária a partir de cada v_i

$$Restante(S, A) = \sum_{i=1}^v \frac{p_i + n_i}{p + n} I\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

bits de informação para classificar o exemplo, onde v é o número de valores distintos do atributo A

Escolha do Atributo

E como afinal escolhemos o atributo?

- Ganho de informação a partir do teste do atributo em A

Escolha do Atributo

E como afinal escolhemos o atributo?

- Ganho de informação a partir do teste do atributo em A
 - Diferença entre a necessidade anterior de informação e a nova

Escolha do Atributo

E como afinal escolhemos o atributo?


- Ganho de informação a partir do teste do atributo em A
- Diferença entre a necessidade anterior de informação e a nova

$$Ganho(S, A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - Restante(S, A)$$

Escolha do Atributo

E como afinal escolhemos o atributo?

- Ganho de informação a partir do teste do atributo em A
- Diferença entre a necessidade anterior de informação e a nova

$$\text{Ganho}(S, A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - \text{Restante}(S, A)$$


Necessidade no pai do nó com A
(ou seja, antes do teste de A). Os
valores de p e n são os que restam
após a escolha feita no pai de A

Escolha do Atributo

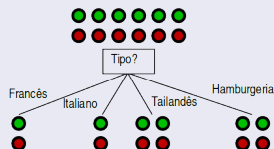
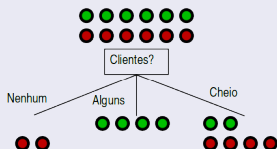
E como afinal escolhemos o atributo?

- A heurística usada para selecionar o atributo é escolher aquele com maior ganho

Escolha do Atributo

E como afinal escolhemos o atributo?

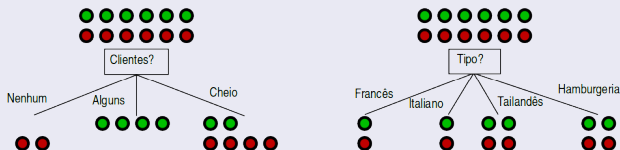
- A heurística usada para selecionar o atributo é escolher aquele com maior ganho
- Ex:



Escolha do Atributo

E como afinal escolhemos o atributo?

- A heurística usada para selecionar o atributo é escolher aquele com maior ganho
- Ex:

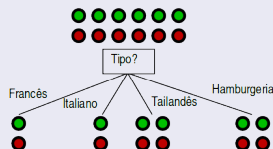
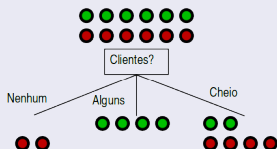


$$\text{Ganho}(\text{clientes}) = 1 - \left[\frac{2}{12} I\left(\frac{0}{2}, \frac{2}{2}\right) + \frac{4}{12} I\left(\frac{4}{4}, \frac{0}{4}\right) + \frac{6}{12} I\left(\frac{2}{6}, \frac{4}{6}\right) \right] \approx 0,541 \text{ bits}$$

Escolha do Atributo

E como afinal escolhemos o atributo?

- A heurística usada para selecionar o atributo é escolher aquele com maior ganho
- Ex:



$$\text{Ganho}(\text{clientes}) = 1 - \left[\frac{2}{12} I\left(\frac{0}{2}, \frac{2}{2}\right) + \frac{4}{12} I\left(\frac{4}{4}, \frac{0}{4}\right) + \frac{6}{12} I\left(\frac{2}{6}, \frac{4}{6}\right) \right] \approx 0,541 \text{ bits}$$

$$\text{Ganho}(\text{tipo}) = 1 - \left[\frac{2}{12} I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{12} I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{12} I\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{4}{12} I\left(\frac{2}{4}, \frac{2}{4}\right) \right] \approx 0 \text{ bit}$$

Construindo Árvores de Decisão

ID3 – Exemplo

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x1 | S | N | N | S | Alg | \$\$\$ | N | S | Fr | 0-10 | S |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x3 | N | S | N | N | Alg | \$ | N | N | Bu | 0-10 | S |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x6 | N | S | N | S | Alg | \$\$ | S | S | It | 0-10 | S |
| x7 | N | S | N | N | Nen | \$ | S | N | Bu | 0-10 | N |
| x8 | N | N | N | S | Alg | \$\$ | S | S | Ta | 0-10 | S |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x11 | N | N | N | N | Nen | \$ | N | N | Ta | 0-10 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |

