Aprendizagem de Dados Simbólicos e/ou Numéricos

Francisco Carvalho

Árvores de Decisão: Dados Probabilistas

- Input
  - n objetos formando uma partição em m classes
  - cada objeto é descrito por variáveis modais de semântica probabilística.
- Objetivo
  - Descrever, na forma de uma árvore binária, as m classes da partição
  - Construir uma regra de decisão capaz de classificar novos objetos (cuja pertinência a uma das classes é desconhecida)

Árvores de Decisão: Dados Probabilistas

- Objetos
  - $E = \{1, \ldots, n\}$ objetos formando uma partição em m classes disjuntas $C_1, \ldots, C_m$
- Variáveis: cada objeto $k \in \Omega$ é descrito por dois tipos de variáveis
  - C: variável que define as classes (variável à explicar) C é uma variável uni-valorada nominal usual
  - $Y_1, \ldots, Y_p$, p variáveis explicativas (variáveis multivaloradas qualitativas e quantitativas com um sistema de pesos associado

Árvores de Decisão: Dados Probabilistas

- Tabela de dados simbólicos
  - A cada objeto k é associado uma descrição

$$a_k = [C(k) = c_k] \wedge [Y_1(k) \sim f_{k1}] \wedge \ldots \wedge [Y_p(k) \sim f_{kp}]$$

  - Descrição para a variável a explicar C: $[C(k) = c_k]$
    A classe de k é descrita na forma usual por um único valor $C(k) = c_k \in \{1,\ldots,m\}$, sem imprecisão
  - Descrição para a variável explicativa Y_j: $[Y_j(k) \sim f_{kj}]$
    Para cada objeto k, a descrição segundo Y_j é realizada por uma distribuição de probabilidade (ou de freqüências)

Árvores de Decisão: Dados Probabilistas

- Tabela de dados Probabilísticos

k	Y_1	…	Y_p	C
1				
:				
k	$Y_1(k) \sim f_{k1}$		$Y_p(k) \sim f_{p1}$	$C(k)=c_k$
:				
n				

Árvores de Decisão: Dados Probabilistas

- Dois tipos de descrições probabilísticas f_{kj}
  - Y_j é contínua
    O suporte de Y_j para k é um intervalo $V_{kj} \subset O_j \subseteq \Re$
    Nesse caso, $Y_j(k)$ tem uma distribuição uniforme f_{kj} definida em V_{kj}: $[Y_j(k) \sim U(V_{kj})]$
  - Exemplo
    $Y_j(k)$: tamanho de k
    Se $Y_j(k)$ é distribuída uniformemente e tem suporte [18,27]: $[Y_j(k) \sim U([18,27])]$

Árvores de Decisão: Dados Probabilistas

- Dois tipos de descrições probabilísticas f_{kj}
  - Y_j é discreta
    O suporte de Y_j para k é o conjunto $V_{kj} \subset O_j$
    Nesse caso, $Y_j(k)$ tem uma distribuição de probabilidade ou de freqüências f_{kj} definida em V_{kj}
  - Exemplo (variável multivalorada nominal)
    $Y_j(k)$: cor de k e [cor(k)~{1/4(branco), 3/4(azul)}]
  - Exemplo (variável quantitativa discreta)
    [número de pétalas~{1/2(7),1/3(8),1/6(9)}]

Árvores de Decisão: Dados Probabilistas

- Dados Boleanos
  - Associar um sistema de pesos a variável multivalorada
  - Exemplos
    [tamanho R [18,27]] → [tamanho ~ U ([18,27])]
    [cor R {azul, verde}] → [cor ~ {0.5 (azul), 0.5 (verde)}]
    [no de pétalas R {15,...,20}] → [no de pétalas ~ {1/6(15), ...,1/6(20)}]
    [cor R {azul} → [cor ~{1(azul)}]

Árvores de Decisão: Dados Probabilistas

- Exemplo de tabela de dados probabilistas

k	Cor	Tamanho	No de pétalas	C
1	1 (verde)	U([60,95])	1 (0.5), 2 (0.5)	1
2	0.5(azul), 0.5 verde)	U([30,40])	1 (0.3), 2 (0.4), 3 (0.30)	2
3	0.3(preto), 0.4 verde, 0.3(branco)	U([35,60])	3 (0.5), 4 (0.5)	1

Árvores de Decisão: Dados Probabilistas

- Construção de consultas binárias
  - Algoritmo recursivo para subdividir o conjunto E de objetos
  - A etapa de divisão é baseada em uma consulta binária de resposta sim/não baseada em uma única variável (e nos seus valores observados)

Árvores de Decisão: Dados Probabilistas

- Construção de consultas binárias
  - Dois tipos de questões binárias segundo se O_j é ordenado ou não

Variável	Nominal	Ordinal	Contínua
descrição consulta	distribuição $[Y \in V]?$	distribuição $[Y \leq c]?$	intervalo $[Y \leq c]?$

Árvores de Decisão: Dados Probabilistas

- Consultas do tipo $[Y_j \leq c]$
  - Caso discreto
    Se O_j tem n_j valores, constrói-se $n_j - 1$ questões binárias (o limiar c é escolhido entre o valores $v_{j,1}$, …,$v_{j,nj-1}$)
  - Exemplo
    Y_j número de pétalas e $O_j = \{1,2,3,4\}$
    Nesse caso, $c \in \{1,2,3\}$

Árvores de Decisão: Dados Probabilistas

- Consultas do tipo $[Y_j \leq c]$
  - Caso contínuo
    As descrições são do tipo $[m_{kj}, M_{kj}]$
    O limiar é o ponto médio entre dois limites m_{kj} e M_{kj} quaisquer ($k=1,\ldots,n$)
    Existe no máximo $2n-1$ questões binárias (limiares)
  - Exemplo: Na tabela anterior
    Y_j Tamanho
    Nesse caso, $c \in \{32.5, 37.5, 50, 60, 77.5\}$

