

# Algoritmo $k$ -means Baseado na Distância de Fisher-Rao na Subvariedade das Distribuições Gaussianas Diagonais

Julianna Pinele<sup>1</sup>

Instituto de Matemática, Estatística e Computação Científica, UNICAMP, Campinas, SP

João Eloir Strapasson<sup>2</sup>

Faculdade de Ciências Aplicadas, UNICAMP, Limeira, SP

**Resumo.** Uma fórmula explícita da distância de Fisher-Rao para modelos de distribuição em geral é muito difícil de determinar, especialmente no caso das distribuições Gaussianas multivariadas. Neste trabalho apresentamos um limitante superior para a distância de Fisher-Rao no espaço das distribuições Gaussianas multivariadas usando a distância na subvariedade formada por distribuições gaussianas cuja matriz de covariância é diagonal. Introduzimos também uma adaptação do algoritmo  $k$ -means baseado na distância de Fisher-Rao para simplificar misturas compostas por distribuições Gaussianas diagonais. Aplicamos o algoritmo para simplificar misturas Gaussianas no contexto de segmentação de imagens.

**Palavras-chave.** Distância de Fisher-Rao, Limitante Superior, Mistura Gaussiana, Algoritmo  $k$ -means, Segmentação de Imagens.

## 1 Introdução

A abordagem de geometria diferencial em teoria de probabilidade foi apresentada no trabalho pioneiro de C. Rao em 1945. Ele introduziu uma estrutura Riemanniana no espaço formado por distribuições de probabilidade parametrizadas utilizando a métrica dada por R. Fisher em 1921. A distância associada à essa métrica será chamada aqui de distância de Fisher-Rao. Esta, assim como outras medidas de dissimilaridade entre duas distribuições de probabilidade, tem sido aplicada em várias áreas tais como inferência estatística [5], processamento de imagens, processamento de sinais [3] entre outras.

Uma fórmula explícita da distância de Fisher-Rao para modelos de distribuição em geral são muito difíceis de determinar, especialmente no caso das distribuições Gaussianas multivariadas. Nesse espaço uma fórmula fechada para a distância de Fisher-Rao é conhecida apenas no caso univariado e em algumas subvariedades. A distância de Fisher-Rao na subvariedade composta por distribuições Gaussianas, cuja matriz de covariância é diagonal, fornece um limitante superior para a distância de Fisher-Rao no caso geral.

Muitas aplicações contendo distância ou divergência entre distribuições envolvem análise de agrupamento, também conhecido como *clustering*. Clustering tem como objetivo

---

<sup>1</sup>jpinele@ime.unicamp.br

<sup>2</sup>joao.strapasson@fca.unicamp.br

colocar em um mesmo grupo objetos que sejam semelhantes de acordo com alguma função de similaridade. Um dos algoritmos bastante utilizado na área de agrupamento é o *k-means* proposto por S. Lloyd em 1957.

Modelos de mistura Gaussiana são frequentemente utilizados para modelar pontos de um conjunto de dados. O algoritmo mais conhecido para estimar parâmetros de uma mistura Gaussiana é o EM (*Expectation Maximization*). Em muitas aplicações que envolvem modelos de mistura, o custo computacional é muito alto devido ao grande número de componentes da mistura. Esse custo pode ser fortemente diminuído se reduzirmos o número de componentes da mistura. Garcia e Nielsen propuseram em [3] uma simplificação para a mistura Gaussiana usando o *Bregman hard clustering*, uma adaptação do *k-means* para misturas de famílias exponenciais usando a divergência de Bregman.

Neste trabalho, apresentamos uma adaptação do algoritmo *k-means* baseado na distância de Fisher-Rao para agrupar distribuições Gaussianas diagonais e aplicamos o algoritmo para simplificar misturas Gaussianas diagonais no contexto de segmentação de imagens.