Exemplos: {x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x11, x12}
Atributos: Alt, Bar, Fri, Hun, Pat, Price, Rain, Res, Type, Est
Padrão: N (Não esperar)

Construindo Árvores de Decisão

ID3 – Exemplo

- O ID3 determina o ganho de informação para cada atributo candidato, e seleciona aquele com maior ganho.

| Atributo | Ganho |
|-------------|-------|
| Alternativa | 0,0 |
| Bar | 0,0 |
| Sex/Sáb | 0,021 |
| Faminto | 0,196 |
| Cientes | 0,541 |
| Preço | 0,196 |
| Chuva | 0,0 |
| Reserva | 0,021 |
| Tipo | 0,0 |
| Estimativa | 0,208 |

Construindo Árvores de Decisão

ID3 – Exemplo

- O ID3 determina o ganho de informação para cada atributo candidato, e seleciona aquele com maior ganho.
- “Clientes” é o atributo com melhor previsão do atributo-alvo (se esperaremos para comer)

| Atributo | Ganho |
|-------------|-------|
| Alternativa | 0,0 |
| Bar | 0,0 |
| Sex/Sáb | 0,021 |
| Faminto | 0,196 |
| Clientes | 0,541 |
| Preço | 0,196 |
| Chuva | 0,0 |
| Reserva | 0,021 |
| Tipo | 0,0 |
| Estimativa | 0,208 |

Construindo Árvores de Decisão

ID3 – Exemplo

- O ID3 determina o ganho de informação para cada atributo candidato, e seleciona aquele com maior ganho.
- “Clientes” é o atributo com melhor previsão do atributo-alvo (se esperaremos para comer)
- Separemos então o conjunto de dados para cada um dos valores de “Clientes”

| Atributo | Ganho |
|-------------|-------|
| Alternativa | 0,0 |
| Bar | 0,0 |
| Sex/Sáb | 0,021 |
| Faminto | 0,196 |
| Clientes | 0,541 |
| Preço | 0,196 |
| Chuva | 0,0 |
| Reserva | 0,021 |
| Tipo | 0,0 |
| Estimativa | 0,208 |

Construindo Árvores de Decisão

ID3 – Exemplo

| Atributo | Ganho |
|-------------|-------|
| Alternativa | 0,0 |
| Bar | 0,0 |
| Sex/Sáb | 0,021 |
| Faminto | 0,196 |
| Clientes | 0,541 |
| Preço | 0,196 |
| Chuva | 0,0 |
| Reserva | 0,021 |
| Tipo | 0,0 |
| Estimativa | 0,208 |

Construindo Árvore de Decisão

ID3 – Exemplo

| Atributo | Ganho |
|-------------|-------|
| Alternativa | 0,0 |
| Bar | 0,0 |
| Sex/Sáb | 0,021 |
| Faminto | 0,196 |
| Cientes | 0,541 |
| Preço | 0,196 |
| Chuva | 0,0 |
| Reserva | 0,021 |
| Tipo | 0,0 |
| Estimativa | 0,208 |

| Ex | Atributos | | | | | | | | | | Obj |
|----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x1 | S | N | N | S | Alg | \$\$\$ | N | S | Fr | 0-10 | S |
| x3 | N | S | N | N | Alg | \$ | N | N | Bu | 0-10 | S |
| x6 | N | S | N | S | Alg | \$\$ | S | S | It | 0-10 | S |
| x8 | N | N | N | S | Alg | \$\$ | S | S | Ta | 0-10 | S |

Construindo Árvores de Decisão

ID3 – Exemplo

| Atributo | Ganho |
|-------------|-------|
| Alternativa | 0,0 |
| Bar | 0,0 |
| Sex/Sáb | 0,021 |
| Faminto | 0,196 |
| Cientes | 0,541 |
| Preço | 0,196 |
| Chuva | 0,0 |
| Reserva | 0,021 |
| Tipo | 0,0 |
| Estimativa | 0,208 |

| Ex | Atributos | | | | | | | | | | Obj |
|----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x1 | S | N | N | S | Alg | \$\$\$ | N | S | Fr | 0-10 | S |
| x3 | N | S | N | N | Alg | \$ | N | N | Bu | 0-10 | S |
| x6 | N | S | N | S | Alg | \$\$ | S | S | It | 0-10 | S |
| x8 | N | N | N | S | Alg | \$\$ | S | S | Ta | 0-10 | S |

