Modelos de Markov Ocultos no Reconhecimento da Língua LIBRAS

Prof. Dr. Hemerson Pistori

GPEC - Grupo de Pesquisa em Engenharia e Computação
UCDB - Universidade Católica Dom Bosco

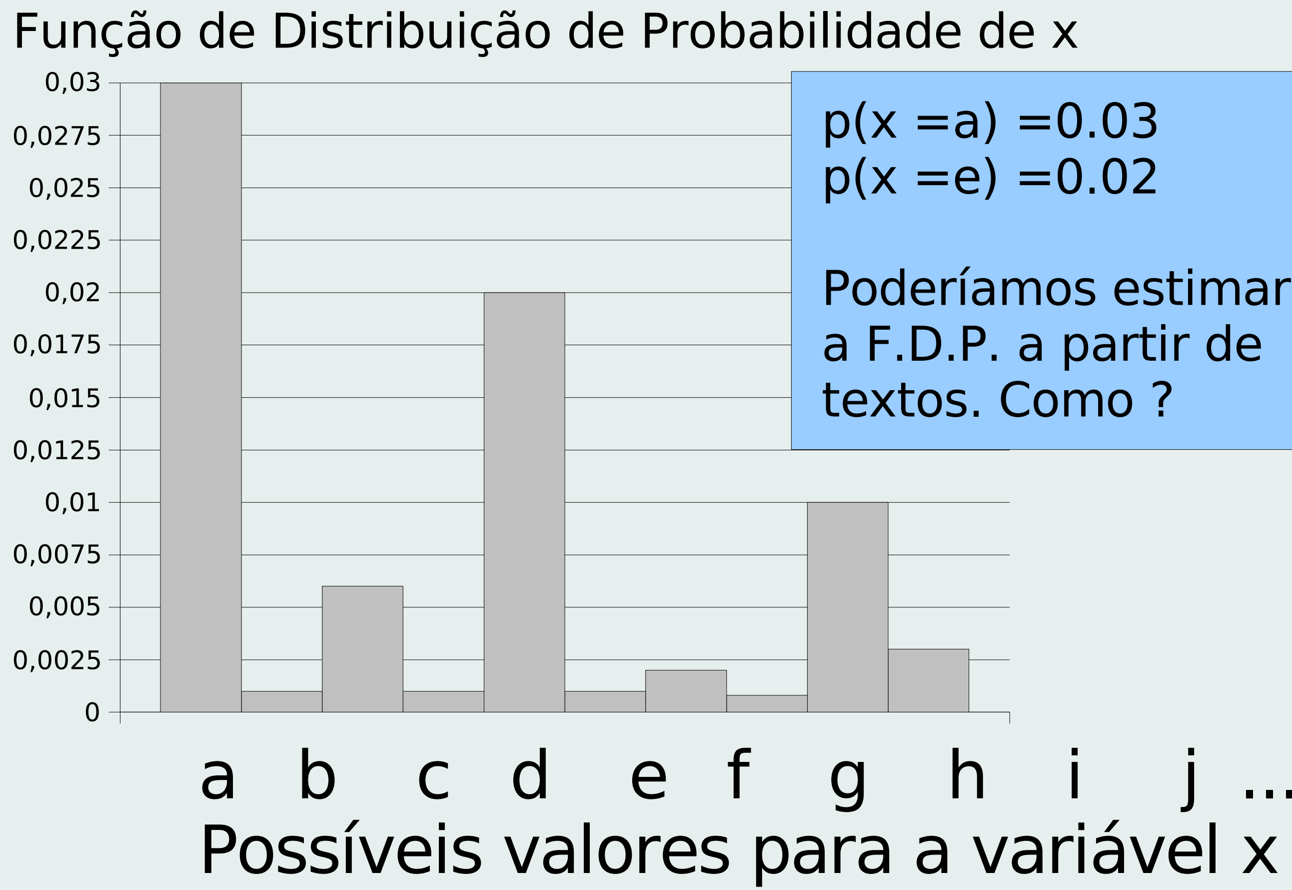
Sumário

- Conceitos Fundamentais de Probabilidade e Estatística
- Redes Bayesianas
- Processos Estocásticos e Modelos de Markov
- Reconhecimento de Língua de Sinais
- Modelos de Markov Ocultos
- Principais Algoritmos
- Conclusão

Conceitos Fundamentais

- Variáveis Aleatórias Discretas
- Funções de Distribuição de Probabilidade
- Probabilidade Condicional
- Independência
- Distribuição Conjunta

Variáveis Aleatórias Discretas



p(x =a) =0.03
p(x =e) =0.02

Poderíamos estimar a F.D.P. a partir de textos. Como ?

Variável Aleatória Discreta

Exemplo: **x** – ocorrência de uma determinada letra do alfabeto em um texto escrito em português.

