Correções em slides dos arquivos Aula5.pdf e Aula6.pdf:

Aula5.pdf, slide 3: $\ln(1-x) \simeq -x$, se $x \simeq 0$

Aula5.pdf, slide 5: (terceira linha): $g_k \propto \dfrac{T_0 k}{T_0^2 + k^2 x^2} \underset{k \to \infty}{\simeq} \dfrac{T_0}{k x^2}$

Aula5.pdf, slide 5: leitura dos artigos de Szu & Hartley

Seção 3 (Junho 1987)

seção II (Novembro 1987)

Aula6.pdf, slide 3: $y_2 = \dfrac{x(1) + x(2)}{N_2}$

$x(1) \in$ cluster 2

$x(2) \in$ cluster 2

nº de vetores $x(n) \in$ cluster 2

UFRJ — COPPE — PEE — CPE723 — Otimização Natural

Simulated Annealing — Aula 07 — Deterministic Annealing

$$J = \sum_x p_x \sum_y p_{y|x} d_{xy} + T \sum_x p_x \sum_y p_{y|x} \log p_{y|x} + K - \underbrace{\sum_x \gamma_x \sum_y p_{y|x}}_{\sum_x \gamma_x \sum_y p_{y|x} = K}$$

$$\left(\frac{d(x \log x)}{dx} = 1 + \log x \right)$$

(otimizando...)

$$\frac{dJ}{dp_{y|x}} = p_x d_{xy} + T p_x (1 + \log p_{y|x}) - \gamma_x \overset{\vee}{=} 0$$

$$p_x \Big(d_{xy} + T \log p_{y|x} + \underbrace{T - \frac{\gamma_x}{p_x}}_{\text{"} T \log z_x \text{"}} \Big) = 0 \implies d_{xy} + T \log p_{y|x} + T \log z_x = 0$$

Então: $\log p_{y|x} = -\frac{d_{xy}}{T} - \log z_x$

$$\boxed{p_{y|x} = \frac{e^{-\frac{d_{xy}}{T}}}{z_x}}$$, Condição da Partição (DA) (Etapa 1),

onde $z_x = \sum_y e^{-\frac{d_{xy}}{T}}$ $\longleftarrow$ fator de normalização, para que $\sum_y p_{y|x} = 0 \quad \forall x$.

Exemplo: $X = \begin{bmatrix} 0.4 & -2.1 & 0.5 & -0.5 \\ 1.7 & 0.7 & -1.5 & 0.4 \end{bmatrix}$, $Y = \begin{bmatrix} 3.8 & -1.6 \\ 2.5 & 2.0 \end{bmatrix}$ e $T = 10$

$$e^{-\frac{d_{xy}}{10}} = \begin{bmatrix} 0.30 & 0.02 & 0.07 & 0.10 \\ 0.66 & 0.82 & 0.19 & 0.69 \end{bmatrix} \Rightarrow p_{y|x} = \begin{array}{c|c} \begin{bmatrix} 0.31 & 0.03 & 0.26 & 0.13 \\ 0.69 & 0.97 & 0.74 & 0.87 \end{bmatrix} & \begin{array}{c} (y_1) \\ (y_2) \end{array} \\ \begin{array}{cccc} (x(1) & x(2) & x(3) & x(4)) \end{array} & \end{array}$$

(z_1) $\Sigma = 0.69$ $\Sigma = 0.84$ $\Sigma = 0.26$ $\Sigma = 0.79$ (z_4)

Substituindo $p_{y|x} = \left(\frac{1}{z_x}\right) e^{\frac{-d_{xy}}{T}}$ na expressão $J = D - TH$, encontramos $J_{min}\Big|_T$:

$$J = \sum_x p_x \sum_y p_{y|x} d_{xy} + T \sum_x p_x \sum_y p_{y|x} \log p_{y|x} \underbrace{\left(+ \kappa - \sum_x \gamma_x \sum_y p_{y|x} \right)}_{0}$$

$$J_{min}\Big|_T = \cancel{\sum_x p_x \sum_y \frac{e^{\frac{-d_{xy}}{T}}}{z_x} d_{xy}} + T \sum_x p_x \sum_y \frac{e^{\frac{-d_{xy}}{T}}}{z_x} \left(\cancel{-\frac{d_{xy}}{T}} - \log z_x \right)$$

$$J_{min}\Big|_T = -T \sum_x p_x \log z_x$$

Exemplo: $J_{min}\Big|_{T=10} = -10\left(\frac{\log(0.69)}{4} + \frac{\log(0.84)}{4} + \frac{\log(0.26)}{4} + \frac{\log(0.79)}{4} \right) = ?$

$= 5.32$ ↲

Etapa 2 — Condição do Centróide (DA)

$$D = \sum_x p_x \sum_y p_{y|x} d_{xy} = \swarrow \qquad , \left((x(n)-y_k)^T (x(n)-y_k) = x^T(n)x(n) - 2x^T(n)y_k + y_k^T y_k \right)$$

$$= \frac{1}{4}\Big(p_{1|1} \|x(1)-y_1\|^2 + p_{1|2} \|x(2)-y_1\|^2 + p_{1|3} \|x(3)-y_1\|^2 + p_{1|4} \|x(4)-y_1\|^2 +$$

$$+ p_{2|1} \|x(1)-y_2\|^2 + p_{2|2} \|x(2)-y_2\|^2 + p_{2|3} \|x(3)-y_2\|^2 + p_{2|4} \|x(4)-y_2\|^2 \Big)$$

$$\frac{dD}{dy_k} = \frac{1}{4}\Big(p_{k|1}(-2x(1)+2y_k) + p_{k|2}(-2x(2)+2y_k) + p_{k|3}(-2x(3)+2y_k) + p_{k|4}(-2x(4)+2y_k) \Big) = 0$$

(otimizando...)