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x7 | N | S | N | N | Nen | \$ | S | N | Bu | 0-10 | N |
| x11 | N | N | N | N | Nen | \$ | N | N | Ta | 0-10 | N |

Construindo Árvores de Decisão

ID3 – Exemplo

| Atributo | Ganho |
|-------------|-------|
| Alternativa | 0,0 |
| Bar | 0,0 |
| Sex/Sáb | 0,021 |
| Faminto | 0,196 |
| Cientes | 0,541 |
| Preço | 0,196 |
| Chuva | 0,0 |
| Reserva | 0,021 |
| Tipo | 0,0 |
| Estimativa | 0,208 |

| Ex | Atributos | | | | | | | | | | Obj |
|----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x1 | S | N | N | S | Alg | \$\$\$ | N | S | Fr | 0-10 | S |
| x3 | N | S | N | N | Alg | \$ | N | N | Bu | 0-10 | S |
| x6 | N | S | N | S | Alg | \$\$ | S | S | It | 0-10 | S |
| x8 | N | N | N | S | Alg | \$\$ | S | S | Ta | 0-10 | S |

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x7 | N | S | N | N | Nen | \$ | S | N | Bu | 0-10 | N |
| x11 | N | N | N | N | Nen | \$ | N | N | Ta | 0-10 | N |

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |

Construindo Árvores de Decisão

ID3 – Exemplo

| Atributo | Ganho |
|-------------|-------|
| Alternativa | 0,0 |
| Bar | 0,0 |
| Sex/Sáb | 0,021 |
| Faminto | 0,196 |
| Cientes | 0,541 |
| Preço | 0,196 |
| Chuva | 0,0 |
| Reserva | 0,021 |
| Tipo | 0,0 |
| Estimativa | 0,208 |

| Ex | Atributos | | | | | | | | | | Obj |
|----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x1 | S | N | N | S | Alg | \$\$\$ | N | S | Fr | 0-10 | S |
| x3 | N | S | N | N | Alg | \$ | N | N | Bu | 0-10 | S |
| x6 | N | S | N | S | Alg | \$\$ | S | S | It | 0-10 | S |
| x8 | N | N | N | S | Alg | \$\$ | S | S | Ta | 0-10 | S |

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x7 | N | S | N | N | Nen | \$ | S | N | Bu | 0-10 | N |
| x11 | N | N | N | N | Nen | \$ | N | N | Ta | 0-10 | N |

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |

Note que para todo exemplo em que $Cli = Alg$, $Obj = S$; e para todo em que $Cli = Nen$, $Obj = N$

Construindo Árvores de Decisão

ID3 – Exemplo

| Atributo | Ganho |
|-------------|-------|
| Alternativa | 0,0 |
| Bar | 0,0 |
| Sex/Sáb | 0,021 |
| Faminto | 0,196 |
| Cientes | 0,541 |
| Preço | 0,196 |
| Chuva | 0,0 |
| Reserva | 0,021 |
| Tipo | 0,0 |
| Estimativa | 0,208 |

Esses nós da árvore tornam-se folhas com as respectivas classificações

| Ex | Atributos | | | | | | | | | | Obj |
|----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x1 | S | N | N | S | Alg | \$\$\$ | N | S | Fr | 0-10 | S |
| x3 | N | S | N | N | Alg | \$ | N | N | Bu | 0-10 | S |
| x6 | N | S | N | S | Alg | \$\$ | S | S | It | 0-10 | S |
| x8 | N | N | N | S | Alg | \$\$ | S | S | Ta | 0-10 | S |

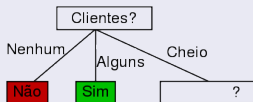
| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x7 | N | S | N | N | Nen | \$ | S | N | Bu | 0-10 | N |
| x11 | N | N | N | N | Nen | \$ | N | N | Ta | 0-10 | N |