## 2 Distância de Fisher-Rao

Em 1945, Rao propôs um método para calcular a distância entre distribuições de probabilidade introduzindo uma métrica Riemanniana, em termos da chamada matriz de informação de Fisher, em uma família paramétrica de distribuições de probabilidade.

**Definição 2.1.** (*Matriz de informação de Fisher*) Seja  $S = \{p_{\theta}; \theta \in \Theta\}$  um modelo estatístico de dimensão  $n$  com espaço de parâmetros  $\Theta$ . Dado um ponto  $\theta \in \Theta$ , a matriz de informação de Fisher de  $S$  em  $\theta$  é a matriz  $G = [g_{ij}(\theta)]$  de ordem  $n$ , tal que

$$g_{ij}(\theta) = E_{\theta} \left( \frac{\partial}{\partial \theta_i} \log p(x; \theta) \frac{\partial}{\partial \theta_j} \log p(x; \theta) \right) = \int \frac{\partial}{\partial \theta_i} \log p(x; \theta) \frac{\partial}{\partial \theta_j} \log p(x; \theta) p(x; \theta) dx$$

( $E_{\theta}$  é a esperança com respeito à distribuição  $p_{\theta}$ ).

Seja  $\langle \mathbf{x}, \mathbf{y} \rangle_G = \mathbf{x}^t [g_{ij}(\theta)] \mathbf{y}$  o produto interno dado pela matriz da métrica  $G$ , o comprimento de arco de uma curva  $\alpha$  ligando  $\alpha(t_1) = \theta_1$  a  $\alpha(t_2) = \theta_2$  é dado por

$$\ell(\alpha) = \int_{t_1}^{t_2} (\langle \alpha'(t), \alpha'(t) \rangle_G)^{\frac{1}{2}} dt.$$

A distância de Fisher-Rao entre duas distribuições  $p_{\theta_1}$  e  $p_{\theta_2}$  é dada pelo menor comprimento de arco de uma curva em  $\Theta$  conectando  $\theta_1$  e  $\theta_2$  [1]. Nesse texto, vamos nos referir à distância entre as distribuições  $p_{\theta_1}$  e  $p_{\theta_2}$  como a distância entre os parâmetros  $\theta_1$  e  $\theta_2$ .

Uma distribuição Gaussiana multivariada é definida por

$$p(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \frac{(2\pi)^{-\left(\frac{n}{2}\right)}}{\sqrt{\text{Det}(\Sigma)}} \exp \left( -\frac{(\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}{2} \right),$$

na qual  $\mathbf{x}^t = (x_1, \dots, x_n) \in \mathbb{R}^n$  é o vetor de variáveis,  $\boldsymbol{\mu}^t = (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$  é vetor de médias e  $\Sigma \in P_n(\mathbb{R})$  é a matriz de covariância ( $P_n(\mathbb{R})$  é o conjunto das matrizes simétricas

positivas definidas de ordem  $n$ ). Consideremos  $\mathcal{M} = \{p_{\theta}; \theta \in \Theta = \mathbb{R}^n \times P_n(\mathbb{R})\}$  o modelo estatístico de dimensão  $\left(n + \frac{n(n+1)}{2}\right)$  formado por essas distribuições.

A equação da métrica da informação de Fisher de  $\mathcal{M}$ , ver referência [1], é

$$ds^2 = d\boldsymbol{\mu}^t \Sigma^{-1} d\boldsymbol{\mu} + \frac{1}{2} tr[(\Sigma^{-1} d\Sigma)^2],$$

na qual  $d\boldsymbol{\mu}^t = (d\mu_1, \dots, d\mu_n) \in \mathbb{R}^n$  e  $d\Sigma = [d\sigma_{ij}] \in P_n(\mathbb{R})$ .