Probabilidade Condicional e Independência

$\mathbf{x_k}$ – ocorrência de uma letra na posição k

$\mathbf{x_{k-1}}$ – ocorrência de uma letra na posição k-1

$$p(\mathbf{x_k} = \text{'a'}) > p(\mathbf{x_k} = \text{'a'} \mid \mathbf{x_{k-1}} = \text{'s'}) \quad ?$$

$$p(\mathbf{x_k} = \text{'a'} \mid \mathbf{x_{k-1}} = \text{'s'}) > p(\mathbf{x_k} = \text{'a'} \mid \mathbf{x_{k-1}} = \text{'a'}) \quad ?$$

$$p(\mathbf{x_k} = \text{'a'} | \mathbf{x_{k-1}} = \text{'z'}) > p(\mathbf{x_k} = \text{'a'} | \mathbf{x_{k-1}} = \text{'z'}, \mathbf{x_{k-2}} = \text{'z'}, \mathbf{x_{k-3}} = \text{'i'}, \mathbf{x_{k-4}} = \text{'p'})?$$

$$p(\mathbf{x_k} = \text{'a'}) > p(\mathbf{x_k} = \text{'a'} | \mathbf{x_{k-2000}} = \text{'s'}) \quad ?$$

Probabilidade Condicional e Independência

$\mathbf{x_s}$ – sexo ('f' – feminino, 'm' – masculino')

$\mathbf{x_c}$ – curso ('c' – computação, 'd' – direito, 'l' - letras)

	MD	
MC	MD	
MC	MD	FL
MC	FD	FL
FC	FD	
	FD	

Situação 1

$p(\mathbf{x_s} = \text{'f'}) = 1/2$

$p(\mathbf{x_c} = \text{'c'}) = 1/3$

$p(\mathbf{x_s} = \text{'f'} | \mathbf{x_c} = \text{'c'}) = 1/4$

	MD	
MC	MD	
MC	MD	ML
FC	FD	FL
FC	FD	
	FD	

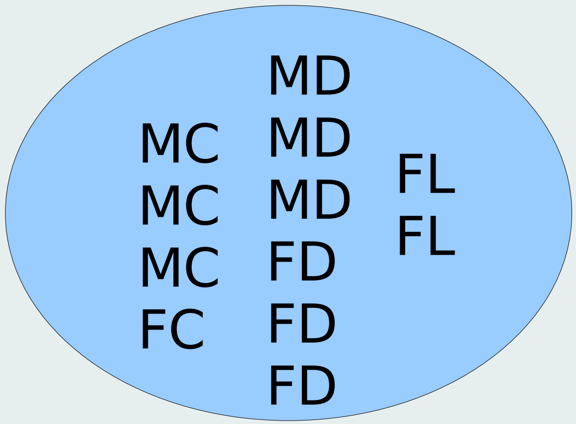
Situação 2

$p(\mathbf{x_s} = \text{'f'}) = 1/2$

$p(\mathbf{x_c} = \text{'c'}) = 1/3$

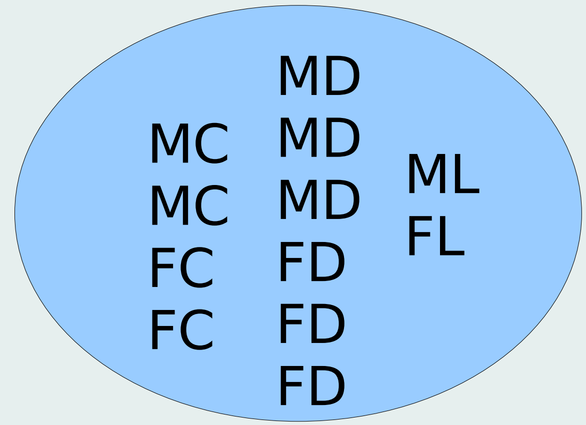
$p(\mathbf{x_s} = \text{'f'} | \mathbf{x_c} = \text{'c'}) = 1/2$

Probabilidade Condicional e Independência



Situação 1

$p(\mathbf{x_s} = \text{'f'} | \mathbf{x_c} = \text{'c'}) = 1/4 =$
$(1/12) / (4/12) \quad =$
$p(\mathbf{x_s} = \text{'f'}, \mathbf{x_c} = \text{'c'}) / p(\mathbf{x_s} = \text{'c'})$



Situação 2

$p(\mathbf{x_s} = \text{'f'} | \mathbf{x_c} = \text{'c'}) = 1/2 =$
$p(\mathbf{x_s} = \text{'f'})$

Caso geral.......: p(x|y) = p(x,y) / p(y)
Independência.: p(x|y) = p(x)

Outro exemplo: p(x|y1,y2,y3,y4,y5) = p(x|y1,y2)

Distribuição Conjunta

MD
MC MD FL
MC MD FL
MC FD
FC FD
FD

$$p(\mathbf{x_s} = \text{'f'}, \mathbf{x_c} = \text{'l'})$$

	Masc.	Femin.	
Comput.	0,25	0,08	0,33
Direito	0,25	0,25	0,5
Letra	0	0,17	0,17
	0,5	0,5	

$$p(\mathbf{x_s} = \text{'m'}) = p(\mathbf{x_s} = \text{'m'}, \mathbf{x_c} = \text{'c'}) + p(\mathbf{x_s} = \text{'m'}, \mathbf{x_c} = \text{'d'}) + p(\mathbf{x_s} = \text{'m'}, \mathbf{x_c} = \text{'l'})$$

Distribuição Conjunta

$$p(\mathbf{x_k} = 'a' | \mathbf{x_{k-1}} = 'z', \mathbf{x_{k-2}} = 'z', \mathbf{x_{k-3}} = 'i', \mathbf{x_{k-4}} = 'p')?$$