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |

Construindo Árvore de Decisão

ID3 – Exemplo

| Atributo | Ganho |
|----------------|--------------|
| Alternativa | 0,0 |
| Bar | 0,0 |
| Sex/Sáb | 0,021 |
| Faminto | 0,196 |
| Cientes | 0,541 |
| Preço | 0,196 |
| Chuva | 0,0 |
| Reserva | 0,021 |
| Tipo | 0,0 |
| Estimativa | 0,208 |



| Ex | Atributos | | | | | | | | | | Obj |
|----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x1 | S | N | N | S | Alg | \$\$\$ | N | S | Fr | 0-10 | S |
| x3 | N | S | N | N | Alg | \$ | N | N | Bu | 0-10 | S |
| x6 | N | S | N | S | Alg | \$\$ | S | S | It | 0-10 | S |
| x8 | N | N | N | S | Alg | \$\$ | S | S | Ta | 0-10 | S |

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x7 | N | S | N | N | Nen | \$ | S | N | Bu | 0-10 | N |
| x11 | N | N | N | N | Nen | \$ | N | N | Ta | 0-10 | N |

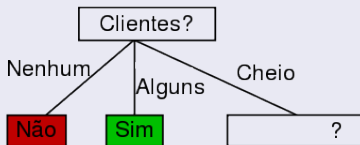
| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |

Construindo Árvores de Decisão

ID3 – Exemplo

- Por outro lado, os descendentes correspondendo a $Cli = Che$ têm entropia não nula

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |

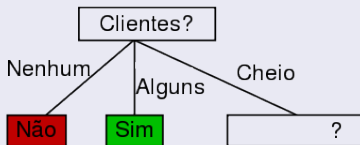


Construindo Árvores de Decisão

ID3 – Exemplo

- Por outro lado, os descendentes correspondendo a $Cli = Che$ têm entropia não nula
- A árvore deve ser desenvolvida abaixo deste nó

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |



Construindo Árvores de Decisão

ID3 – Exemplo

- O processo de selecionar um novo atributo é repetido para cada nó descendente não terminal

Construindo Árvore de Decisão

ID3 – Exemplo

- O processo de selecionar um novo atributo é repetido para cada nó descendente não terminal
 - Usando apenas os exemplos de treino associados a esse nó

Construindo Árvore de Decisão

ID3 – Exemplo

- O processo de selecionar um novo atributo é repetido para cada nó descendente não terminal
 - Usando apenas os exemplos de treino associados a esse nó
 - Atributos mais altos na árvore são descartados – não mais usados


Construindo Árvores de Decisão

ID3 – Exemplo

- O processo de selecionar um novo atributo é repetido para cada nó descendente não terminal
- Usando apenas os exemplos de treino associados a esse nó
- Atributos mais altos na árvore são descartados – não mais usados

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |

Usamos apenas os dados em que Clientes = cheio.


$$I\left(\frac{2}{6}, \frac{4}{6}\right) \approx 0,918$$

Construindo Árvore de Decisão

ID3 – Exemplo

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |

| Atributo | Ganho |
|-------------|-------|
| Alternativa | 0,109 |
| Bar | 0,0 |
| Sex/Sáb | 0,109 |
| Faminto | 0,251 |
| Preço | 0,251 |
| Chuva | 0,044 |
| Reserva | 0,251 |
| Tipo | 0,251 |
| Estimativa | 0,251 |

Construindo Árvores de Decisão

ID3 – Exemplo

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |

- Nesse caso, houve um empate

| Atributo | Ganho |
|-------------|-------|
| Alternativa | 0,109 |
| Bar | 0,0 |
| Sex/Sáb | 0,109 |
| Faminto | 0,251 |
| Preço | 0,251 |
| Chuva | 0,044 |
| Reserva | 0,251 |
| Tipo | 0,251 |
| Estimativa | 0,251 |

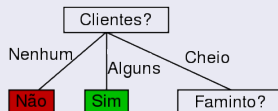
Construindo Árvores de Decisão

ID3 – Exemplo

| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |

- Nesse caso, houve um empate
- Escolhemos uma das opções: “Faminto”

| Atributo | Ganho |
|-------------|-------|
| Alternativa | 0,109 |
| Bar | 0,0 |
| Sex/Sáb | 0,109 |
| Faminto | 0,251 |
| Preço | 0,251 |
| Chuva | 0,044 |
| Reserva | 0,251 |
| Tipo | 0,251 |
| Estimativa | 0,251 |



Construindo Árvores de Decisão

ID3 – Exemplo

- O processo continua para cada nó folha até que uma de duas condições ocorram:

Construindo Árvores de Decisão

ID3 – Exemplo

- O processo continua para cada nó folha até que uma de duas condições ocorram:
 - Todo atributo já foi incluído no caminho; ou