Note que, para todo  $(\mathbf{c}, Q) \in \mathbb{R}^n \times GL_n(\mathbb{R})$ , com  $GL_n(\mathbb{R})$  o espaços das matrizes invertíveis de ordem  $n$ , a aplicação  $(\boldsymbol{\mu}, \Sigma) \mapsto (Q\boldsymbol{\mu} + \mathbf{c}, Q\Sigma Q^t)$  estabelece uma isometria em  $\Theta$ , ver referência [1]. Consequentemente, a distância de Fisher-Rao entre  $\boldsymbol{\theta}_1 = (\boldsymbol{\mu}_1, \Sigma_1)$  e  $\boldsymbol{\theta}_2 = (\boldsymbol{\mu}_2, \Sigma_2)$  em  $\mathcal{M}$  satisfaz, para todo  $(\mathbf{c}, Q) \in \mathbb{R}^n \times GL_n(\mathbb{R})$ ,

$$d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = d((\boldsymbol{\mu}_1, \Sigma_1), (\boldsymbol{\mu}_2, \Sigma_2)) = d((Q\boldsymbol{\mu}_1 + \mathbf{c}, Q\Sigma_1 Q^t), (Q\boldsymbol{\mu}_2 + \mathbf{c}, Q\Sigma_2 Q^t)). \quad (1)$$

Ainda não é conhecida uma expressão explícita para a distância entre duas distribuições multivariadas no caso geral. Uma fórmula fechada para a distância de Fisher-Rao em  $\mathcal{M}$ ,  $d_F$ , é conhecida somente para o caso univariado e em algumas subvariedades de  $\mathcal{M}$ .

Por exemplo, para  $n = 1$ , uma forma fechada para a distância de Fisher-Rao é dada via uma associação com o modelo do plano hiperbólico, ver referências [1, 2]. Consideramos a média  $\mu$  e o desvio padrão  $\sigma$  como parâmetros,  $\boldsymbol{\theta} = (\mu, \sigma)$ . No plano superior média  $\times$  desvio padrão,  $\mathbb{H}_F^2 = \mathbb{R} \times (0, +\infty)$ , uma expressão analítica para  $d_F$  é

$$d_F((\mu_1, \sigma_1), (\mu_2, \sigma_2)) = \sqrt{2} \log \frac{\left| \left( \frac{\mu_1}{\sqrt{2}}, \sigma_1 \right) - \left( \frac{\mu_2}{\sqrt{2}}, -\sigma_2 \right) \right| + \left| \left( \frac{\mu_1}{\sqrt{2}}, \sigma_1 \right) - \left( \frac{\mu_2}{\sqrt{2}}, \sigma_2 \right) \right|}{\left| \left( \frac{\mu_1}{\sqrt{2}}, \sigma_1 \right) - \left( \frac{\mu_2}{\sqrt{2}}, -\sigma_2 \right) \right| - \left| \left( \frac{\mu_1}{\sqrt{2}}, \sigma_1 \right) - \left( \frac{\mu_2}{\sqrt{2}}, \sigma_2 \right) \right|} \quad (2)$$

( $|\cdot|$  é a norma Euclidiana em  $\mathbb{R}^2$ ).

A distância de Fisher-Rao também é conhecida na subvariedade de  $\mathcal{M}$ , de dimensão  $2n$ , formada pelas distribuições cuja matriz de covariância é uma matriz diagonal dada por  $\mathcal{M}_D = \{p_{\theta}; \boldsymbol{\theta} = (\boldsymbol{\mu}, \Sigma), \Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2), \sigma_i > 0, i = 1, \dots, n\}$ . Considerando como parâmetro  $\boldsymbol{\theta} = (\mu_1, \sigma_1, \mu_2, \sigma_2, \dots, \mu_n, \sigma_n)$ , a métrica no espaço dos parâmetros de  $\mathcal{M}_D$  é uma métrica produto no espaço  $(\mathbb{H}_F^2)^n$ . Para  $\boldsymbol{\theta}_1 = (\mu_{11}, \sigma_{11}, \dots, \mu_{1n}, \sigma_{1n})$  e  $\boldsymbol{\theta}_2 = (\mu_{21}, \sigma_{21}, \dots, \mu_{2n}, \sigma_{2n})$ , a distância entre esses parâmetros, definida em [2], é

$$d_D(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \sqrt{\sum_{i=1}^n d_F((\mu_{1i}, \sigma_{1i}), (\mu_{2i}, \sigma_{2i}))^2}. \quad (3)$$

Em [6] propomos um limitante superior baseado na isometria dada em (1) e na distância sobre a subvariedade  $\mathcal{M}_D$ , (3). A Proposição abaixo diz que calcular a distância entre duas distribuições Gaussianas multivariadas quaisquer é o mesmo que calcular a distância entre uma distribuição Gaussiana que tem como média o vetor nulo e como matriz de covariância a identidade e outra distribuição cuja matriz de covariância é uma matriz diagonal.

**Proposição 2.1.** *Para todo par de distribuições Gaussianas multivariadas com parâmetros  $\theta_1 = (\mu_1, \Sigma_1)$  e  $\theta_2 = (\mu_2, \Sigma_2)$ , afirmamos:*

$$d_F(\theta_1, \theta_2) = d_F((\mathbf{0}, I_n), (\mu_3, D_3)), \quad (4)$$

na qual  $\mathbf{0}$  é o vetor de médias nulo,  $I_n$  é a matriz identidade,  $D_3$  é a matriz diagonal dada pelos autovalores de  $A = \Sigma_1^{-(1/2)} \Sigma_2 \Sigma_1^{-(1/2)}$ ,  $Q$  é a matriz ortogonal cujas colunas são os respectivos autovetores de  $A$  ( $A = QD_3Q^t$ ) e  $\mu_3 = Q^t \Sigma_1^{-(1/2)} (\mu_2 - \mu_1)$ .

Assim, considerando a distância na subvariedade  $\mathcal{M}_D$  e a Proposição acima obtemos uma fórmula fechada para um limitante superior para a distância de Fisher-Rao.

**Proposição 2.2.** *A distância de Fisher-Rao entre duas distribuições Gaussianas multivariadas  $\theta_1 = (\mu_1, \Sigma_1)$  e  $\theta_2 = (\mu_2, \Sigma_2)$  é limitada por,*

$$U(\theta_1, \theta_2) = \sqrt{\sum_{i=1}^n d_F((0, 1), (\mu_{3i}, D_{3i}))^2}, \quad (5)$$

sendo que  $\mu_{3i}$  são as coordenadas de  $\mu_3$  e  $D_{3i}$  são os elementos da diagonal de  $D_3$ , como estabelecido na proposição acima.

A prova da Proposição 2.1 e conseqüentemente da Proposição 2.2 encontra-se em [6].

### 3 $k$ -means baseado na distância de Fisher-Rao

O algoritmo  $k$ -means é um método de clustering, proposto por Stuart Lloyd em 1957, que particiona um conjunto de dados em  $k$  subconjuntos. Dado um conjunto de dados  $\mathcal{C}$ , o algoritmo começa com a escolha de  $k$  centros para o conjunto e depois associa-se a cada ponto de  $\mathcal{C}$  seu centro mais próximo em relação à alguma distância ou medida de divergência, formando grupos (clusters). Depois, atualiza-se os centros (centroide) de cada grupo até nenhum elemento mudar de grupo em duas iterações sucessivas.

No algoritmo original, a distância considerada é a distância Euclidiana porém, existem várias adaptações deste algoritmo nas quais usam-se a distância de Mahalanobis ou alguma medida de divergência, como a divergência de Bregman [3].

#### 3.1 Centroide na Subvariedade $\mathcal{M}_D$

Em [4], Schwander e Nielsen propuseram  $k$ -means para simplificar misturas Gaussianas univariadas usando a distância de Fisher-Rao, dada em (2), obtendo bons resultados. Para fazer as iterações do  $k$ -means eles definiram centroides no espaço  $\mathbb{H}_F^2$  usando o modelo centroide, dado por Galperin em 1993, para espaços de curvatura constante.

