Multiple Clustering Views via Constrained Projections

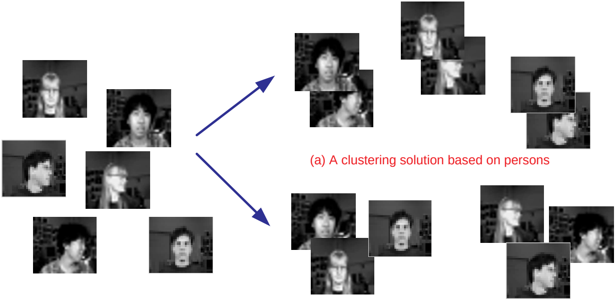
Xuan-Hong Dang, Aarhus University, Denmark
(joint work with Ira Assent, Aarhus University, and James Bailey, University of Melbourne)

28 avril 2012

Outlines

Introduction

- Clustering : categorizes similar objects into same groups
- High dimensional data → Multiple clusterings may exist.



(a) A clustering solution based on persons

(b) A clustering solution based on poses

- Other data : text/document data, gene data,...
- Challenge : How to find all meaningful solutions ?

Alternative Clustering Problem

Several algorithms have been developed.

- Seeking alternative clusterings simultaneously.
  Eg : Maximize $L(\Theta^{(1)}; \mathcal{X}) + L(\Theta^{(2)}; \mathcal{X}) - I(C^{(1)}; C^{(2)}|\Theta)$
- Seeking alternative clusterings in sequence.
  Eg : Maximize $L(\Theta^{(2)}; \mathcal{X}) - I(C^{(1)}; C^{(2)})$
  $\Rightarrow$ model view point : latter approach has limited number of parameters optimized
  $\Rightarrow$ our approach in this work

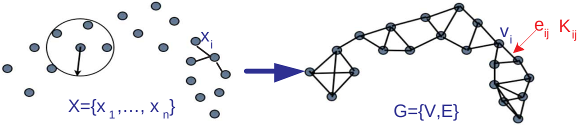
Given $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ in $\mathbb{R}^d$ and $C^{(1)}$ as reference, seek $C^{(2)}$ as an alternative : $\bigcup_i C_i^{(2)} = \mathcal{X}$ and $C_i^{(2)} \cap C_j^{(2)} = \emptyset$ for $\forall i \neq j;\ i, j \leq k$

Objective

Subspace learning :

- un-correlate from $C^{(1)}$ $\Rightarrow$ ensure difference.
- retaining local data proximity $\Rightarrow$ ensure quality.

With graph based approach :



- F : maps $\{\mathbf{x}_i\}_{i=1}^{n}$ into $\{\mathbf{y}_i\}_{i=1}^{n}$ (i.e., $Y = F^T X$)
  $\Rightarrow$ $\mathbf{f}$ in F combines X into 1-dim : $\mathbf{f}^T X = \{y_1, \ldots, y_n\} = \mathbf{y}^T$.
- Define objective :

$$\arg\min_{\mathbf{f}} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\mathbf{f}^T \mathbf{x}_i - \mathbf{f}^T \mathbf{x}_j)^2 K_{ij} \quad \text{s.t. } S^T X^T \mathbf{f} = 0$$

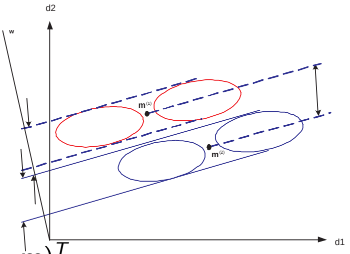
$\Rightarrow$ *S is a feature subspace capturing $C^{(1)}$*

$\Rightarrow$ *Penalize : K_{ij} large but y_i, y_j are mapped far apart*

Learn S with LDA

- Learn S as a subspace best capturing $C^{(1)}$.
  $\Rightarrow$ *$C^{(1)}$'s clusters represented in S are most separable.*

- Fisher LDA is a good choice :



$$\max_{\mathbf{w}} \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}$$

where

$$S_B = \sum_k n_k(\mathbf{m}^{(k)} - \mathbf{m})(\mathbf{m}^{(k)} - \mathbf{m})^T$$

$$S_W = \sum_k \sum_i^{n_k} (\mathbf{x}_i^{(k)} - \mathbf{m}^{(k)})(\mathbf{x}_i^{(k)} - \mathbf{m}^{(k)})^T$$

- Optimal $\mathbf{w}$'s are eigenvectors of $S_W^{-1} S_B$
- S is chosen with leading $\mathbf{w}$'s.