Construindo Árvores de Decisão

ID3 – Exemplo

- O processo continua para cada nó folha até que uma de duas condições ocorram:
 - Todo atributo já foi incluído no caminho; ou
 - Os exemplos de treino associados com essa folha têm todos o mesmo atributo-alvo (sua entropia é 0)

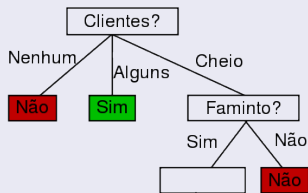
Construindo Árvores de Decisão

ID3 – Exemplo

- O processo continua para cada nó folha até que uma de duas condições ocorram:
 - Todo atributo já foi incluído no caminho; ou
 - Os exemplos de treino associados com essa folha têm todos o mesmo atributo-alvo (sua entropia é 0)

| Ex | Atributos | | | | | | | | | | Obj |
|----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-----|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x5 | S | N | S | N | Che | \$\$\$ | N | S | Fr | >60 | N |
| x9 | N | S | S | N | Che | \$ | S | N | Bu | >60 | N |

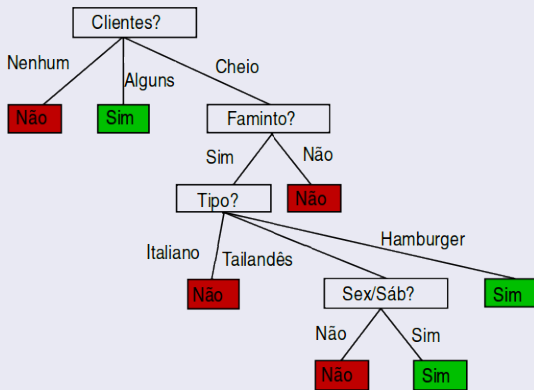
| Ex | Atributos | | | | | | | | | | Obj |
|-----|-----------|-----|-----|-----|-----|--------|-----|-----|-----|-------|-----|
| | Alt | Bar | Sex | Fam | Cli | Pre | Chu | Res | Tip | Est | |
| x2 | S | N | N | S | Che | \$ | N | N | Ta | 30-60 | N |
| x4 | S | N | S | S | Che | \$ | S | N | Ta | 10-30 | S |
| x10 | S | S | S | S | Che | \$\$\$ | N | S | It | 10-30 | N |
| x12 | S | S | S | S | Che | \$ | N | N | Bu | 30-60 | S |



Construindo Árvores de Decisão

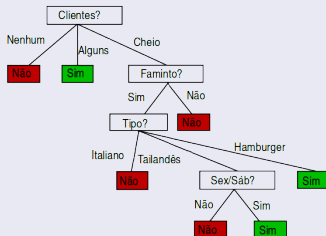
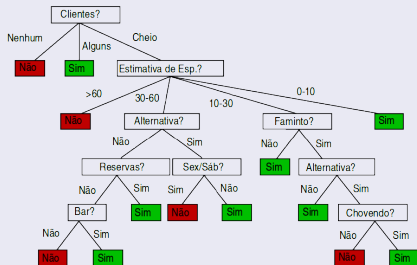
ID3 – Exemplo

- E continuando o processo chegaremos a



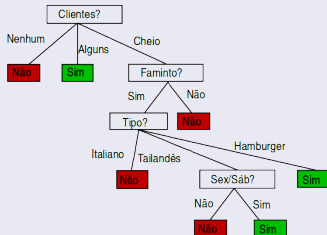
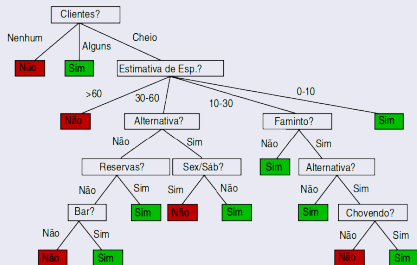
Construindo Árvores de Decisão

Exemplo – Comparando as árvores



Construindo Árvores de Decisão

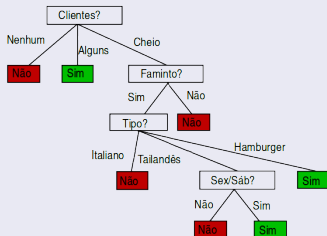
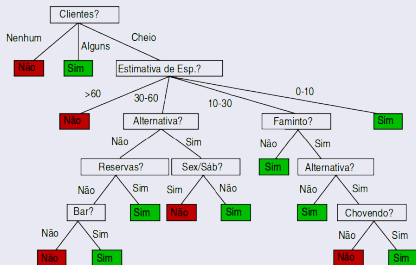
Exemplo – Comparando as árvores