Árvores de Decisão: Dados Probabilistas

- Consultas do tipo $[Y_j \in V]$
  - Se O_j tem n_j valores pode-se construir $2^{nj-1} - 1$ questões binárias, isto é, existe $2^{nj-1} - 1$ maneiras de dividir um domínio O_j com n_j valores
  - Exemplo: Y_j = cor e O_j = {azul, verde, branco}

$$V = \{verde\} \quad \overline{V} = \{branco, azul\} \quad \text{ou}$$

$$V = \{branco\} \quad \overline{V} = \{azul, verde\} \quad \text{ou}$$

$$V = \{azul\} \quad \overline{V} = \{verde, branco\}$$

Árvores de Decisão: Dados Probabilistas

- Construção de uma nova tabela de variáveis binárias
  - Seja $\mathcal{Q}$ = {1, …, Q} a totalidade das questões binárias correspondentes a variável Y_j

k	1	…	q	…	Q
1					
:					
k			p_{kq}		
:					
n					

Árvores de Decisão: Dados Probabilistas

- Construção de uma nova tabela de variáveis binárias
  - Hierarquia binária de classes com subclasses a esquerda e a direita (nó) para cada classe
  - A pertinência ao nó esquerdo depende da resposta positiva a consulta $[Y_j \leq c]$ (ou $[Y_j \in V]$)
  - A pertinência ao nó direito depende da resposta positiva a consulta $[Y_j > c]$ (ou $[Y_j \notin V]$)

Árvores de Decisão: Dados Probabilistas

- Construção de uma nova tabela de variáveis binárias
  - p_{kq} : probabilidade que k seja associado ao nó esquerdo
  - $1 - p_{kq}$: probabilidade que k seja associado ao nó direito

Árvores de Decisão: Dados Probabilistas

- Cálculo de p_{kq}
  - Y_j é contínua
    descrição k : $U([m_k, M_k])$
    questão binária : $[Y \leq c]$

$$p_{kq} = \begin{cases} 0, \text{se } c < m_{kj} \\ (c - m_{kj}) / (M_{kj} - m_{kj}), \text{se } c \in [m_{kj}, M_{kj}] \\ 1, \text{se } c > M_{kj} \end{cases}$$

Árvores de Decisão: Dados Probabilistas

- Cálculo de p_{kq}
  - Y_j é contínua
    Exemplo: Na tabela anterior k=1 e tamanho~U([80,95])
    Se q: [tamanho ≤ 37.5], $p_{1q} = 0$ e $1 - p_{1q} = 1$
    Se q: [tamanho ≤ 87.5], $p_{1q} = 0.5$ e $1 - p_{1q} = 0.5$
  - Y_j é ordinal
    descrição k : f_{kj}
    questão binária : $[Y \le c]$

$$p_{kq} = \sum_{v \le c} f_{kj}(v)$$

Árvores de Decisão: Dados Probabilistas

- Cálculo de p_{kq}
  - Y_j é ordinal
    Exemplo: Na tabela anterior k=2 e Número de pétalas~{1 (0.3), 2 (0.4), 3 (0.30)}
    Se q: [Número de pétalas ≤ 2], $p_{1q} = 0.7$ e $1 - p_{1q} = 0.3$
  - Y_j é nominal
    descrição k : f_{kj}
    questão binária : $[Y \in V]$

$$p_{kq} = \sum_{v \in V} f_{kj}(v)$$

Árvores de Decisão: Dados Probabilistas

- Cálculo de p_{kq}
  - Y_j é nominal

    Exemplo: Na tabela anterior k=3 e Cor~{preto (0.3), verde (0.4), branco (0.30)}

    Se q: [Cor $\in$ {verde}], $p_{1q} = 0.4$ e $1 - p_{1q} = 0.6$

Árvores de Decisão: Dados Probabilistas

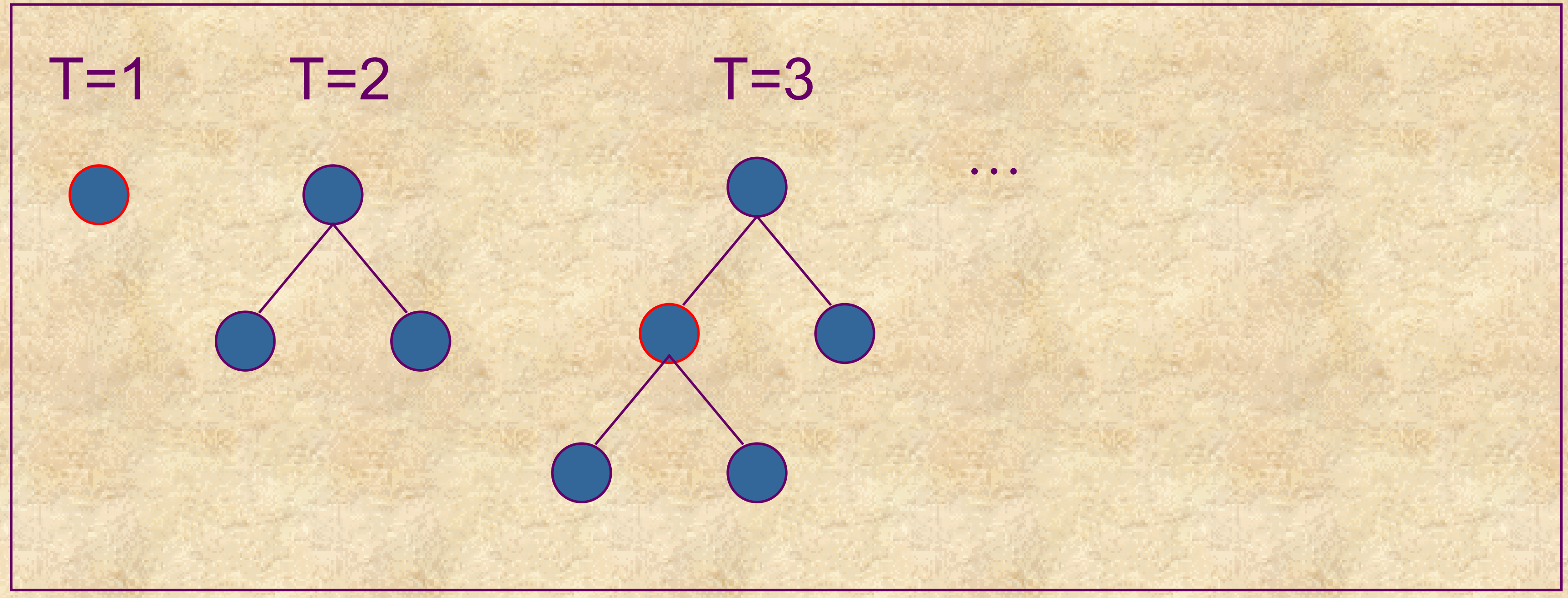
- Algoritmo de Divisão Recursiva
  - Objetivo: construir uma árvore binária que fornecerá dois tipos de informação
    - Uma descrição de cada uma das m classes a priori na forma de uma conjunção de propriedades
    - Uma regra de decisão d capaz de classificar um novo objeto em uma das m classes a priori
  - Dois estágios para obter esses objetivos
    - Estratégia de construção da árvore
    - Seleção da melhor sub-árvore