Dentre os modelos do espaço hiperbólico  $\mathbb{H}^2$  (disco de Poincaré, disco de Klein, Minkowski, semi-plano superior), Galperin calcula o centroide no modelo de Minkowski que é a folha superior do hiperbolóide  $z^2 = 1 + x^2 + y^2$ . Dado um ponto  $\mathbf{p}$ , com coordenadas

$(x_{\mathbf{p}}, y_{\mathbf{p}})$ , no disco de Klein, sua representação no modelo de Minkowski é dada por  $\mathbf{p}'$  com coordenadas:

$$\bar{x}_{\mathbf{p}'} = \frac{x_{\mathbf{p}}}{1 - x_{\mathbf{p}}^2 - y_{\mathbf{p}}^2}, \quad \bar{y}_{\mathbf{p}'} = \frac{y_{\mathbf{p}}}{1 - x_{\mathbf{p}}^2 - y_{\mathbf{p}}^2} \quad \text{e} \quad \bar{z}_{\mathbf{p}'} = \frac{z_{\mathbf{p}}}{1 - x_{\mathbf{p}}^2 - y_{\mathbf{p}}^2}.$$

Depois o centroide é computado e normalizado para pertencer ao modelo de Minkowski,

$$\mathbf{c}'' = \frac{1}{n} \sum_i \mathbf{p}'_i \quad \text{e} \quad \mathbf{c}' = \frac{\mathbf{c}''}{-x_{\mathbf{c}''}^2 - y_{\mathbf{c}''}^2 + z_{\mathbf{c}''}^2}$$

e, através da relação  $x_{\mathbf{c}} = x_{\mathbf{c}'}/z_{\mathbf{c}'}$  e  $y_{\mathbf{c}} = y_{\mathbf{c}'}/z_{\mathbf{c}'}$ , obtemos o centroide  $\mathbf{c}$  no disco de Klein. Para calcular o centroide  $\mathbf{c}$  de um conjunto de ponto  $\mathcal{C} = \{(\mu_i, \sigma_i)\} \subset \mathbb{H}_F^2$ , basta usar a relação entre  $\mathbb{H}_F^2$  e  $\mathbb{H}^2$  e converter os pontos de um modelo para o outro.

Como temos uma fórmula fechada para a distância na subvariedade  $\mathcal{M}_D$ , a qual é um limitante superior para a distância de Fisher-Rao no caso geral, propomos um centroide para distribuições nessa subvariedade. Foi visto na seção anterior que a distância em  $\mathcal{M}_D$  é dada pela métrica produto no espaço  $(\mathbb{H}_F^2)^n$ . Dado um conjunto de pontos  $\mathcal{C} = \{\boldsymbol{\theta}_i\} \subset (\mathbb{H}_F^2)^n$ ,  $\boldsymbol{\theta}_i = (\mu_{1i}, \sigma_{1i}, \dots, \mu_{ni}, \sigma_{ni})$ , definimos o centroide de  $\mathcal{C}$  como

$$\mathbf{c} := (\mathbf{c}_1, \dots, \mathbf{c}_n), \tag{6}$$

no qual  $\mathbf{c}_j$ ,  $j = 1, \dots, n$ , é o centroide de  $\mathcal{C}_j = \{(\mu_{ji}, \sigma_{ji})\} \subset \mathbb{H}_F^2$  definido anteriormente.

### 3.2 Simplificação de Misturas Gaussianas Diagonais

Uma mistura Gaussiana parametrizada  $f$  de  $M$  componentes é uma soma ponderada de  $M$  distribuições Gaussianas, isto é,

$$f(\mathbf{x}) = \sum_{i=1}^M w_i p_i(\mathbf{x}; \boldsymbol{\mu}_i, \Sigma_i),$$

em que  $\mathbf{x} \in \mathbb{R}^n$ ,  $p_i(\mathbf{x}; \boldsymbol{\mu}_i, \Sigma_i)$ ,  $i = 1, \dots, M$ , são as distribuições Gaussianas e  $w_i$ ,  $i = 1, \dots, M$ , são os pesos da mistura que satisfazem a restrição  $\sum_{i=1}^M w_i = 1$ . Chamamos de mistura Gaussiana diagonal àquela cuja matriz de covariância de cada componente é diagonal.

Para modelar um conjunto de dados através de uma mistura Gaussiana o algoritmo mais utilizado é o EM proposto por Dempster em 1977. O EM estima parâmetros de um conjunto de observações via maximização da função de verossimilhança. Neste trabalho, usamos a adaptação do algoritmo EM proposta por Banerjee para estimar parâmetros de uma mistura de famílias exponencial, o *Bregman soft clustering*, [3].

Em muitas aplicações que envolvem modelos de mistura, o custo computacional é muito alto devido o grande número de componentes da mistura. Esse custo pode ser fortemente diminuído se reduzirmos o número de componentes da mistura: dada uma mistura  $f$  de  $M$  componentes queremos encontrar uma mistura  $g$  de  $L$  componentes,  $1 \leq L < M$ , tal que  $g$  seja a melhor aproximação para  $f$  com respeito a alguma medida de similaridade.

A seguir resumimos o algoritmo  $k$ -means para simplificar misturas Gaussianas diagonais baseado na distância de Fisher-Rao na subvariedade  $\mathcal{M}_D$ .

**Inicialização:** Dado um conjunto de dados formado pelos pesos e os parâmetros da mistura,  $\mathcal{C} = \{(w_1, \boldsymbol{\theta}_1), \dots, (w_M, \boldsymbol{\theta}_M)\}$ , escolhamos aleatoriamente  $L$  pontos de  $\mathcal{C}$ ,  $\mathbf{c}_j = (\hat{w}_j, \hat{\boldsymbol{\theta}}_j)$ , com  $j \in \{1, \dots, L\}$ , para os centroides iniciais.

**Agrupamento:** Dado  $L$  centroides  $\mathbf{c}_j = (\hat{w}_j, \hat{\boldsymbol{\theta}}_j)$ , com  $j \in \{1, \dots, L\}$ , dizemos que o ponto  $(w_i, \boldsymbol{\theta}_i)$  pertence ao cluster  $\mathcal{C}_j$  quando  $d_D(\boldsymbol{\theta}_i, \hat{\boldsymbol{\theta}}_j) \leq d_D(\boldsymbol{\theta}_i, \hat{\boldsymbol{\theta}}_m)$ , com  $m \in \{1, \dots, L\}$ .

**Atualização:** Atualizamos o centroide  $\mathbf{c}_j$  de cada cluster  $\mathcal{C}_j$  usando o centroide definido em (6) e fazendo  $\hat{w}_j = \sum_i w_i$ , com  $(w_i, \boldsymbol{\theta}_i) \in \mathcal{C}_j$ .

O algoritmo termina quando o centro de cada grupo não muda em duas iterações sucessivas.

Vamos simplificar misturas Gaussianas diagonais no cenário de segmentação de imagens. Dada uma imagem colorida  $I$  de entrada, adaptamos o algoritmo *Bregman soft clustering*, dado em [3], para gerar uma mistura Gaussiana diagonal  $f$  de 32 componentes que modela os pixels da imagem. Consideramos um pixel  $\boldsymbol{\rho} = (\rho_R, \rho_G, \rho_B)$  como um ponto no  $\mathbb{R}^3$ , no qual  $\rho_R, \rho_G, \rho_B$  são as informações RGB de cada pixel. Para a segmentação da imagem, dizemos que cada pixel da imagem  $\boldsymbol{\rho}$  pertence a classe  $\mathcal{C}_j$  se  $p_j(\boldsymbol{\rho}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) > p_i(\boldsymbol{\rho}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ ,  $\forall i \in \{1, \dots, M\} \setminus \{j\}$ . Sendo assim, a imagem segmentada é dada pela troca da informação de cor do pixel  $\boldsymbol{\rho}$  pela média  $\boldsymbol{\mu}_j$  da Gaussiana  $p_j$ .