- As árvores diferem porque o algoritmo olha para os exemplos, e não para a função correta

Construindo Árvores de Decisão

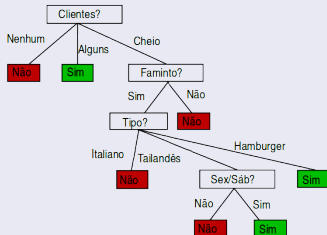
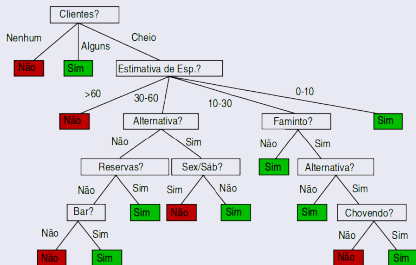
Exemplo – Comparando as árvores



- As árvores diferem porque o algoritmo olha para os exemplos, e não para a função correta
- A árvore resultante é consideravelmente mais simples

Construindo Árvores de Decisão

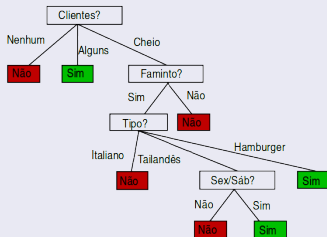
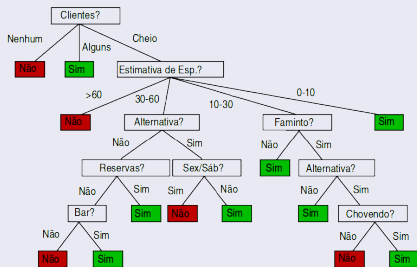
Exemplo – Comparando as árvores



- As árvores diferem porque o algoritmo olha para os exemplos, e não para a função correta
- A árvore resultante é consideravelmente mais simples
 - O algoritmo não tem razão para incluir a Chuva e a Reserva, uma vez que pode classificar todos os exemplos sem elas

Construindo Árvores de Decisão

Exemplo – Comparando as árvores



- Também detectou um padrão não-previsto: o autor da árvore esperará pela comida tailandesa nas sextas e sábados

Árvores de Decisão

Florestas Aleatórias



Fonte: <https://epoca.globo.com/colunas-e-blogs/blog-do-planeta/noticia/2016/08/floresta-perdida-do-brasil.html>

Árvores de Decisão

Florestas Aleatórias

- Nesse algoritmo, são criadas múltiplas árvores de decisão



Fonte: <https://epoca.globo.com/colunas-e-blogs/blog-do-planeta/noticia/2016/08/floresta-perdida-do-brasil.html>

Árvores de Decisão

Florestas Aleatórias

- Nesse algoritmo, são criadas múltiplas árvores de decisão
- Para classificar um novo exemplo:



Fonte: <https://epoca.globo.com/colunas-e-blogs/blog-do-planeta/noticia/2016/08/floresta-perdida-do-brasil.html>

Árvores de Decisão

Florestas Aleatórias

- Nesse algoritmo, são criadas múltiplas árvores de decisão
- Para classificar um novo exemplo:
 - Passe o exemplo a cada uma das árvores na floresta



Fonte: <https://epoca.globo.com/colunas-e-blogs/blog-do-planeta/noticia/2016/08/floresta-perdida-do-brasil.html>

Árvores de Decisão

Florestas Aleatórias

- Nesse algoritmo, são criadas múltiplas árvores de decisão
- Para classificar um novo exemplo:
 - Passe o exemplo a cada uma das árvores na floresta
 - O resultado da floresta será a classificação conforme o resultado da maioria das árvores nela



Fonte: <https://epoca.globo.com/colunas-e-blogs/blog-do-planeta/noticia/2016/08/floresta-perdida-do-brasil.html>

Árvores de Decisão

Florestas Aleatórias

- Nesse algoritmo, são criadas múltiplas árvores de decisão
- Para classificar um novo exemplo:
 - Passe o exemplo a cada uma das árvores na floresta
 - O resultado da floresta será a classificação conforme o resultado da maioria das árvores nela
 - Votação simples ou ponderada