Árvores de Decisão: Dados Probabilistas

- Estratégia de construção da árvore
  - Em um dado passo do algoritmo, a estratégia consiste em dividir o nó terminal que fornece o melhor critério
  - O passo T do algoritmo consiste no exame dos T-1 nós terminais de A_{T-1} para selecionar a melhor divisão binária que produz a melhor árvore A_T com T nós terminais
  - Esse procedimento é aplicado recursivamente até ue uma condição de parada seja satisfeita

Árvores de Decisão: Dados Probabilistas

- Estratégia de construção da árvore
  - Constrói-se uma seqüência de árvores binárias encaixadas $A_1, \ldots, A_T, \ldots, A_{max}$



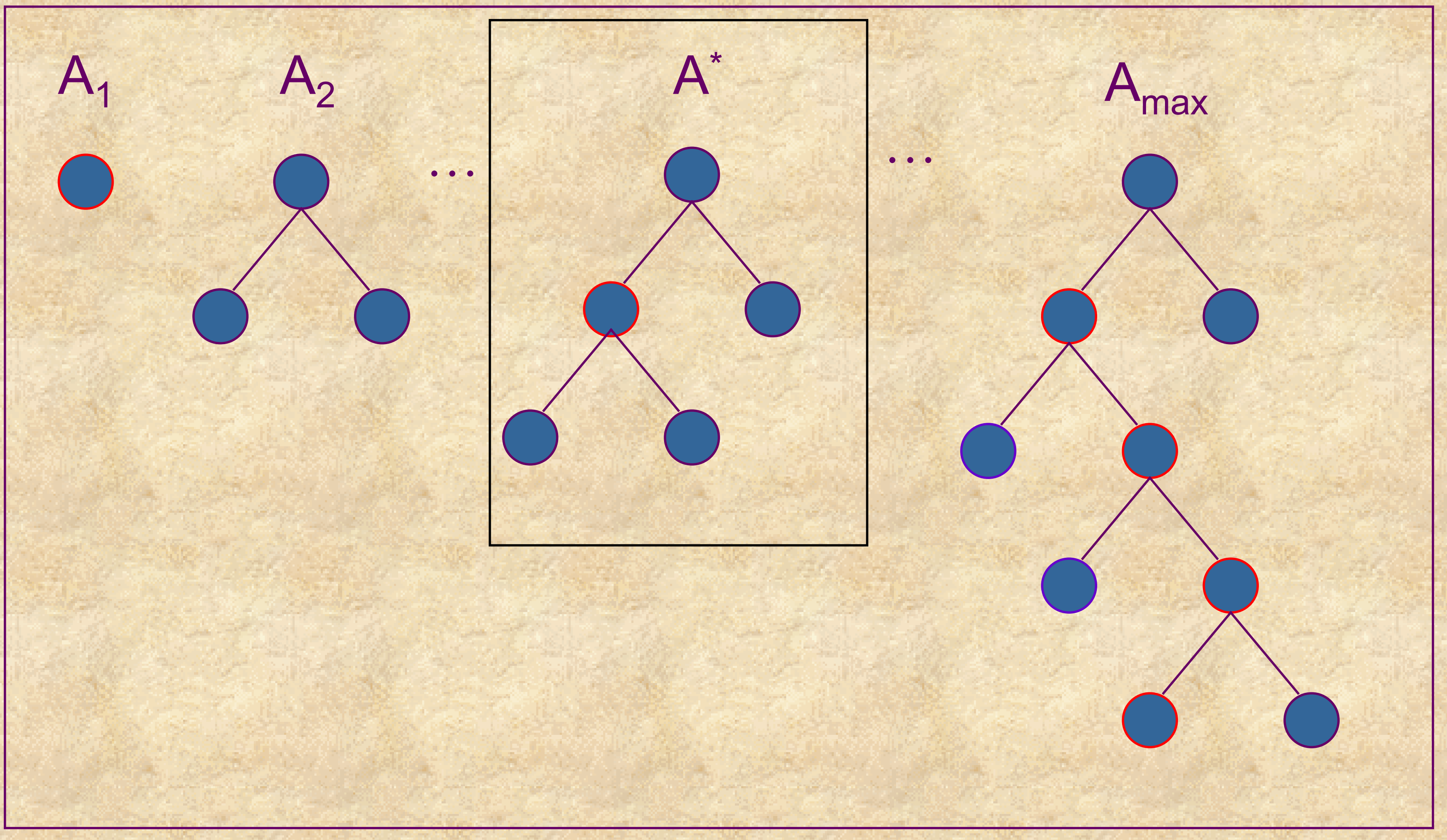
Árvores de Decisão: Dados Probabilistas

- Seleção da melhor sub-árvore
  - Objetivo: fornecer uma árvore A^* associada com a melhor regra de decisão
  - A qualidade de uma regra de decisão é função da taxa de acertos.
  - $R(A_T)$ é a taxa associada a árvore A_T
  - Esse estágio consiste em selecionar a melhor sub-árvore na seqüência prévia de arvores encaixadas

$$A^* = \underset{A \in \{A_1, \ldots, A_{\max}\}}{\arg\min} R(A)$$

Árvores de Decisão: Dados Probabilistas

- Seleção da melhor sub-árvore



Árvores de Decisão: Dados Probabilistas

- Notação
  - T_A: conjunto de nós terminais admissíveis para a divisão em um dado estágio do algoritmo (nós para os quais a condição de parada não é satisfeita)
  - T_D: conjunto de nós terminais definitivos durante a fase de crescimento da árvore (que satisfazem a condição de parada)
  - T_C: conjunto de todos os nós terminais em um dado estágio do algoritmo: $T_C = T_A \cup T_D$. Consideraremos também o conjunto de nós terminais correntes em um dado estágio do algoritmo

