$$M = \begin{bmatrix} 0.2 & 0.4 \\ 0.8 & 0.6 \end{bmatrix} = \begin{bmatrix} p(x_n = 0 \mid x_{n-1} = 0) & p(x_n = 0 \mid x_{n-1} = 1) \\ p(x_n = 1 \mid x_{n-1} = 0) & p(x_n = 1 \mid x_{n-1} = 1) \end{bmatrix}$$

$$M = \begin{bmatrix} p(0|0) & p(0|1) \\ p(1|0) & p(1|1) \end{bmatrix}$$ (abreviando $p_{x_n | x_{n-1}}$)

$$\left(\hookrightarrow p_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right)$$

Considere que "sempre começamos" do estado $x_0 = \mathbf{0}$: $\begin{cases} p(x_0 = 0) = p(0) = 1 \\ p(x_0 = 1) = p(1) = 0 \end{cases}$

$$\left(P(A \cap B) = P(B|A) \cdot P(A) \quad \text{e} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B) \right)$$

$$p(x_1 = 0) = p(0|0)\,p(0) + p(0|1)\,p(1) = 0.2 \times 1 + 0.4 \times 0 = 0.2$$

$$p(x_1 = 1) = p(1|0)\,p(0) + p(1|1)\,p(1) = 0.8 \times 1 + 0.6 \times 0 = 0.8$$

$$\boxed{p_1 = M p_0}$$

UFRJ — COPPE — PEE — CPE723 — Otimização Natural

Simulated Annealing — Aula 04 — Análise de Convergência

Teorema A ("Stochastic Relaxation")

$$\pi(\omega) = \frac{1}{Z} e^{-\frac{J(\omega)}{T}}$$

$\{s_1, s_2, \ldots, s_N\}$

Assuma que, para todo s ("site" = componente do estado x) $\in S$ (conjunto de componentes do estado x), a sequência $\{n_t, t \geq 1\}$ contenha s infinitamente frequentemente. Então, para qualquer estado inicial $\eta \in \Omega$ (conjunto de ...

... todos os estados possíveis) e para qualquer $\omega \in \Omega$, temos:

$$\lim_{t \to \infty} P(x(t) = \omega \mid x(0) = \eta) = \pi(\omega) \qquad (*)$$

Obs.: η_t $(t = 1, 2, \ldots)$ é a sequência de componentes de x ("sites") que são sucessivamente perturbadas ("visitadas") no processo de atualização.

O conteúdo de cada componente de x ("site") tem L valores possíveis: $(s_i = x_i)$ $x_i \in \Lambda$; $\Lambda = \{0, 1, \ldots, L-1\}$ e $\Omega = \Lambda^N$.

Obs. 2: o resultado $(*)$ é análogo ao do artigo de Metropolis, 1953.

Vamos estudar o Teorema A através de um exemplo:

Exemplo: $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$; $x_i \in \underbrace{\{0, 1, 2, 3\}}_{\Lambda}$; $\tilde{\Lambda} = \{0, 1, 2, 3\}^4$

Função-custo: $\frac{1}{4}$ da distância quadrática ao vetor $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

$$J(x) = \frac{1}{4} \left\| x - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\|^2$$

Perturbações (para o "loop interno, de Metropolis"): $x_i := x_i + 1$, $+2$ ou $+3$ (cada uma com prob. $\frac{1}{3}$)

$x_i := x_i + 0, +1, +2$ ou $+3$ (cada uma com prob. $\frac{1}{4}$)

As perturbações do tipo +1, +2 ou +3 são implementadas por meio de

$$\lceil U(0,1) * 3 \rceil$$

Ou seja, sorteamos i de forma tal que $P(i=1) = P(i=2) = P(i=3) = P(i=4) = \frac{1}{4}$ (perturbações equiprováveis) e, em seguida, fazemos:

$$x_i = np.remainder(\, x_i + \lceil U(0,1) * 3 \rceil, 4)$$

Se $x = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$ e $i = 2$:

$$\begin{pmatrix} 1 \\ 3 \\ 0 \\ 1 \end{pmatrix} \xleftarrow{+1} \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix} \xrightarrow{+2} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix} \xrightarrow{+3} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

Valores que podem ser assumidos pela função-custo $J(x)$:

0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1	$\frac{5}{4}$	$\frac{3}{2}$	$\frac{7}{4}$	2	$\frac{9}{4}$	$\frac{5}{2}$	x	3	$\frac{13}{4}$	x	x	4	$J(x)$
1	8	24	32	20	24	48	32	6	24	24		4	8			1	N_J
0.39	3.13	9.38	12.5	7.81	9.38	18.75	12.5	2.34	9.38	9.38		1.56	3.13			0.39	N_J (%)
3.35	16.25	29.22	23.38	8.59	6.38	7.82	3.20	0.32	0.85	0.53		0.049	0.04			0.002	$P(J)$ (Metrop.)
3.29	1.99	1.21	0.73	0.45	0.27	0.16	0.10	0.06	0.04	0.02		0.008	0.005			0.001	$\frac{e^{-J/T}}{Z}$ ($T = 0.5$)