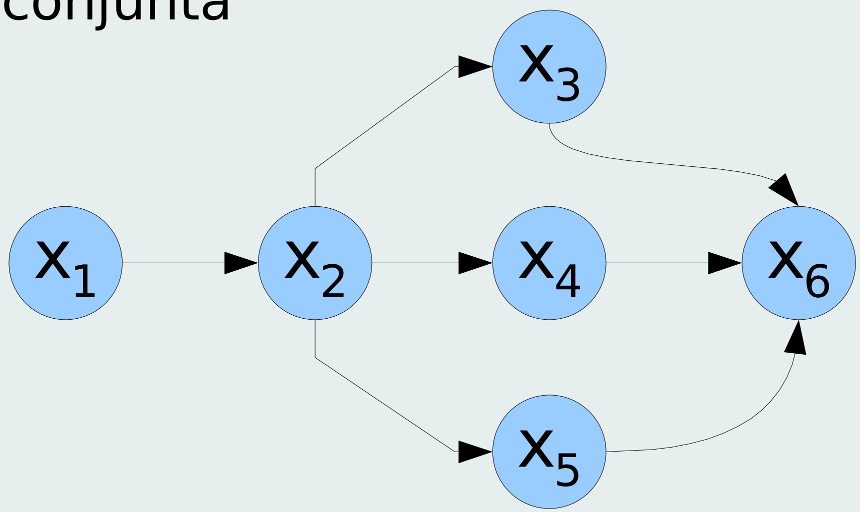
Caso as variáveis não sejam independentes teremos que manter os valores da distribuiçao conjunta de 5 variáveis.

Considerando 23 diferentes letras:

Precisaremos de um matriz com $\mathbf{23^5}$ células:
~ 6 Mega - 24 Mb (4 bytes por célula)

Redes Bayesianas

Utiliza informações sobre independência para diminuir custo em espaço para armazenar distribuição conjunta



$$p(x_1,x_2,x_3,x_4,x_5,x_6) = p(x_1).p(x_2|x_1).p(x_3|x_2).p(x_4|x_2).p(x_5|x_2).p(x_6|x_3,x_4,x_5)$$

$$10^6 \longrightarrow 10 + 10^2 + 10^2 + 10^2 + 10^2 + 10^4$$

~4 Mb ~40 Kb

- 4 bytes por célula
- 10 valores por variável

Teorema de Bayes

$p(x \mid y) = p(x,y) / p(y)$

$p(y \mid x) = p(y,x) / p(x)$

$p(x,y) = p(y,x)$ Conjunção é comutativa.

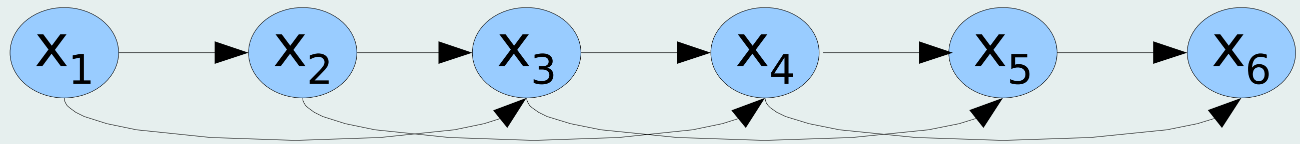
$\mathbf{p(x \mid y) = (p(y \mid x) * p(x)) / p(y)}$

p(d = 'meningite' | s = 'dor na nuca') = ?
p(d = 'torcicolo' | s = 'dor na nuca') = ?

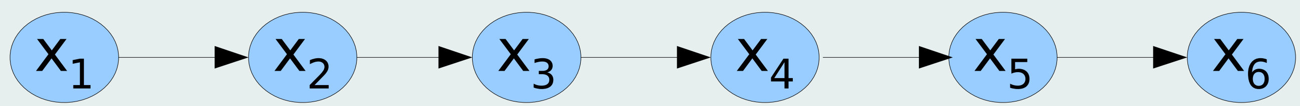
meningite ou torcicolo ?

$p(n|m)*p(m) > p(n|t)*p(t)$

Modelos de Markov de ordem n



$$p(x_k | x_{k-1}, x_{k-2}, \ldots, x_1) = p(x_k | x_{k-1}, x_{k-2})$$



$$p(x_k | x_{k-1}, x_{k-2}, \ldots, x_1) = p(x_k | x_{k-1})$$

Um tipo de modelo utilizado para representar **Processos Estocásticos.**

Modelos de Markov de Ordem n

$\mathbf{x_k}$ – ocorrência de uma letra na posição k

Qual a ordem do modelo de Markov que melhor representaria o problema de "adivinhar a próxima letra"

Compromisso:

Complexidade no Espaço X Margem de Erro

Outras Limitações

$\mathbf{x_k}$ – ocorrência de uma letra na posição i

O estado pode não ser diretamente observável

Atributos: Tot. Extremidades, Comprim. do Contorno, Cruzamentos.

[4 9 1]' , [1 2 0]' , [2 5 0]' , [3 7 1]' , [1 6 0]' , [0 3 0]' , [1 2 0]'

P I S T O R I

Reconhecimento de Língua de Sinais

FRUTAS EM LIBRAS

Fonte: Dicionário de Libras Ilustrado (Governo do Estado de São Paulo)

Reconhecimento de Língua de Sinais

Exemplos de Atributos

(1) Total pixels de pele visível, (2) Total pixels mão direita
(3) Total pixels mão esquerda, (4) Pontas de dedos visíveis,
(5) Distância entre centros de massa das mãos
(6) Distância entre centro de massa da mão esquerda e face