Árvores de Decisão: Dados Probabilistas

- Notação
  - $Q_A(t)$: conjunto de consultas admissíveis em um nó t.
  - $Q_C(t)$: conjunto de consultas candidatas em um nó t. Consultas binárias cujos nós filhos “não são tão pequenos” e que fornecem “um mínimo de informação”
  - $P_t(i)$: probabilidade condicional de observar a classe i no nó t
  - $p_k(t)$: probabilidade de pertinência do objeto k ao nó i
  - p_{kq}: probabilidade que k satisfaça a consulta q

Árvores de Decisão: Dados Probabilistas

- Notação
  - $W(t,q)$: informação fornecida pela consulta q no nó t (qualidade da divisão binária produzida por q)
  - q^*_t: melhor consulta no nó t
  - q^*: melhor consulta entre todos os nós terminais em uma dada etapa da divisão
  - l, n_l: nó esquerdo criado pela divisão e seu efetivo associado
  - r, n_r: nó direito criado pela divisão e seu efetivo associado

Árvores de Decisão: Dados Probabilistas

- Algoritmo Geral da Árvore de Decisão
- Inicie com a raíz
- WHILE[1] (o tamanho da árvore é admissível) AND (existe nós terminais admissíveis) $T_A \neq \emptyset$
  - (para cada nó admissível t) $\forall t \in T_A$
    - (para cada questão admissível q no nó t) $\forall q \in Q_A(t)$
      - [2] Divida t em dois nós terminais temporários l e r: $T_C \leftarrow T_C \setminus \{t\} \cup \{l,r\}$
      - [3] calcule os efetivos dos nós l e r: n_l e n_r
      - IF [4] ambos os nós são admissíveis: $(n_l > \text{limiar})$ AND $(n_r > \text{limiar})$

Árvores de Decisão: Dados Probabilistas

- Algoritmo Geral da Árvore de Decisão
      - (5) calcule a qualidade da divisão: $W(t,q)$
      - IF a qualidade é suficiente: $W(t,q) > \text{limiar}$
      - THEN q é candidato para o nó t: $q \in Q_C(t)$
      - ELSE renuncie a consulta q
    - ELSE renuncie a questão q
  - END q

Árvores de Decisão: Dados Probabilistas

- Algoritmo Geral da Árvore de Decisão
    - IF não existe consulta candidata no nó t: $Q_C(t)=\emptyset$
        - THEN t é um nó terminal: $T_D \leftarrow T_D \cup \{t\}$
    - ELSE (6) selecione a melhor questão no nó t: q^*_t
- END t
- IF não existe nó t para dividir: $Q_C(t)=\emptyset, \forall t \in T_A$
    - THEN pare
- ELSE (7) selecione a melhor divisão no melhor nó: q^*

Árvores de Decisão: Dados Probabilistas

- Atualize os nós da árvore
  - IF o efetivo do nó filho l é suficiente AND l não é um nó puro THEN $T_A \leftarrow T_A \cup \{l\}$
  - ELSE l é um nó terminal: $T_D \leftarrow T_D \cup \{l\}$
  - IF o efetivo do nó filho r é suficiente AND r não é um nó puro THEN $T_A \leftarrow T_A \cup \{r\}$
  - ELSE r é um nó terminal: $T_D \leftarrow T_D \cup \{r\}$
  - $T_C \leftarrow T_A \cup T_D$

Árvores de Decisão: Dados Probabilistas

- (8) Calcule a informação descritiva associada aos novos nós
- (9) Estime a taxa R(A) para a árvore corrente

☛ END WHILE

Árvores de Decisão: Dados Probabilistas

- Descrição detalhada das diferentes etapas
- Tamanho e admissibilidade dos nós filhos *l,r* com os pesos n_l, n_r (passo 4 do algoritmo)
  - calculo da probabilidade de pertinência de um objeto k aos nós filhos *l,r* do nó *t*

$$p_k(l) = p_k(t)\, p_{kq}$$

$$p_k(r) = p_k(t)\left(1 - p_{kq}\right)$$

  - No primeiro passo do algoritmo, $p_k(t) = 1$, $k = 1, \ldots, n$, pois todos os indivíduos pertencem completamente a raiz
  - nos passos seguintes as probabilidades p_{kq} são computadas uma vez escolhida a melhor divisão

Árvores de Decisão: Dados Probabilistas

- Tamanho e admissibilidade dos nós filhos *l*,*r* com os pesos n_l, n_r (passo 4 do algoritmo)
  - calculo de n_l, n_r

$$n_l = \sum_{k=1}^{n} p_k(l)$$

$$n_r = \sum_{k=1}^{n} p_k(r)$$

Árvores de Decisão: Dados Probabilistas

- Tamanho e admissibilidade dos nós filhos l,r com os pesos n_l,n_r (passo 4 do algoritmo)
  - Condição de admissibilidade para o tamanho das classes n_l,n_r
    A qualidade da divisão binária q do nó t (a ser definida) é calculada apenas se

$$n_l \geq \text{limiar minimo} \quad \text{e} \quad n_r \geq \text{limiar minimo}$$

onde *limiar mínimo* é escolhido a priori

Árvores de Decisão: Dados Probabilistas

- Busca da melhor divisão (passos 6 e 7 do algoritmo)
  - Busca da melhor questão *q* (melhor divisão) em um dado nó *t*
  - Passo preliminar: atualização do conjunto de nós terminais correntes T_C: dois nós filhos *l* e *r* substituem o nó t
  - A qualidade da divisão é calculada em relação ao novo conjunto

$$T_C \leftarrow T_C \setminus \{t\} \cup \{l, r\}$$

Árvores de Decisão: Dados Probabilistas

- Qualidade *W(t,q)* de uma divisão *q* no nó *t* (passo 5 do algoritmo)
  - A qualidade da divisão induzida pela questão *q* no nó *t* é definida pela seguinte log-verossimilhança

$$W(t,q) = \log \prod_{k=1}^{n} \sum_{s \in T_C} p_k(s) P_s(c_k)$$

$p_k(s)$: probabilidade de que o objeto k seja afetado ao nó s

$P_s(c_k)$ é a probabilidade de que a classe $C(k)=c_k$ é observada, condicionalmente ao nó *s*