Obs.: $Z = \sum_{n=1}^{256} \exp\left(\frac{-J(x(n))}{T}\right)$

probabilidade individual de um estado x, dado o seu J

Note que $P(J) \simeq N_J \times \frac{\exp(-J/T)}{Z}$

probabilidade de J

número de "ocorrências" de J em $J(x(n))$, $n = 1, 2, \ldots, 256$

"estados $x(n) \mid J(x(n)) = J$"

Matriz de Transição M_1: (referente a perturbações na primeira componente de x)

← ... 256 origens ... →

	(1) 0000	(2) 0001	(3) 0002	(4) 0003	(5) 0010	...	(i) abcd	...	(256) 3333
(1) 0000		0							
(2) 0001		$\frac{1}{3}\left(1-\exp\left(\frac{-0.75}{0.5}\right)\right)$							
⋮		⋮							
(66) 1001		$\frac{1}{3}$							
⋮		⋮							
(130) 2001		$\frac{1}{3}$							
⋮		⋮							
(194) 3001		$\frac{1}{3}\exp\left(\frac{-0.75}{0.5}\right)$							
⋮									
(256) 3333									

256 destinos

$(i = 64a + 16b + 4c + d + 1)$

$\Delta J = 0.75$

$\frac{1}{3}$ (= prob("aceitar 1001 a partir de 0001") × prob("sortear $\hat{x} = 1001$ a partir de 0001"))

prob("aceitar 3001 a partir de 0001")

prob("sortear $\hat{x} = 3001$ a partir de 0001")

x	$J(x)$
0001	0.75
1001	0.50
2001	0.75
3001	1.50

Da mesma forma que calculamos M_1, podemos calcular também M_2, M_3 e M_4.

As quatro matrizes compartilham o mesmo vetor invariante. Ele pode ser calculado através de $\pi(\omega) = \frac{e^{-\frac{J(\omega)}{T}}}{Z}$ e verificado através de $M\pi = \pi$ ($M = M_1$, M_2, M_3 ou M_4) ou, para matrizes pequenas, através da solução da equação do vetor invariante: $M\pi = \pi \longrightarrow (M - I)\pi = 0 \longrightarrow \pi$.

Começando com $x(0)$ sorteado de uma distribuição de probabilidades qualquer (p_0), temos $\underbrace{M_4 M_2 M_1 M_3 M_3 M_1 M_2 \cdots M_4}_{\text{aplicação das matrizes } M \text{ em ordem aleatória.}} p_0 \cong \pi$.

Lema 1: Temperatura T fixa (constante)

$$
t \longleftarrow \quad
\begin{array}{c|cccccc|ccccc|c}
M_4 & M_1 & M_2 & M_2 & M_3 & M_3 & M_4 & M_3 & M_1 & M_2 & M_4 & M_1 & P_0 \\
12 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0
\end{array}
$$

T_2 $\qquad$ T_1 $\qquad$ T_0

$k=2$ $\qquad$ $k=1$ $\qquad$ $k=0$ $\qquad$ $(k(t) = \sup\{k : T_k < t\})$

Existe uma constante r, $0 \leq r < 1$, tal que para $t = 1, 2, \ldots$:

$$
\sup_{\omega, \eta', \eta''} \left| P(X(t) = \omega \mid X(0) = \eta') - P(X(t) = \omega \mid X(0) = \eta'') \right| \leq r^{k(t)}
$$

Obs. 1: Lema Δ + ($M\pi = \pi$) = Teorema A

Obs. 2: $r = 1 - L^N \delta^N$

No nosso exemplo de hoje $\left(x_i \in \underbrace{\{0,1,2,3\}}_{L=4},\ i = \underbrace{1,2,3,4}_{N=4}\right)$, temos:

δ : menor número dentre as probabilidades locais (em M_i , $i = 1,2,3,4$)

$M_1(193,65)$: $\underset{J=0.75}{1000} \longrightarrow \underset{J=1.75}{3000}$ $\quad \underset{T=0.5}{\Delta J = 1}$ $\quad \frac{1}{3} e^{-2} = 0.0451$

$\delta = 0.0451$ (também nas matrizes M_2, M_3 e M_4)

$\delta^4 = 4.1415 \times 10^{-6}$ $\qquad L^4 = 256$ $\qquad \delta^4 L^4 = 1.06 \times 10^{-3}$

$r = 1 - L^N \delta^N = 0.9984$

$r^{10} = 0.989$ ("10 visitas completas") ; $r^{100} = 0.899$ etc.

Teorema B ("Annealing")

$\lim_{t \to \infty} T(t) = 0$ e $T(t) \geq \frac{N\Delta}{\log t}$, $t \geq 2$ ($T(t)$ é uma sequência de temperaturas)

$\Delta = J_{max} - J_{min}$ $\left(\pi_{T(t)}(\omega)\right)$

π_0 : $\lim_{T \to 0}$ "$\pi(\omega)$" — limite da sequência de vetores invariantes das matrizes de transição, à medida em que $T \to 0$.

Para qualquer estado inicial $\eta \in \Omega$ e para qualquer $\omega \in \Omega$, temos:

$$\lim_{t \to \infty} P\left(x(t) = \omega \,\middle|\, x(0) = \eta\right) = \pi_0(\omega) \quad \left(= \begin{cases} 0, \text{ se } \omega \notin \Omega_0 \\ \frac{1}{|\Omega_0|}, \text{ se } \omega \in \Omega_0 \end{cases} \right)$$