Então: $\left(\sum_x p_{k|x}\right) y_k = \sum_x p_{k|x} x$

$$\boxed{y_k = \frac{\sum_x x p_{k|x}}{\sum_x p_{k|x}}}$$, Condição do Centróide (Etapa 2)

Comparações entre SA, GLA e DA (problema: "clustering")

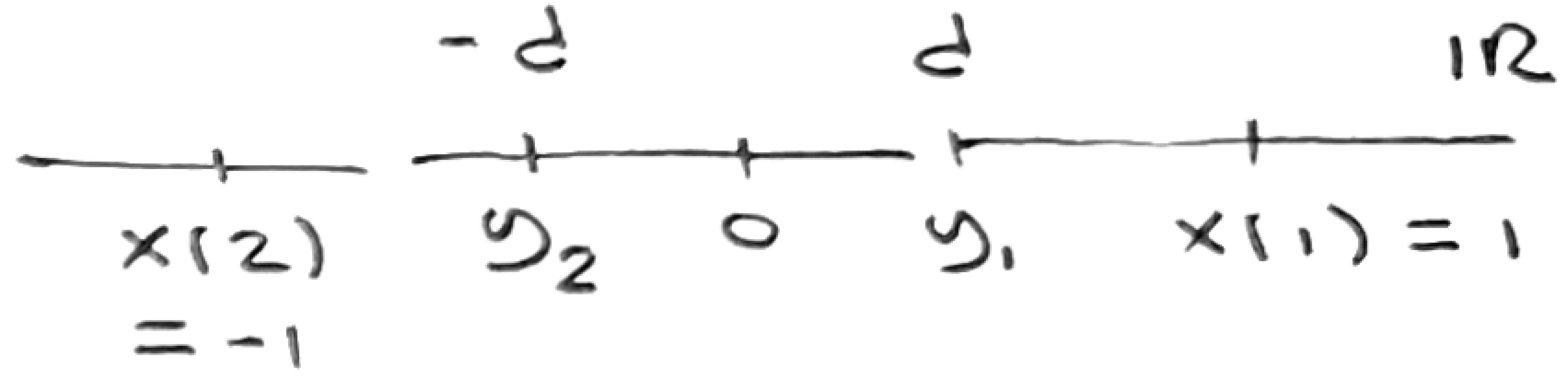
SA: evita mínimos locais, estocástico (lento), parâmetros a ajustar (N, ε, κ e T_0), analogia com resfriamento lento de materiais.

GLA: não evita mínimos locais, determinístico (rápido), simples (sem parâmetros a ajustar. Analogia: métodos gradientes $\longrightarrow$ "descida".

DA: evita mínimos locais, determinístico (rápido), parâmetros T e α (para redução de T). Analogia: mudanças de fase em torno de temperaturas críticas.

Mudanças de fase da solução conforme a temperatura

Prova de 2008*, Questão 4 (use $d(x,y) = \| x - y \|^2$)



$$d(x,y) \longrightarrow d_{xy} = \begin{bmatrix} d^2 - 2d + 1 & d^2 + 2d + 1 \\ d^2 + 2d + 1 & d^2 - 2d + 1 \end{bmatrix}$$

Partição:

$$\left(\alpha = e^{\frac{2d}{T}} \right)$$

$$P_{Y|X} = \begin{bmatrix} \dfrac{e^{\frac{-d^2+2d-1}{T}}}{e^{\frac{-d^2+2d-1}{T}} + e^{\frac{-d^2-2d-1}{T}}} & \dfrac{e^{\frac{-d^2-2d-1}{T}}}{e^{\frac{-d^2+2d-1}{T}} + e^{\frac{-d^2-2d-1}{T}}} \\[2em] \dfrac{e^{\frac{-d^2-2d-1}{T}}}{e^{\frac{-d^2+2d-1}{T}} + e^{\frac{-d^2-2d-1}{T}}} & \dfrac{e^{\frac{-d^2+2d-1}{T}}}{e^{\frac{-d^2+2d-1}{T}} + e^{\frac{-d^2-2d-1}{T}}} \end{bmatrix}$$

(* também de 2009 e 2011)

Substituindo $e^{\frac{2d}{T}}$ por α em $P_{Y|X}$, temos:

$$P_{Y|X} = \begin{bmatrix} \dfrac{\alpha}{\alpha+\alpha^{-1}} & \dfrac{\alpha^{-1}}{\alpha+\alpha^{-1}} \\[2ex] \dfrac{\alpha^{-1}}{\alpha+\alpha^{-1}} & \dfrac{\alpha}{\alpha+\alpha^{-1}} \end{bmatrix}$$

Condição da partição: novo y_1 (note que $y_2 = -y_1$)

$$y_1 = \left(\frac{1}{1}\right) \times \left(\frac{\alpha}{\alpha+\alpha^{-1}}\, x(1) + \frac{\alpha^{-1}}{\alpha+\alpha^{-1}}\, x(2)\right) = \frac{\alpha-\alpha^{-1}}{\alpha+\alpha^{-1}} = \frac{e^{\frac{2d}{T}} - e^{-\frac{2d}{T}}}{e^{\frac{2d}{T}} + e^{-\frac{2d}{T}}} = \tanh\left(\frac{2d}{T}\right)$$

Atualizando $d(n)$, temos:

$$d(n+1) = \tanh\left(\frac{2d(n)}{T}\right).$$ Numericamente: $$\lim_{n\to\infty} d(n) \begin{cases} = 1, & \text{se } T < 0.5 \\ = 0, & \text{se } T > 2 \\ \in [0,1], & \text{se } T \in [0.5, 2] \end{cases}$$