Árvores de Decisão: Dados Probabilistas

- Qualidade *W(t,q)* de uma divisão *q* no nó *t* (passo 5 do algoritmo)

$$P_s(c_k) = P_s(i) \quad \text{se}\; k \in C_i$$

ou

$$P_s(c_k) = P_s(1)^{c_{k1}} \times \ldots \times P_s(K)^{c_{km}} \quad \text{onde} \quad c_{ki} = 1 \Leftrightarrow k \in C_i$$

  - Como o interesse é apenas na informação associada com os dois nós filhos

$$W(t,q) = \log \prod_{k=1}^{n} \left(p_k(l) P_l(c_k) + p_k(r) P_r(c_k) + \lambda_k \right)$$

Árvores de Decisão: Dados Probabilistas

- Qualidade *W(t,q)* de uma divisão *q* no nó *t* (etapa 5 do algoritmo)
  - λ_k é fixo e depende apenas de $p_k(s)$ e $P_s(c_k)$ associadas aos nós *s* previamente definidos, mas que não são essenciais no passo de divisão

$$\lambda_s = \sum_{\substack{s \in T_C \\ s \notin \{l,r\}}} p_k(s) P_s(c_k)$$

  - Para calcular a qualidade da divisão é necessário estimar, para cada *t* e *q* fixados, as probabilidades condicionais $P_l(1),\ldots, P_l(m)$ e $P_r(1),\ldots, P_r(m)$ das m classes

Árvores de Decisão: Dados Probabilistas

- O Algoritmo EM para estimar as probabilidades condicionais $P_l(i), P_r(i)$
  - Idéia básica: encontrar os valores dos parâmetros que maximizam a verossimilhança iterativamente
  - Os parâmetros são as probabilidades condicionais

$$P_l^h(i) \quad \text{e} \quad P_r^h(i), \quad i = 1, \ldots, K$$

  - Cada iteração do Algoritmo EM consiste na repetição de duas etapas: Expectation (E) e Maximisation (M)

Árvores de Decisão: Dados Probabilistas

- O Algoritmo EM para estimar as probabilidades condicionais $P_l(i), P_r(i)$
  - Etapa E: Calcula-se a probabilidade de que k pertence ao nó l e r na iteração (h) do algoritmo

$$E^{(h)}(k,l) = \frac{P_l^{(h)}(c_k)\,p_k(l)}{P_l^{(h)}(c_k)\,p_k(l) + \lambda_k}, \quad k = 1,\ldots,n$$

$$E^{(h)}(k,r) = \frac{P_r^{(h)}(c_k)\,p_k(r)}{P_r^{(h)}(c_k)\,p_k(r) + \lambda_k}, \quad k = 1,\ldots,n$$

Árvores de Decisão: Dados Probabilistas

- O Algoritmo EM para estimar as probabilidades condicionais $P_l(i), P_r(i)$
  - Etapa M: Calcula-se as novas probabilidades condicionais

$$P_l^{(h+1)}(i) = \frac{\sum_{k=1}^{n} c_{ki} E^{(h)}(k,l)}{\sum_{k=1}^{n} E^{(h)}(k,l)} = \frac{\sum_{k \in C_i} E^{(h)}(k,l)}{\sum_{k=1}^{n} E^{(h)}(k,l)}, \quad i = 1,\ldots,m$$

$$P_r^{(h+1)}(i) = \frac{\sum_{k=1}^{n} c_{ki} E^{(h)}(k,r)}{\sum_{k=1}^{n} E^{(h)}(k,r)} = \frac{\sum_{k \in C_i} E^{(h)}(k,r)}{\sum_{k=1}^{n} E^{(h)}(k,r)}, \quad i = 1,\ldots,m$$

Árvores de Decisão: Dados Probabilistas

- O Algoritmo EM para estimar as probabilidades condicionais $P_l(i), P_r(i)$
  - Calcula-se o novo valor do critério $W^{(h+1)}(t,s)$ usando-se as estimativas das probabilidades condicionais $P_l^{(h+1)}$ e $P_r^{(h+1)}$
  - Regra de Parada: O algoritmo EM para quando

$$W^{(h+1)}(t,q) - W^{(h)}(t,q) < \varepsilon$$

onde $\varepsilon > 0$ é próximo de zero

Árvores de Decisão: Dados Probabilistas

- O Algoritmo EM para estimar as probabilidades condicionais $P_l(i), P_r(i)$
  - Valores iniciais das probabilidades condicionais para o algoritmo EM

$$P_l^{(0)}(i) = \frac{\sum_{k=1}^{n} c_{ki} p_k(l)}{\sum_{k=1}^{n} p_k(l)} = \frac{\sum_{k \in C_i} p_k(l)}{n_l} \quad i = 1, \ldots, m$$

$$P_r^{(0)}(i) = \frac{\sum_{k=1}^{n} c_{ki} p_k(r)}{\sum_{k=1}^{n} p_k(r)} = \frac{\sum_{k \in C_i} p_k(r)}{n_r} \quad i = 1, \ldots, m$$

Árvores de Decisão: Dados Probabilistas

- O Algoritmo EM para estimar as probabilidades condicionais $P_l(i),P_r(i)$
  - Condições de admissibilidade baseada na qualidade da divisão
    Guarda-se provisoriamente a questão q se a informação W(t,q) é maior do que um dado limiar

$$\text{if} \quad [W(t,q) > \text{limiar min}] \quad \text{então} \quad [q \in Q_C(t)]$$

diz-se então que q é um candidato para o nó t

Árvores de Decisão: Dados Probabilistas

☛ Seleção da melhor divisão

Essa seleção é realizada em duas etapas:

- Seleção da melhor questão q em cada nó terminal t: q_t^* (etapa 6 do algoritmo)
- Seleção da melhor questão q_t^* entre todos os nós terminais t: q^* (etapa 7 do algoritmo)