Fonte: <https://epoca.globo.com/colunas-e-blogs/blog-do-planeta/noticia/2016/08/floresta-perdida-do-brasil.html>

Florestas Aleatórias – Crescimento

- Defina $m' \ll m$, onde m é o número de atributos (variáveis a serem consideradas) do problema

Florestas Aleatórias – Crescimento

- Defina $m' \ll m$, onde m é o número de atributos (variáveis a serem consideradas) do problema
- O valor de m' é mantido constante durante o crescimento da floresta

Florestas Aleatórias – Crescimento

- Defina $m' \ll m$, onde m é o número de atributos (variáveis a serem consideradas) do problema
 - O valor de m' é mantido constante durante o crescimento da floresta
- Para cada árvore na floresta:

Florestas Aleatórias – Crescimento

- Defina $m' \ll m$, onde m é o número de atributos (variáveis a serem consideradas) do problema
 - O valor de m' é mantido constante durante o crescimento da floresta
- Para cada árvore na floresta:
 - Tome uma amostra aleatória, com reposição, de n elementos do conjunto de treino, onde n é o tamanho do próprio conjunto de treino

Florestas Aleatórias – Crescimento

- Defina $m' \ll m$, onde m é o número de atributos (variáveis a serem consideradas) do problema
 - O valor de m' é mantido constante durante o crescimento da floresta
- Para cada árvore na floresta:
 - Tome uma amostra aleatória, com reposição, de n elementos do conjunto de treino, onde n é o tamanho do próprio conjunto de treino
 - Essa amostra será o conjunto de treino da árvore

Florestas Aleatórias – Crescimento

- Para cada árvore na floresta (cont.):
 - Em cada nó, selecione aleatoriamente m' atributos dos m possíveis

Florestas Aleatórias – Crescimento

- Para cada árvore na floresta (cont.):
 - Em cada nó, selecione aleatoriamente m' atributos dos m possíveis
 - Verifique cada um deles no nó, e escolha o melhor desses m' para expandir o nó

Florestas Aleatórias – Crescimento

- Para cada árvore na floresta (cont.):
 - Em cada nó, selecione aleatoriamente m' atributos dos m possíveis
 - Verifique cada um deles no nó, e escolha o melhor desses m' para expandir o nó
 - Continue o processo até não conseguir crescer mais a árvore

Florestas Aleatórias – Crescimento

- Para cada árvore na floresta (cont.):
 - Em cada nó, selecione aleatoriamente m' atributos dos m possíveis
 - Verifique cada um deles no nó, e escolha o melhor desses m' para expandir o nó
 - Continue o processo até não conseguir crescer mais a árvore
 - Cada árvore cresce até seu máximo \rightarrow não há poda

Referências

- Russell, S.; Norvig P. (2010): Artificial Intelligence: A Modern Approach. Prentice Hall. 3a ed.
 - Slides do livro: aima.eecs.berkeley.edu/slides-pdf/
- Mitchell, T.M.: Machine Learning. McGraw-Hill. 1997.
- Alpaydin, E.: Introduction to Machine Learning. 2 ed. MIT Press. 2010.
- Breiman, L.; Friedman, J.; Olshen, R.; Stone, C.J.: Classification and Regression Trees. Wadsworth. 1984.
- Fayyad, U.M.: On the induction of decision trees for multiple concept learning. Tese de Doutorado. EECS Department, University of Michigan. 1991.
- Harrington, P.: Machine Learning in Action. Manning. 2012.
- Hyafil, L.; Rivest, R. Constructing Optimal Binary Decision Trees is NP-complete. Information Processing Letters 5(1), 15–17. 1976.
- Nilsson, N.J.: Introduction to Machine Learning. 1998.
- Quinlan, J.R.: Induction of Decision Trees. Machine Learning, 1(1), 81–106. 1986.

Referências

- Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann. 1993.
- Shannon, C. E.; Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press. 1949.
- Theodoridis, S.; Koutroumbas, K.: Pattern Recognition. 4 ed. Academic Press. 2009.
- Webb, A.R.; Copsey, k.D.: Statistical Pattern Recognition. 3 ed. Wiley. 2011.
- <https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd>
- <https://medium.com/machina-sapiens/o-algoritmo-da-floresta-aleat%C3%B3ria-3545f6babdf8>
- <https://www.datacamp.com/community/tutorials/random-forests-classifier-python>
- https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm