Optimal interval clustering: Application to Bregman clustering and statistical mixture learning

arXiv 1403.2485

Frank Nielsen, *Sony Computer Science Laboratories Inc.*
Richard Nock, *NICTA - The Australian National University*

Partition $\mathcal{X} = \{x_1, ..., x_n\}$ into k pairwise disjoint clusters $\mathcal{C}_1 \subset \mathcal{X}, ..., \mathcal{C}_k \subset \mathcal{X}$: $\mathcal{X} = \uplus_{i=1}^{k} \mathcal{C}_i$

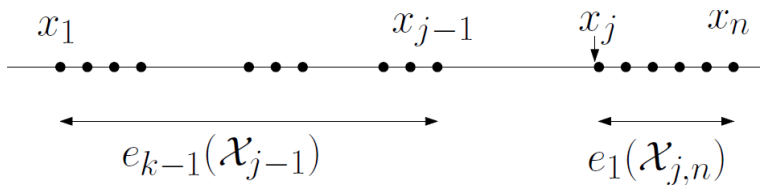
k-means: Minimize the sum of intra-cluster variances: $\min_{p_1,...,p_k} \sum_{i=1}^{n} \min_{j=1}^{k} \|x_i - p_j\|^2$

k-means is NP-hard when $d > 1$ and $k > 1$, $O(n^2 k)$ when $d = 1$

Interval clustering (contiguous clustering)

$\min e_k(\mathcal{X}) = \oplus_{j=1}^{k} e_1(\mathcal{C}_j)$

$$\underbrace{[x_1 ... x_{l_2-1}]}_{\mathcal{C}_1} \underbrace{[x_{l_2} ... x_{l_3-1}]}_{\mathcal{C}_2} ... \underbrace{[x_{l_k} ... x_n]}_{\mathcal{C}_k}$$



Solve by **Dynamic Programming**

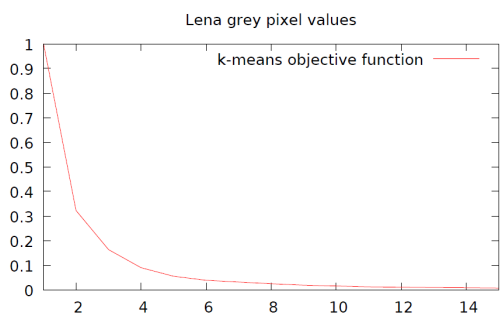
$$e_{i,m} = \min_{m \leq j \leq i} \{e_{j-1,m-1} \oplus e_1(\mathcal{X}_{j,i})\}$$

Optimization: Look-up-tables or summed area tables

Cluster size constraints

$$e_{i,m} = \min_{\max\{1+\sum_{l=1}^{m-1} n_l^-, i+1-n_m^+\} \leq j \; j \leq i+1-n_m^-} \{e_{j-1,m-1} \oplus e_1(\mathcal{X}_{j,i})\},$$

Model selection from the DP table



→ **A Voronoi cell condition for DP optimality**

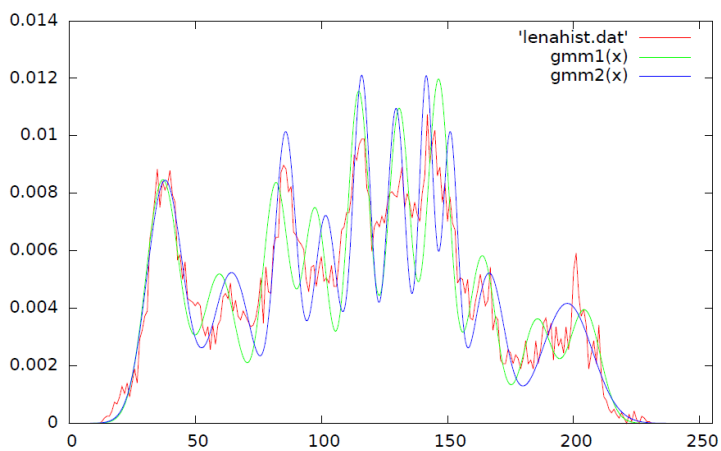
$\mathcal{F} = \{p_F(x;\theta) \; : \; \theta \in \Theta\}$ an exponential family:
$p_F(x|\theta) = \exp\left(t(x)\theta - F(\theta) + k(x)\right)$

Mixture maximizing the complete log-likelihood:

$$l_c(\mathcal{X}; L, \Omega) = \sum_{i=1}^{n} \log(\alpha_{l_i} p(x_i; \theta_{l_i})),$$

$L = \{l_i\}_i$:hidden labels.

$$\max l_c \equiv \min_{\theta_1,...,\theta_k} \sum_{i=1}^{n} \min_{j=1}^{k} (-\log p(x_i;\theta_j) - \log \alpha_j).$$



Duality exponential family ↔ Bregman divergence