- Em conseqüência, a melhor divisão é aquela que produz a melhor árvore de (T+1) nós, dada uma árvore de T nós
- "Melhor" significa um valor de máximo verossimilhança W

Árvores de Decisão: Dados Probabilistas

- Informações Descritivas calculadas para os novos nós (etapa 8 do algoritmo)
  - Os seguintes tipos de informação são calculados para cada novo nó *t*
  - *Informação relacionada com a descrição do nó* $[Y_j \in V]$, $[Y_j < c]$, …
  - *Informações relativas aos indivíduos*
    - Probabilidade de pertinência dos objetos k ao nó t: $p_k(t)$
    - O conjunto de indivíduos afetados a *t* com base na regra da maioria

$$\text{no}(t) = \{k \mid p_k(t) > p_k(s), \forall s \neq t\}$$

Árvores de Decisão: Dados Probabilistas

- Informações Descritivas calculadas para os novos nós (etapa 8 do algoritmo)
  - *Informações das Classes de uma dada partição a priori* Define-se a seguinte tabela que fornece informações gerais sobre as K classes a priori no nó *t*:

	Tamanho	Prob.	d
classe 1	$n_t(1)$	$P_t(1)$	
…	…	…	
classe i	$n_t(i)$	$P_t(i)$	♣
…	…	…	
classe m	$n_t(m)$	$P_t(m)$	

Árvores de Decisão: Dados Probabilistas

- Informações Descritivas calculadas para os novos nós (etapa 8 do algoritmo)
  - *Informações das Classes de uma dada partição a priori*
    Onde
    - $P_t(i)$: probabilidade condicional de observar a classe i no nó *t*
    - $n_t(i)$: tamanho da classe *i* no nó *t*: isso é calculado com base na lista de objetos afetados ao nó *t*: $n_t(i)=\{k \mid k \in$ nó(t) e $k \in C_i\}$
    - a ultima coluna indica a classe majoritária no nó (♣) segundo as probabilidades condicionais: $[d(t) = i] \Leftrightarrow [P_t(i) > P_t(l), \forall l \neq i]$

Árvores de Decisão: Dados Probabilistas

- Informações Descritivas calculadas para os novos nós (etapa 8 do algoritmo)
  - *Informações das Classes de uma dada partição a priori* Onde
    - O cálculo do tamanho total do nó n_t e a probabilidade p(t) de alcança-lo é como segue:

$$n(t) = \sum_{i=1}^{m} n_t(i)$$

$$p(t) = \sum_{k=1}^{n} p_k(t)\, p(k) = \frac{1}{n} \sum_{k=1}^{n} p_k(t)$$

Supõe-se que cada objeto tem um peso p(k) = 1/n

Árvores de Decisão: Dados Probabilistas

Condições de Parada

- O crescimento da árvore é interrompido se ao menos uma das condições abaixo for satisfeita:
  - A árvore é muito grande: card(T_C) > limiar max
  - Não é possível realizar novas divisões: $T_A = \emptyset$

    Essa condição ocorre quando os nós terminais são puros (inclui indivíduos de apenas uma classe) ou quando o tamanho deles é muito pequeno

Árvores de Decisão: Dados Probabilistas

- Componentes da Decisão
  - Um dos objetivos desse método é fornecer uma regra de decisão que é capaz de classificar novos objetos com um mínimo de erros
  - Para isso é necessário selecionar a melhor sub-árvore (e a sua regra de decisão associada) dentre a seqüência de árvores embutidas.
  - Questões preliminares:
    - Como construir uma regra de decisão
    - Como medir a qualidade da mesma

Árvores de Decisão: Dados Probabilistas

☛ Regra de Decisão

- Seja D o conjunto de todos os vetores de descrições simbólicas relevantes para a situação em estudo
- Uma regra de decisão *d* associada com uma árvore binária é uma função

$$d : D \to \{1,\ldots,i,\ldots,m\}$$

que permite afetar cada objeto *k* de vetor de descrição $x_k \in D$ a uma classe *i* da partição a priori.

- Nesse método, vale a seguinte regra:

$$[d(x_k) = i] \quad \Leftrightarrow \quad [p_k(i) > p_k(j) \quad \text{para} \quad j = 1,\ldots,m,\ j \neq i]$$

Árvores de Decisão: Dados Probabilistas

- Regra de Decisão
  - A probabilidade de pertinência $p_k(i)$ é calculada como a média ponderada (em relação a todos os nós terminais) das probabilidades condicionais $P_t(i)$, com pesos $p_k(t)$

$$p_k(i) = \sum_{t=1}^{T} P_t(i)\, p_k(t)$$

  - Ela consiste na soma das probabilidades de que um individuo pertença a classe *i*, para todos os nós terminais *t*, essas probabilidades sendo ponderadas pelas probabilidades de que esse objeto alcance as diferentes folhas

Árvores de Decisão: Dados Probabilistas

- Qualidade de uma regra de decisão d
  - Taxa de previsões corretas
    A qualidade da regra pode ser definida através de uma *taxa de previsões corretas*

$$R(A) = \frac{n_A}{n_t}$$

onde n_A é número de objetos corretamente classificados pela árvore A e n_t é o número total de objetos a ser classificado

Árvores de Decisão: Dados Probabilistas

- Qualidade de uma regra de decisão d
  - Taxa de previsões corretas
    Um objeto k pertencente a classe *i* é dito ser *corretamente classificado* pela regra de decisão *d* associada com a árvore A se

$$[d(x_k) = i] \quad \text{e} \quad [\mathrm{c}_\mathrm{k} = i]$$

    são ambos verdadeiros, isto é, se *k* é corretamente afetado a classe i com base no seu vetor de descrição x_k

Árvores de Decisão: Dados Probabilistas

- Qualidade de uma regra de decisão d
  - População ou Amostra
    Duas situações consideradas
    - o conjunto de objetos pode ser considerado como uma população exaustiva.
      Nesse caso é calculada a taxa de previsão correta R(A) classificando-se os n (= n_t) objetos de E
    - o conjunto de objetos pode ser considerado como uma amostra de uma população mais geral
      Nesse caso, a taxa R(A) é calculada com base em um conjunto de teste: n_t é o tamanho do conjunto de teste e n_A é igual a totalidade de casos que são classificados corretamente