Usando o  $k$ -means baseado na distância de Fisher-Rao na subvariedade  $\mathcal{M}_D$ , simplificamos a mistura  $f$  em uma mistura Gaussiana  $g$  de  $L$  componentes com  $L = \{2, 4, 8, 16\}$ , cada mistura nos dá uma segmentação da imagem. As imagens utilizados foram Babon e Lena. A Figura 1 ilustra a segmentação, o número de cores em cada figura é igual ao número de componentes da  $g$ .



Figura 1: Ilustração da simplificação da mistura em  $L$  componentes.

Analisamos a qualidade da simplificação da mistura em função de  $L$ , através da aproximação da divergência Kullback-Leibler,  $d_{KL}(f||g)$ , pelo método de Monte-Carlo. Observamos que a qualidade aumenta (isto é,  $d_{KL}(f||g)$  diminui) quando  $L$  aumenta em ambas as figuras, ver Figura (2). Analisamos também a qualidade da segmentação da imagem pelo clássico índice PSNR (Peak Signal-to-Noise Ratio). A Figura (2) mostra que qualidade da segmentação também aumenta quando aumentamos  $L$ , como era de se esperar.

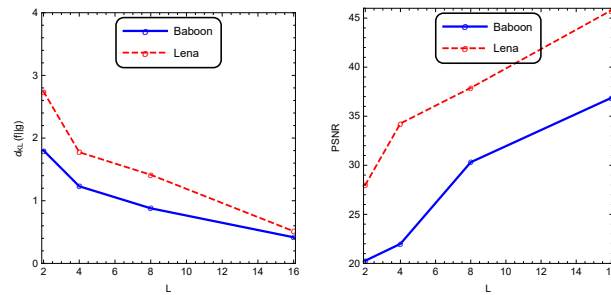


Figura 2: O gráfico à esquerda ilustra a qualidade da simplificação ( $d_{KL}(f||g)$ ) em função de  $L$  e o gráfico à direita ilustra a qualidade da segmentação (PSNR) em função da  $L$ .

O nosso algoritmo é similar ao Bregman hard clustering [3] em termos de eficiência computacional uma vez que ambos usam apenas fórmulas fechadas em cada iterações. Neste trabalho, usamos uma distância como medida de similaridade e, em [3], os autores usam uma divergência.

## Agradecimentos

Agradecemos ao CNPq pelo suporte financeiro e ao IMECC pelo incentivo à pesquisa.

## Referências

- [1] J. Burbea. Informative geometry of probability spaces, *Expositiones Mathematica*, 4:347–378, 1986.
- [2] S. I. Costa, S. A. Santos and J. E. Strapasson. Fisher information distance: a geometrical reading, *Discrete Applied Mathematics*, 197:59–96, 2015.
- [3] V. Garcia and F. Nielsen. Simplification and hierarchical representations of mixtures of exponential families, *Signal Processing*, 90:3197–3212, 2010.
- [4] O. Schwander and F. Nielsen. Model centroids for the simplification of kernel density estimators, Acoustics, Speech and Signal Processing (ICASSP), *IEEE International Conference on*, volume 1, 2012. DOI: 10.1109/ICASSP.2012.6287989.
- [5] L. T. Skovgaard, A Riemannian geometry of the multivariate normal model, *Scand. J. Statist.*, 11:211–223, 1984.
- [6] J. E. Strapasson, J. Porto e S. I. Costa. *On bounds for the Fisher-Rao distance between multivariate normal distributions*, Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MAXENT 2014), Vol. 1641, AIP Publishing, 2015.