Lista de Exercícios 3

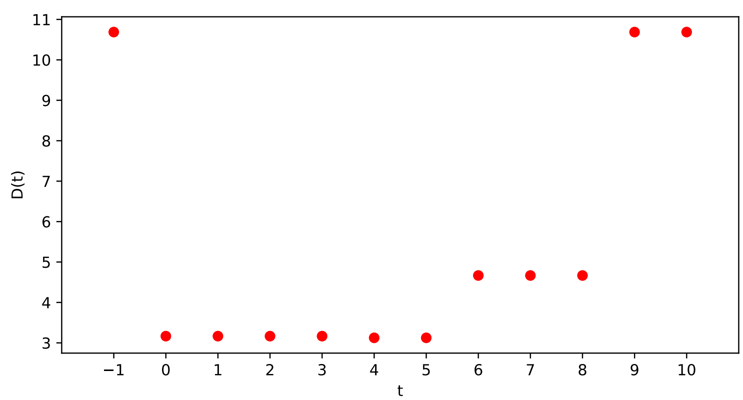
Otimização Natural

Patrícia Kovaleski

2 de abril de 2018

Questão 1) *D.A. para Quantização Escalar com 1 Bit*

(a) Gráfico de $D(t)$ para $t \in [-1, 10]$.



(b) Matriz não-normalizada p', calculada utilizando $\epsilon^{-d_{xy}/T}$:

$$p' = \begin{bmatrix} 1.2341 \times 10^{-4} & 3.6788 \times 10^{-1} & 1.2341 \times 10^{-4} & 2.3195 \times 10^{-16} \\ 9.5402 \times 10^{-6} & 6.9768 \times 10^{-1} & 1.1592 \times 10^{-3} & 2.4017 \times 10^{-14} \end{bmatrix}$$

Vetor μ, calculado pela soma das colunas da matriz não-normalizada:

$$\mu = \begin{bmatrix} 1.3295 \times 10^{-4} & 1.0655 & 1.2826 \times 10^{-3} & 2.4249 \times 10^{-14} \end{bmatrix}$$

Matriz $p(y|x)$, calculada pela divisão da matriz p' pelo vetor μ :

$$p(y|x) = \begin{bmatrix} 0.9282 & 0.3452 & 0.0962 & 0.0096 \\ 0.0717 & 0.6547 & 0.9038 & 0.9904 \end{bmatrix}$$

(c)

$$D = \sum_{\mathbf{X}} p(x) \sum_{\mathbf{Y}} p(y|x) d(x, y)$$

$$D = 0.25 \times (d_{x_1y_1} \times p(y_1|x_1) + d_{x_1y_2} \times p(y_2|x_1) + d_{x_2y_1} \times p(y_1|x_2) + d_{x_2y_2} \times p(y_2|x_2) + \\ d_{x_3y_1} \times p(y_1|x_3) + d_{x_3y_2} \times p(y_2|x_3) + d_{x_4y_1} \times p(y_1|x_4) + d_{x_4y_2} \times p(y_2|x_4))$$

$$D = 0.25 \times (9.0 \times 0.9282 + 11.56 \times 0.0717 + 1.0 \times 0.3452 + 0.36 \times 0.6547 + \\ 9.0 \times 0.0962 + 6.76 \times 0.9038 + 36.0 \times 0.0096 + 31.36 \times 0.9904)$$

$$D = 12.036$$

(d) Os novos centróides são calculados de acordo com:

$$y_i = \frac{\sum_x p(y_i|x)x}{\sum_x p(y_i|x)}$$

Logo,

$$y_1 = 1.4822$$
$$y_2 = 6.4698$$

(e) Repetindo as letras (b), (c) e (d) para T = 0.1

$$p(y|x) = \begin{bmatrix} 1.000 & 1.659 \times 10^{-3} & 1.870 \times 10^{-10} & 7.059 \times 10^{-21} \\ 7.622 \times 10^{-12} & 0.998 & 1.000 & 1.000 \end{bmatrix}$$

$$D = 11.8703$$

$$y_1 = 0.0066$$
$$y_2 = 6.3346$$

(f) Repetindo as letras (b), (c) e (d) para T = 50.0

$$p(y|x) = \begin{bmatrix} 0.5128 & 0.4968 & 0.4888 & 0.4768 \\ 0.4872 & 0.5032 & 0.5112 & 0.5232 \end{bmatrix}$$

$$D = 13.0881$$

$$y_1 = 4.6635$$
$$y_2 = 4.8344$$

(g) Comparando os items (d), (e) e (f), vemos que quanto menor a temperatura, menor o erro encontrado em D. Porém, utilizar de primeira um valor muito baixo de temperatura não costuma nos levar ao mínimo global da função, uma vez que não permite aumentar o valor de D, apenas diminuí-lo. Para T = 0.1, foi encontrado o menor valor para D, com probabilidades condicionais próximas a 1 ou 0 para cada cluster. Isto evidencia um *Hard Clustering* e muito provavelmente o resultado encontrado trata-se de um mínimo local.

Já utilizando um valor muito alto de temperatura, como T = 50, a distribuição de probabilidade tornou-se uniforme entre os clusteres. Com isso, há máxima incerteza sobre a distribuição dos clusteres. Este pode ser visto como uma das etapas do *Soft Clustering*, que, após maximizar a incerteza e encontrar o mínimo nessas condições, diminui o valor da temperatura e recalcula as partições e centróides para a nova condição. Nessa abordagem não necessariamente será encontrado o mínimo global da função, mas permitirá escapar de alguns mínimos locais.

Questão 2)