Árvores de Decisão: Dados Probabilistas

- Qualidade de uma regra de decisão d
  - Seleção de sub-árvores
    Tendo-se definido a qualidade de uma árvore, é possível selecionar a “melhor árvore” A^* da seqüência de árvores encaixadas $A_1, \ldots, A_T, \ldots, A_{max}$, isto é, a árvore associada com a maior taxa R(A)
    - Tratando-se de uma população, o valor máximo de R(A) é alcançado quando $A = A_{max}$

      Refinando-se a árvore binária sempre se consegue melhorar a taxa de previsões corretas no conjunto de aprendizagem

Árvores de Decisão: Dados Probabilistas

- Qualidade de uma regra de decisão d
  - Seleção de sub-árvores
    - Tratando-se de uma amostra, a melhor árvore binária geralmente é muito menor do que A_{max} por conta do sobre-ajustamento

      Após uma dada etapa do algoritmo, o crescimento da árvore leva a uma regra de decisão mais eficiente no conjunto de treinamento e uma degradação da sua qualidade em relação a toda população

Exemplo 3

Tabela de dados probabilisticos

K_k	Y_1			Y_2			Y_3		
	1	2	3	1	2	3	1	2	3
1	0.10	0.30	0.60	0.40	0.30	0.30	0.05	0.70	0.25
2	0.20	0.40	0.40	0.10	0.50	0.40	0.10	0.80	0.10
3	0.50	0.20	0.30	0.15	0.35	0.50	0.40	0.20	0.40
4	0.40	0.10	0.50	0.30	0.30	0.40	0.35	0.20	0.45

Tabela de questões binárias

	1	2	3	4	5	6
K	$Y_1 \leq 1$	$Y_1 \leq 2$	$Y_2 \leq 1$	$Y_2 \leq 2$	$Y_3 \leq 1$	$Y_3 \leq 2$
1	0.10	0.40	0.40	0.70	0.05	0.75
2	0.20	0.60	0.10	0.60	0.10	0.90
3	0.50	0.70	0.15	0.50	0.40	0.60
4	0.40	0.50	0.30	0.60	0.35	0.55

Etapa 1

$t = 1, \quad Y_1 \le 1$

$$p_k(e) = p_k(t) \cdot p_{kq} \quad k = 1,2,3,4 \qquad p_k(d) = p_k(t) \cdot (1 - p_{kq}) \quad k = 1,2,3,4$$

$$p_1(e) = p_1(t) \times p_{1q} = 1 \times 0.10 = 0.10$$

$$p_1(d) = p_1(t) \times (1 - p_{1q}) = 1 \times 0.90 = 0.9($$

$$p_2(e) = p_2(t) p_{2q} = 1 \times 0.20 = 0.20$$

$$p_2(d) = p_2(t) \cdot (1 - p_{2q}) = 1 \times 0.80 = 0.8($$

$$p_3(e) = p_3(t) \cdot p_{3q} = 1 \times 0.50 = 0.50$$

$$p_3(d) = p_3(t) \cdot (1 - p_{3q}) = 1 \times 0.50 = 0.5($$

$$p_4(e) = p_4(t) \cdot p_{4q} = 1 \times 0.40 = 0.40$$

$$p_4(d) = p_4(t) \cdot (1 - p_{4q}) = 1 \times 0.60 = 0.6($$

$$P_e^0(1)=\frac{\sum_{k\in C_1} p_k(e)}{\sum_{k=1}^{n} p_k(e)}=\frac{0.10+0.20}{0.1+0.20+0.5+0.40}=0.25$$

$$P_e^0(2)=1-P_e^0(1)=0.75$$

$$P_d^0(1)=\frac{\sum_{k\in C_2} p_k(e)}{\sum_{k=1}^{n} p_k(e)}=\frac{0.90+0.80}{0.90+0.80+0.50+0.60}=0.60$$

$$P_d^0(2)=1-P_d^0(1)=0.40$$

$$\begin{aligned}
W(t,q) &= \log\prod_{k=1}^{n}(p_k(e)\cdot P_e(c_k)+p_k(d)\cdot P_d(c_k)) \\
&= \log(\,(0.10\times 0.25+0.90\times 0.60)\cdot \\
&\qquad (0.20\times 0.25+0.80\times 0.60)\cdot \\
&\qquad (0.50\times 0.75+0.50\times 0.40)\cdot \\
&\qquad (0.40\times 0.75+0.60\times 0.40)\cdot \\
&= \log(0.0913)=-2.394
\end{aligned}$$

Resultados de W(t,q) para Etapa 1

$Y_1 \leq 1$
W(t,q) =-2.394

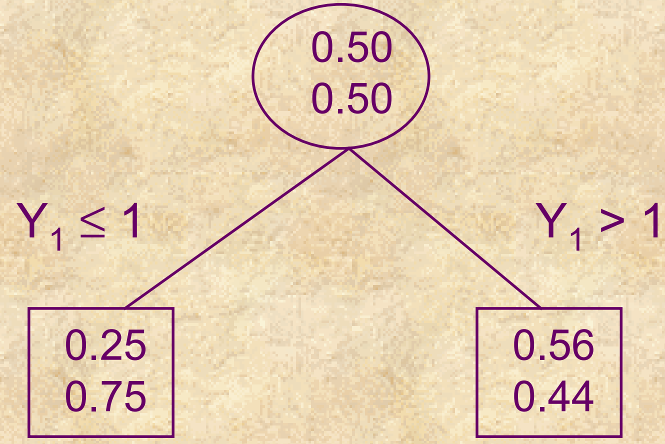
$Y_1 \leq 2$
W(t,q) =-2.7806

$Y_2 \leq 1$
W(t,q) =-2.975

$Y_2 \leq 2$
W(t,q) =-2.682

$Y_3 \leq 1$
W(t,q) =-2.763

$Y_3 \leq 2$
W(t,q) =-2.521



$Y_1 \leq 1$

$$P_t(1) = \frac{0.10 + 0.20}{0.10 + 0.20 + 0.50 + 0.40} = 0.25$$

$$P_t(2) = 1 - 0.25 = 0.75$$

$Y_1 > 1$

$$P_t(1) = \frac{0.30 + 0.60 + 0.40 + 0.40}{0.30 + 0.60 + 0.4 + 0.40 + 0.20 + 0.30 + 0.50 + 0.10} = 0.56$$