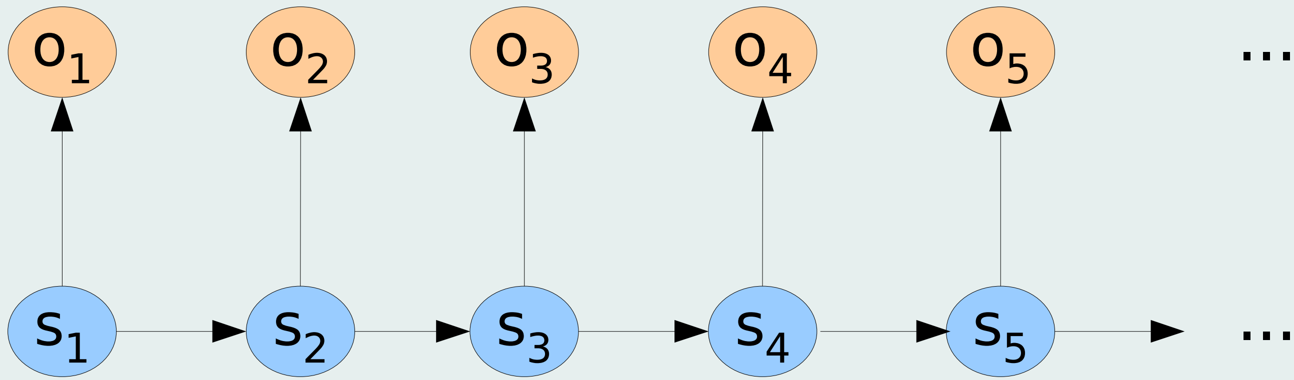
Observações (Seqüências de Vetores de Atributos):

[200 50 0 5 20 2]' , [200 70 0 5 20 4]' , [200 60 0 2 20 3]' , ...

Estados (exemplos)

1 - Mão direita na configuração “C”, próxima ao queixo, e mão direita fora da área de interesse.
2 – Duas mãos na configuração “V”, na altura do peito, com mão direita ao lado direito do tronco e mão esquerda...

Modelos de Markov Ocultos

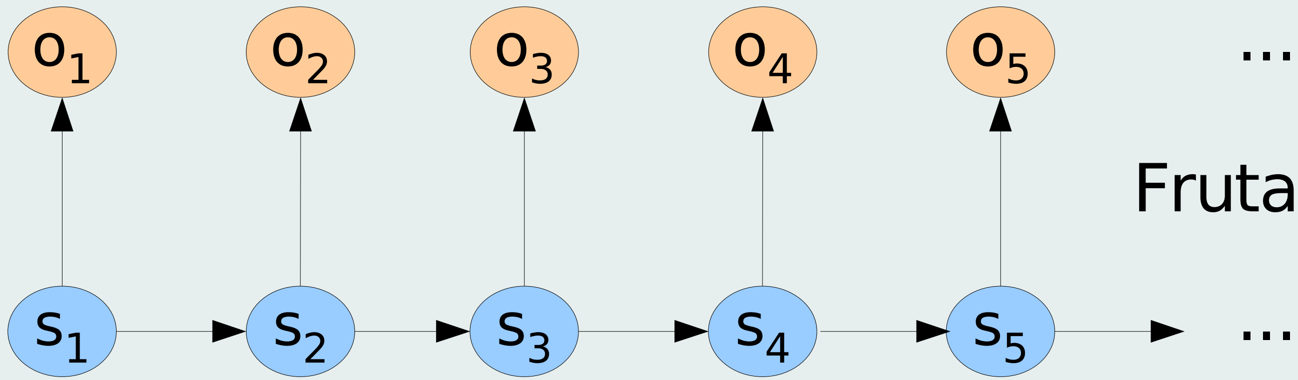


$p(s_k | s_{k-1}) =$ Probabilidade de termos duas mãos na configuração "V", na altura do peito, com mão direita ao lado direito do tronco e mão esquerda do lado esquerdo do tronco, **no quadro k**, sabendo que tínhamos a mão direita na configuração "C", próxima ao queixo, e mão direita fora da área de interesse, **no quadro k-1**.

$p(o_k | s_k) =$ Probabilidade de observarmos um vetor de atributos [200 50 0 5 20 2]' sabendo que a mão direita está na configuração "C", próxima ao queixo, e a mão direita está fora da área de interesse.

Modelos de Markov Ocultos

$\lambda =$



Frutas

o_1 = [200 50 0 5 20 2]', o_2 = [200 70 0 5 20 4]' , o_3 = [200 60 0 2 20 3]' , ...

$$p(\lambda | o_1, o_2, \ldots, o_T) = \frac{p(o_1, o_2, \ldots, o_T | \lambda) * p(\lambda)}{p(o_1, o_2, \ldots, o_T)}$$