Solving Constrained Function(1)

- Define D with $D_{ii} = \sum_j K_{ij}$ and $L = D - K$
- Deploying summation :

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{f}^T\mathbf{x}_i - \mathbf{f}^T\mathbf{x}_j)^2 K_{ij} = \mathbf{f}^T XLX^T\mathbf{f}$$

- Adding $\mathbf{f}^T XDX^T\mathbf{f} = 1$ to remove $\mathbf{f}$'s freedom :

$$\mathcal{L}(\alpha,\beta,\mathbf{f}) = \mathbf{f}^T XLX^T\mathbf{f} - \alpha(\mathbf{f}^T XDX^T\mathbf{f} - 1) - \beta S^T X^T\mathbf{f}$$

- For simplicity :

$$\begin{cases} \widetilde{L} = XLX^T \\ \widetilde{D} = XDX^T \\ \widetilde{S} = XS \end{cases}$$

Solving Constrained Function(2)

- $\widetilde{D}$ is symmetric, pos.semi-definite. Change $\mathbf{f} = \widetilde{D}^{-1/2}\mathbf{z}$:

$$\mathbf{f}^T\widetilde{L}\mathbf{f} = \mathbf{z}^T\widetilde{D}^{-1/2}\widetilde{L}\widetilde{D}^{-1/2}\mathbf{z} = \mathbf{z}^TQ\mathbf{z}$$

and two constraints : $\begin{cases} \mathbf{f}^T\widetilde{D}\mathbf{f} = \mathbf{z}^T\mathbf{z} = 1 \\ \widetilde{S}^T\mathbf{f} = \widetilde{S}^T\widetilde{D}^{-1/2}\mathbf{z} = 0 \end{cases}$

- Lagrange function can be re-written :

$$\mathcal{L}(\alpha, \beta, \mathbf{z}) = \frac{1}{2}\mathbf{z}^TQ\mathbf{z} - \frac{1}{2}\alpha(\mathbf{z}^T\mathbf{z} - 1) - \beta U^T\mathbf{z}$$

where $U^T = \widetilde{S}^T\widetilde{D}^{-1/2}$.

- Taking derivative and with little algebra :

$$\begin{aligned} \alpha\mathbf{z} &= Q\mathbf{z} - U(U^TU)^{-1}U^TQ\mathbf{z} \\ &= \left(I - U(U^TU)^{-1}U^T\right)Q\mathbf{z} \\ &= PQ\mathbf{z} \end{aligned}$$

$\Rightarrow$ *eigenvalue problem*

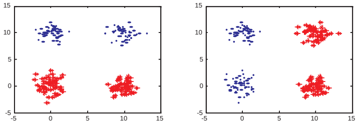
Solving Constrained Function(3)

- Solving : $\alpha\mathbf{z} = PQ\mathbf{z}$
- Notice PQ might not be symmetric ; yet, $\alpha(PQ) = \alpha(PQP)$ due to $P^T = P$ and $P^2 = P$.
  $\Rightarrow$ *not solving* $PQ\mathbf{z} = \alpha\mathbf{z}$ *but* $PQP\mathbf{v} = \alpha\mathbf{v}$, *with* $\mathbf{v} = P^{-1}\mathbf{z}$
- Eigenvalues of PQP are non-negative :
  $\Rightarrow \alpha_0 = 0$ is smallest
  $\Rightarrow \mathbf{v}_0 = P^{-1}\widetilde{D}^{1/2}\mathbf{1}$ is trivial
- Optimal direction $\mathbf{f}$:

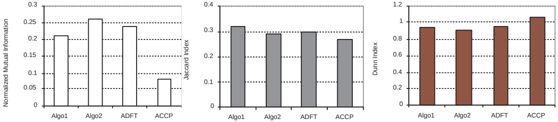
$$\mathbf{f} = \widetilde{D}^{-1/2}P\mathbf{v}$$

  with corresponding smallest non-zero eigenvalue α.
  $\Rightarrow F$ is formed based on q leading eigenvectors of PQP corresponding to smallest non-zero α's.

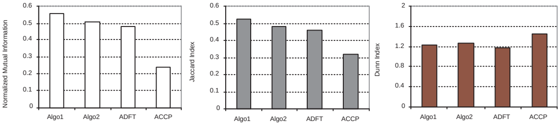
Initial experimental results



(a) Synthetic data with 4 Gaussians



(b) Cloud data from UCI repository



(c) Housing data from UCI repository

Conclusions

- Novel approach from subspace learning
  - not only being uncorrelated from provided clustering
  - but also retaining local geometrical data proximity
- Global optimum solution can be achieved
- Capability of seeking multiple clusterings (adding more subspaces into S).
- The approach is extendable for non-linear cases.
- Future work :
  - More experiments required on diverse datasets
  - Soft constraint with tradeoff factor (subspace independence vs. local data structure retaining)
  - Alternative clustering interpretation.