$$P_t(2) = 1 - 0.56 = 0.44$$

Cálculo do R(A)

$$p_1(1) = 0.10 \times 0.25 + 0.90 \times 0.56 = 0.529$$

Objeto 1 $\rightarrow$Classe 1

$$p_1(2) = 0.10 \times 0.75 + 0.90 \times 0.44 = 0.471$$

$$p_2(1) = 0.498$$

Objeto 2 $\rightarrow$Classe 2

$$p_2(2) = 0.502$$

$$p_3(1) = 0.40$$

Objeto 3 $\rightarrow$Classe 2

$$p_2(2) = 0.502$$

$$p_4(1) = 0.43$$

Objeto 4 $\rightarrow$Classe 2

$$p_4(2) = 0.77$$

R(A) =75%

Etapa 2

Partição $Y_1 \leq 1$

$Y_2 \leq 1$

$$p_1(e) = p_1(t) \cdot p_{1q} = 0.10 \times 0.4 = 0.04$$

$$p_1(d) = p_1(t) \cdot (1 - p_{1q}) = 0.10 \times 0.60 = 0.06$$

$$p_3(e) = p_3(t) \cdot p_{3q} = 0.075$$

$$p_3(d) = p_3(t) \cdot (1 - p_{3q}) = 0.425$$

$$p_2(e) = p_2(t) \cdot p_{2q} = 0.20 \times 0.10 = 0.02$$

$$p_2(d) = p_2(t) \cdot (1 - p_{2q}) = 0.20 \times 0.90 = 0.18$$

$$p_4(e) = p_4(t) \cdot p_{4q} = 0.12$$

$$p_4(d) = p_4(t) \cdot (1 - p_{4q}) = 0.28$$

$$P_e^0(1) = \frac{0.04 + 0.02}{0.04 + 0.02 + 0.075 + 0.12} = 0.34 \qquad P_e^0(2) = 1 - P_e^0(1) = 0.56$$

$$P_d^0(1) = 0.25 \qquad P_d^0(2) = 1 - P_d^0(2) = 0.75$$

W(t,q) = -8.873

Resultados de W(t,q) para Etapa 2

Partição $Y_1 \leq 1$

$Y_2 \leq 1$
W(t,q) =-8.873

$Y_2 \leq 2$
W(t,q) =-8.805

$Y_3 \leq 1$
W(t,q) =-8.740

$Y_3 \leq 2$
W(t,q) =-6.265

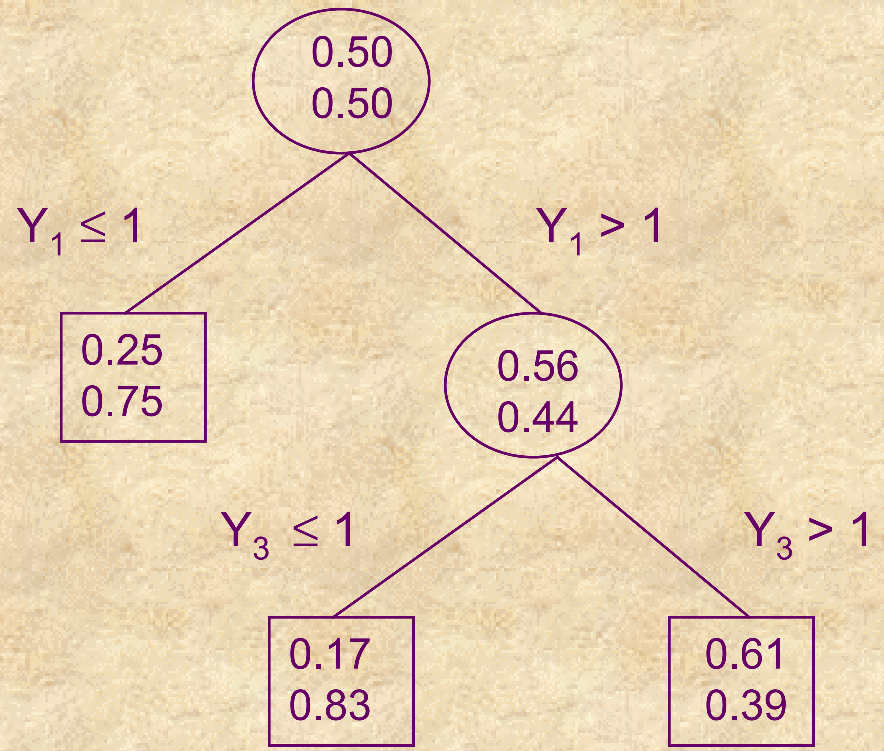
Partição $Y_1 > 1$

$Y_2 \leq 1$
W(t,q) =-4.398

$Y_2 \leq 2$
W(t,q) =-3.963

$Y_3 \leq 1$
W(t,q) =-3.849

$Y_3 \leq 2$
W(t,q) =-4.135



Classe 1 [$Y_1 \in [2,3] \wedge Y_3 \in [2,3]$]

Classe 2 [$Y_1 \in [0,1]$]

[$Y_1 \in [2,3] \wedge Y_3 \in [0,1]$]

Cálculo do R(A)

$$p_1(1) = 0.25 \times 0.10 + 0.61 \times [(0.30 + 0.60) \times (0.70 + 0.25) + 0.17 \times [(0.30 + 0.60) \times 0.05] = 0.55$$

Objeto 1 $\rightarrow$Classe 1

$$p_1(2) = 0.45$$

$$p_2(1) = 0.504$$

Objeto 2 $\rightarrow$Classe 2

$$p_2(2) = 0.496$$

$$p_3(1) = 0.342$$

Objeto 3 $\rightarrow$Classe 2

$$p_2(2) = 0.658$$

$$p_4(1) = 0.369$$

Objeto 4 $\rightarrow$Classe 2

$$p_4(2) = 0.631$$

R(A) =100%