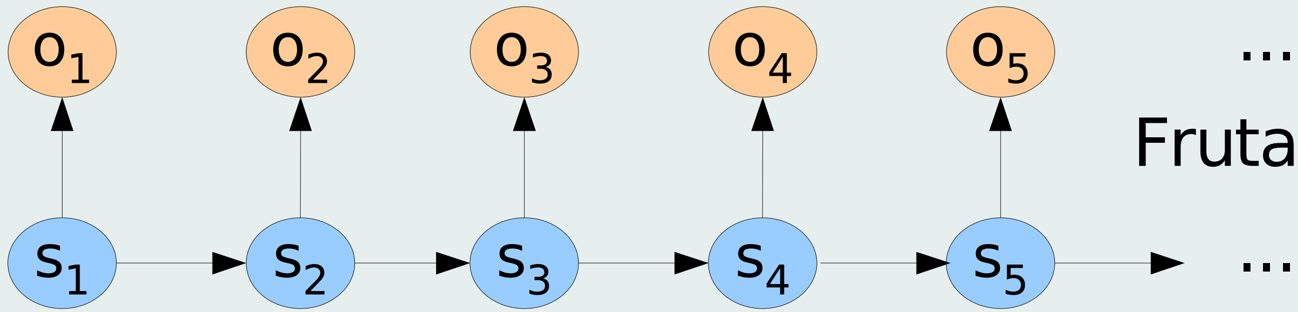
Fácil de Estimar

Pode ser Ignorado

$p(o_1, o_2, \ldots, o_T | \lambda)$ – Parte Interessante

Modelos de Markov Ocultos

$\lambda =$

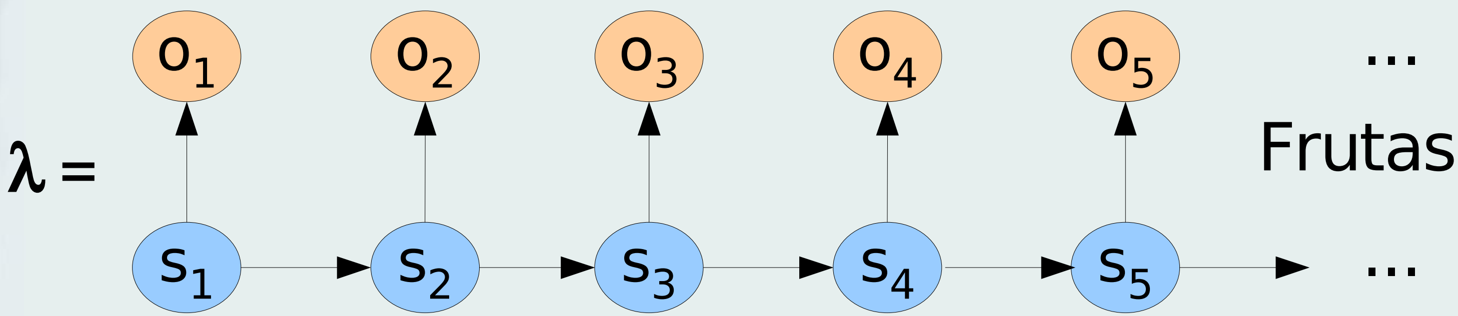


$$p(o_1,o_2,\ldots,o_T|\lambda) = \sum_{V\, s_1,\ldots,s_n} p(o_1,\ldots,o_T,s_1,\ldots,s_T)$$

$$p(o_1,\ldots,o_T,s_1,\ldots,s_T) =$$
$$p(s_1).p(o_1|s_1).p(o_2|s_2)\ldots p(o_T|s_T).p(s_2|s_1).p(s_3|s_2)\ldots p(s_T|s_{T-1})$$

$$p(o_1,\ldots,o_T,s_1,\ldots,s_T) = p(s_1).\prod_{k=1}^{T} p(o_k|s_k).p(s_{k+1}|s_k)$$

Modelos de Markov Ocultos



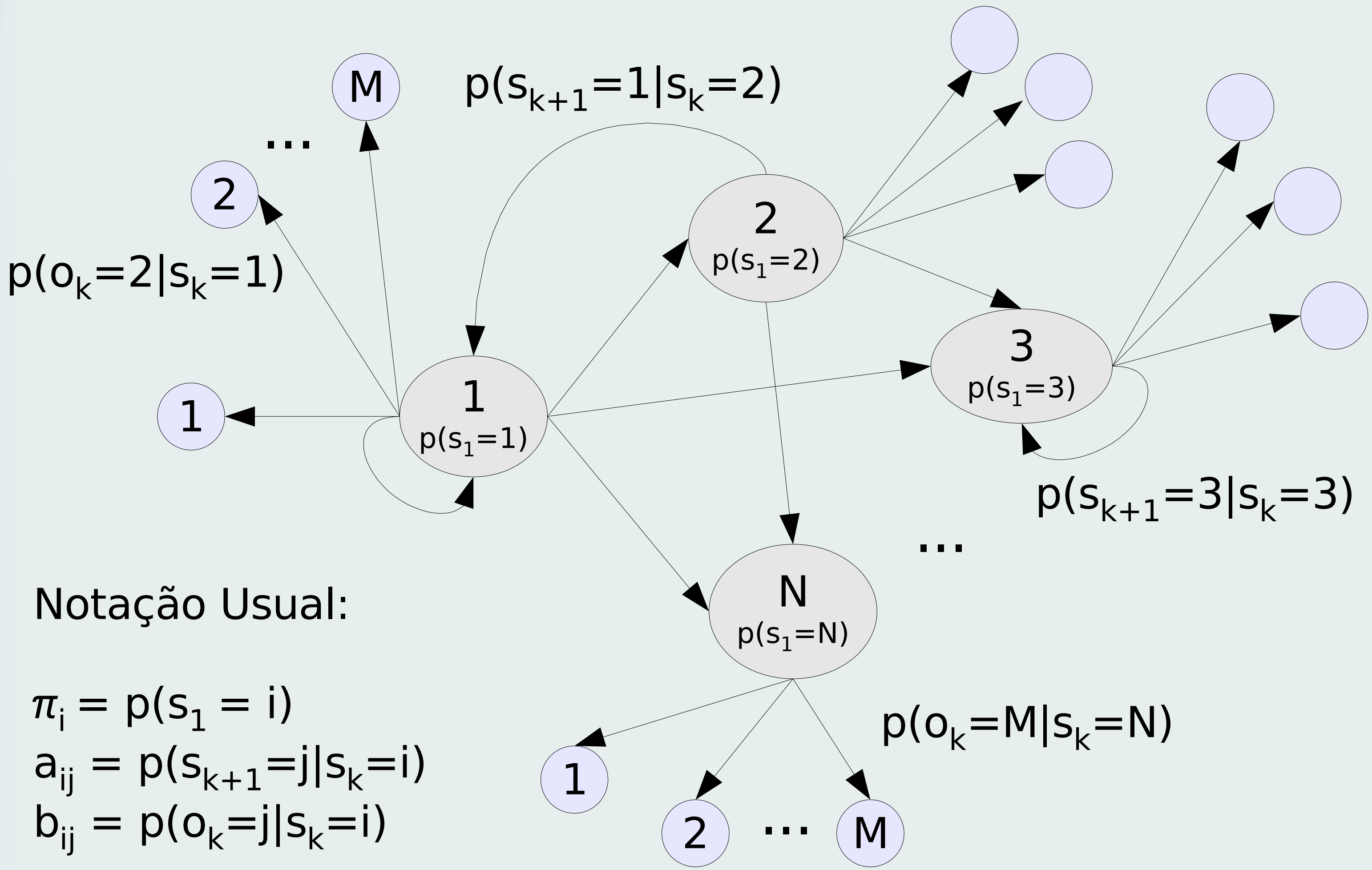
$$p(o_1,o_2,\ldots,o_T|\lambda) = \sum_{\forall\, s_1,\ldots,s_T} p(s_1).\prod_{k=1}^{T} p(o_k|s_k).p(s_{k+1}|s_k)$$

Homogeneidade no "Tempo" – Mesmas distribuições

$S = \{1,\ldots,N\} \quad O = \{1,\ldots,M\}$

$p(o_1=2|s_1=3) = p(o_3=2|s_3=3) \quad p(s_2=1|s_1=2) = p(s_5=1|s_4=2)$

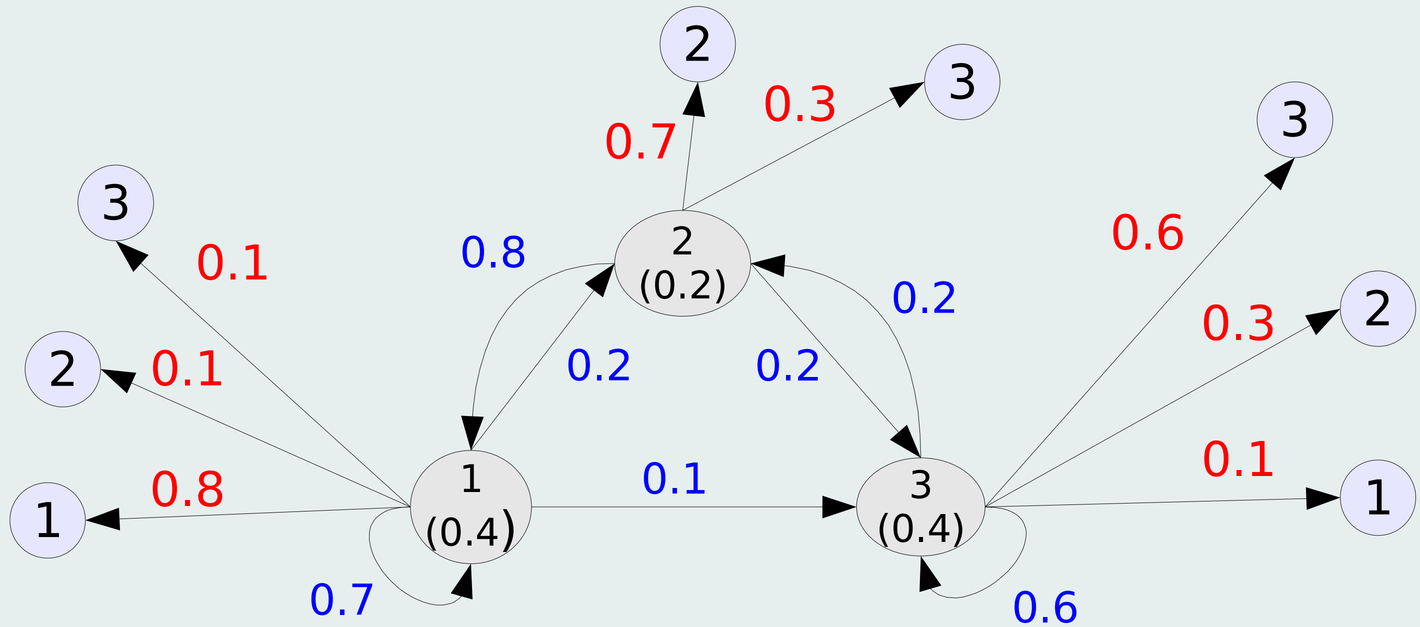
Modelos de Markov Ocultos



Notação Usual:

$\pi_i = p(s_1 = i)$
$a_{ij} = p(s_{k+1}=j|s_k=i)$
$b_{ij} = p(o_k=j|s_k=i)$

Modelos de Markov Ocultos

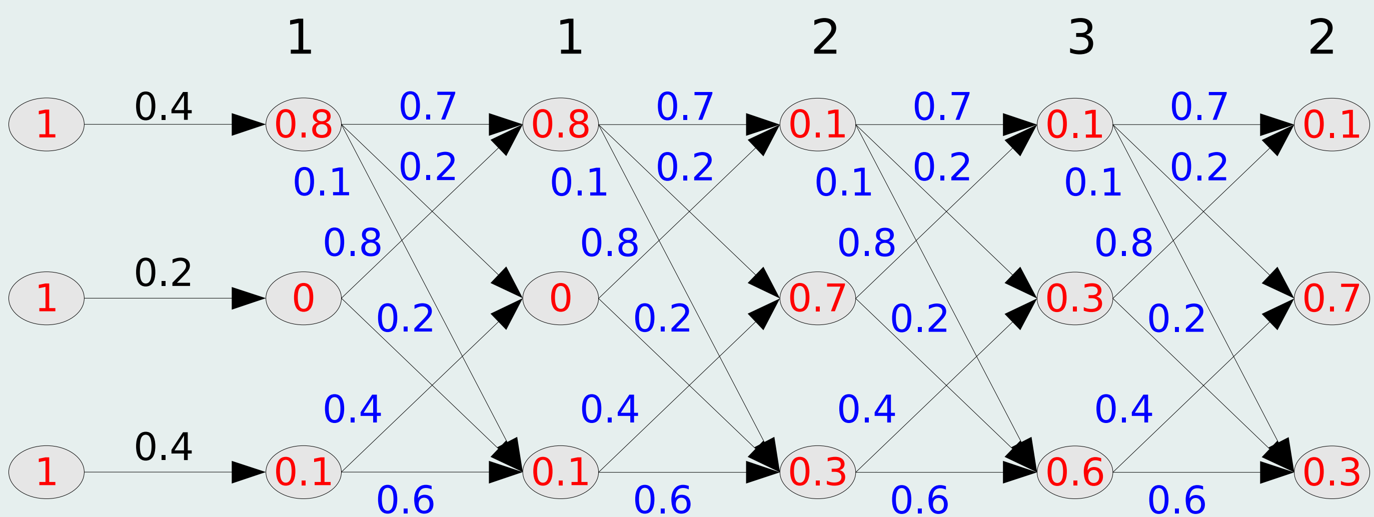


$$\pi = \begin{bmatrix} 0.4 \\ 0.2 \\ 0.4 \end{bmatrix} \quad A = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.8 & 0 & 0.2 \\ 0 & 0.4 & 0.6 \end{bmatrix} \quad B = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0 & 0.7 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}$$

Modelos de Markov Ocultos

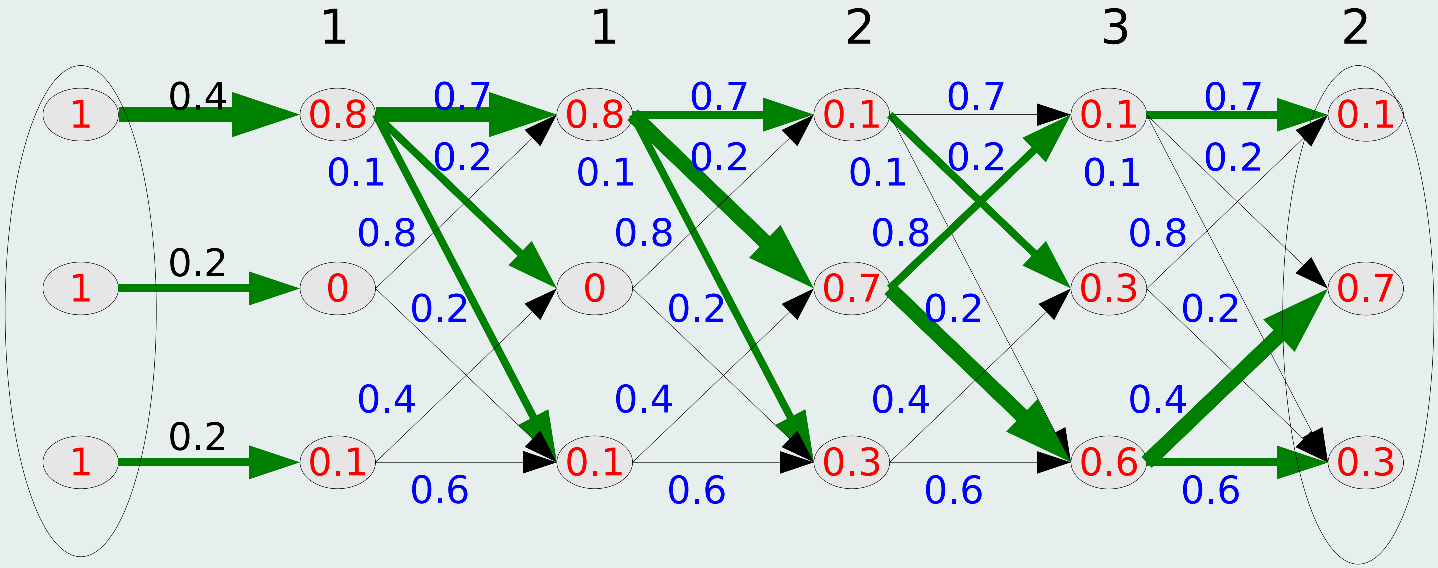
$$\pi = \begin{bmatrix} 0.4 \\ 0.2 \\ 0.4 \end{bmatrix} \quad A = \begin{array}{c} \scriptstyle k+1 \\ \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.8 & 0 & 0.2 \\ 0 & 0.4 & 0.6 \end{bmatrix} \end{array} {}^{k} \quad B = \begin{array}{c} \scriptstyle o \\ \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0 & 0.7 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{bmatrix} \end{array} {}^{s}$$

$$p(\lambda | o_1=1, o_2=1, o_3=2, o_4=3, o_5=2) \ ?$$



Treliça

Modelos de Markov Ocultos



V_I

$p(\lambda | o_1=1, o_2=1, o_3=2, o_4=3, o_5=2)$?

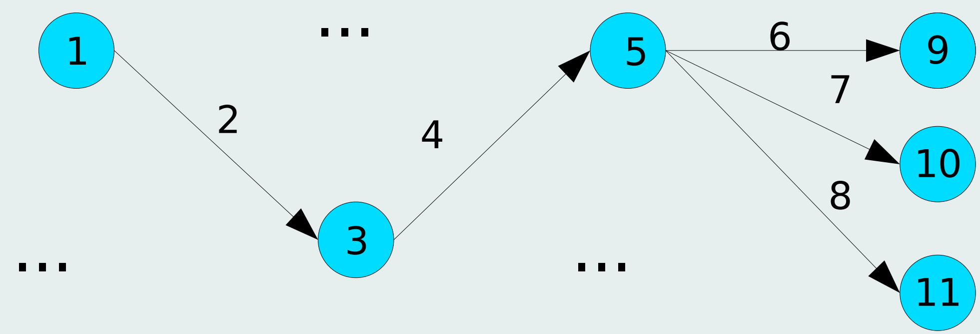
V_F

$$p(o_1, o_2, \ldots, o_T | \lambda) = \sum_{\forall s_1, \ldots, s_T} p(s_1) . \prod_{k=1}^{T} p(o_k | s_k) . p(s_{k+1} | s_k)$$

Problema: Somar o custo (multiplicação dos pesos de vértices e arestas) de cada caminho ligando um vértice de V_I a um de V_F

Algoritmo Forward/Backward

- Reduz complexidade de $O(TN^T)$ para $O(T^2N)$
- Baseado em Programação Dinâmica (Tópico de Otimização e Pesquisa Operacional)



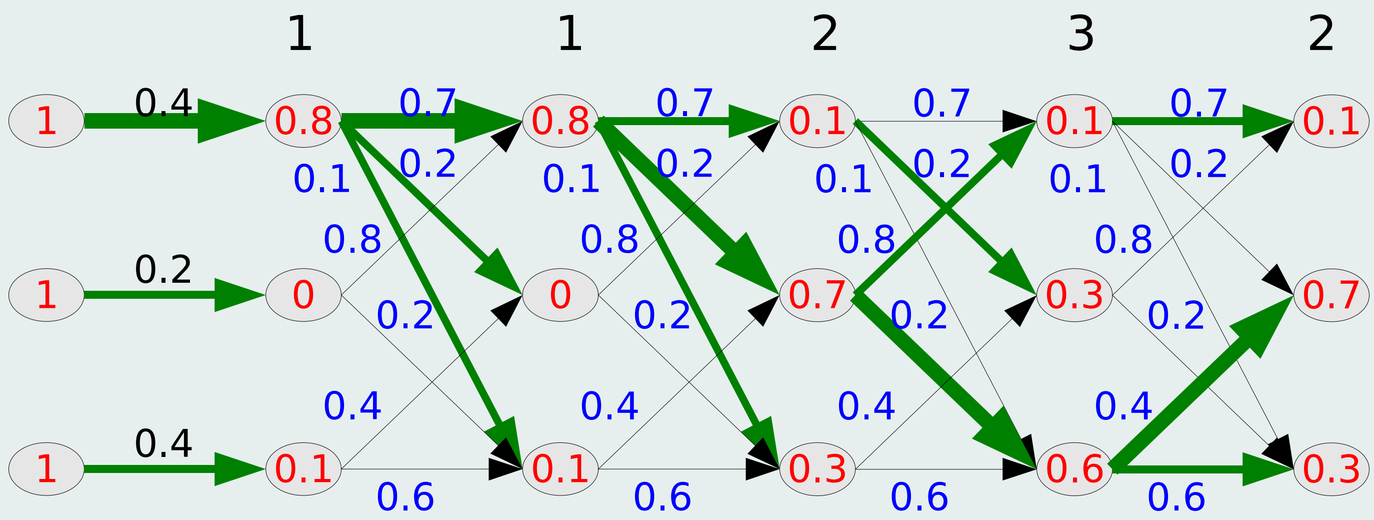
$1*2*3*4*5*6*9 + 1*2*3*4*5*7*10 + 1*2*3*4*5*8*11$

a_{ij} = Acumulador tempo i estado j

$1*2*3*4*5*(6*9 + 7*10 + 8*11)$

Algoritmo Viterbi

- Encontra seqüência mais provável de estados
- Modificação simples do Forward/Backward
- Armazenar caminhos de maior valor enquanto calcula valores a_{ij}



Algoritmo Baum-Welch

- Calcula os valores das matrizes p, A e B.
- Aprendizagem Supervisionada (amostras pré-classificadas)
- Amostras com informações “incompletas” (diferente do caso dos cursos e sexo).
- Algoritmo EM (expectation-maximization) aplicado aos modelos de Markov ocultos.

Observações (Seqüências de Vetores de Atributos):

[200 50 0 5 20 2]' , [200 70 0 5 20 4]' , [200 60 0 2 20 3]' , ...

Estados (manualmente identificados)

1 1 1 1 2 2 3 3 4 4

Considerações Finais

- Áreas envolvidas: probabilidade, estatística, processos estocásticos, teoria dos grafos, programação dinâmica, otimização combinatória, pesquisa operacional, complexidade de algoritmos, aprendizagem de máquina, redes bayesianas, inteligência artificial.
- Conceito análogo – Filtros de Kalman (distribuições contínuas)
- Conceitos relacionados – Modelos de Markov de Tempo Contínuo. Campos Aleatórios de Markov. Modelos Gráficos.