On Spectral Clustering: Analysis and an algorithm

Andrew Ng, Michael Jordan, Yair Weiss

Presenter: Robert Luke

The Algorithm

$S = \{s_i,\ldots,s_n\}$

1. Affinity Matrix-
   $A_{ij}=\exp(-||s_i-s_j||^2/2\sigma^2)$ if $i \neq j$
   $A_{ii}=0$
2. $M_i=\Sigma_{j=1,n} A_{ij}$
   $D=\text{diag}(M)$
   $L = D^{-1/2}AD^{-1/2}$

   $D^{-1/2}=\text{diag}(\text{sqrt}(1/M))$

Algorithm (contd.)

3. Find k largest eigenvectors of L. Concatenate the vectors into a matrix X.
4. $Y_{ij} = X_{ij} / (\Sigma_{j=1,n} X_{ij}^{2})^{½}$
5. Cluster Y using your favorite algorithm (they use K-means)
6. Assign the original point s_i to cluster j if row i of Y was assigned to cluster j.

Analysis

- Ideal Case
  $S = S_1 \mathbf{U}\, S_2 \mathbf{U}\, S_3 \quad n = n_1 + n_2 + n_3$
  Clusters are "infinitely" far apart

$$\hat{A} = \begin{bmatrix} A^{(11)} & 0 & 0 \\ 0 & A^{(22)} & 0 \\ 0 & 0 & A^{(33)} \end{bmatrix} \quad \hat{L} = \begin{bmatrix} \hat{L}^{(11)} & 0 & 0 \\ 0 & \hat{L}^{(22)} & 0 \\ 0 & 0 & \hat{L}^{(33)} \end{bmatrix}$$

$$\hat{A}^{(ii)} = A^{(ii)} \in \mathbb{R}^{n_i \times n_i}.$$

$$\hat{L}^{(ii)} = (\hat{D}^{(ii)})^{-1/2} A^{(ii)} (\hat{D}^{(ii)})^{-1/2}.$$

Analysis

- Get 3 none zero eigenvalues (k=3)

$$\hat{X} = \begin{bmatrix} x_1^{(1)} & \vec{0} & \vec{0} \\ \vec{0} & x_1^{(2)} & \vec{0} \\ \vec{0} & \vec{0} & x_1^{(3)} \end{bmatrix} \in \mathbb{R}^{n \times 3}$$

$$\hat{Y} = \begin{bmatrix} \hat{Y}^{(1)} \\ \hat{Y}^{(2)} \\ \hat{Y}^{(3)} \end{bmatrix} = \begin{bmatrix} \vec{1} & \vec{0} & \vec{0} \\ \vec{0} & \vec{1} & \vec{0} \\ \vec{0} & \vec{0} & \vec{1} \end{bmatrix} R$$

- R is based on the eigen solver and is therefore not unique. But still a unique subspace.

Analysis

Proposition 1 *Let $\hat{A}$'s off-diagonal blocks $\hat{A}^{(ij)}$, $i \neq j$, be zero. Also assume that each cluster S_i is connected.*[2] *Then there exist k orthogonal vectors $r_1, \ldots, r_k$ ($r_i^T r_j = 1$ if $i = j$, 0 otherwise) so that $\hat{Y}$'s rows satisfy*

$$\hat{y}_j^{(i)} = r_i \tag{4}$$

for all $i = 1, \ldots, k$, $j = 1, \ldots, n_i$.

- There are k mutually orthogonal points on the surface of the unit k-sphere around which Ŷ's rows are will cluster. They also correspond to the true clustering of the original data.

Analysis

- General Case
  - A's off-diagonal blocks will be non-zero but small.
  - A=Â+E (perturb the "ideal" by some amount)
  - When will the eigenvectors of L be "close" to the eigenvectors of Ĺ?

Analysis

- Matrix perturbation theory.
- Stability of eigenvectors is based on eigengap. $\delta=|\lambda_3-\lambda_4|$
- If $\lambda_j^{(i)}$ is the j-th largest eigenvalue of $\acute{L}^{(ii)}$, $\lambda_4=\max_i \lambda_2^{(i)}$.
- Want λ_4 to be near zero.

The Assumptions

- A cluster should not have sub-clusters.

Assumption A1. There exists $\delta > 0$ so that, for all $i = 1, \dots, k$, $\lambda_2^{(i)} \leq 1 - \delta$.

Assumption A1.1. Define the *Cheeger constant* [3] of the cluster S_i to be

$$h(S_i) = \min_{\mathcal{I}} \frac{\sum_{j \in \mathcal{I}, k \notin \mathcal{I}} A_{jk}^{(ii)}}{\min\{\sum_{j \in \mathcal{I}} \hat{d}_j^{(i)}, \sum_{k \notin \mathcal{I}} \hat{d}_k^{(i)}\}}. \quad (5)$$

where the outer minimum is over all index subsets $\mathcal{I} \subseteq \{1, \dots, n_i\}$. Assume that there exists $\delta > 0$ so that $(h(S_i))^2/2 \geq \delta$ for all i.

The Assumptions

- Between Class Affinity

Assumption A2. There is some fixed $\epsilon_1 > 0$, so that for every $i_1, i_2 \in \{1, \ldots, k\}$, $i_1 \neq i_2$, we have that

$$\sum_{j \in S_{i_1}} \sum_{k \in S_{i_2}} \frac{A_{jk}^2}{\hat{d}_j \hat{d}_k} \leq \epsilon_1. \tag{6}$$

- Within Class Affinity

Assumption A3. For some fixed $\epsilon_2 > 0$, for every $i = 1, \ldots, k$, $j \in S_i$, we have:

$$\frac{\sum_{k: k \notin S_i} A_{jk}}{\hat{d}_j} \leq \epsilon_2 \left(\sum_{k,l \in S_i} \frac{A_{kl}^2}{\hat{d}_k \hat{d}_l} \right)^{-1/2} \tag{7}$$

The Assumptions

- Don’t want outliers.

Assumption A4. There is some constant $C > 0$ so that for every $i = 1, \ldots, k$, $j = 1, \ldots, n_i$, we have $\hat{d}_j^{(i)} \geq (\sum_{k=1}^{n_i} \hat{d}_k^{(i)})/(Cn_i)$.

Theorem

Theorem 2 *Let assumptions A1, A2, A3 and A4 hold. Set* $\epsilon = \sqrt{k(k-1)\epsilon_1 + k\epsilon_2^2}$. *If* $\delta > (2+\sqrt{2})\epsilon$, *then there exist* k *orthogonal vectors* $r_1, \ldots, r_k$ ($r_i^T r_j = 1$ *if* $i = j$, *0 otherwise) so that* Y*'s rows satisfy*

$$\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} ||y_j^{(i)} - r_i||_2^2 \leq 4C\left(4 + 2\sqrt{k}\right)^2 \frac{\epsilon^2}{(\delta - \sqrt{2}\epsilon)^2}. \quad (8)$$

- The rows of Y will form tight clusters around k well-separated points on the surface of the k-sphere.

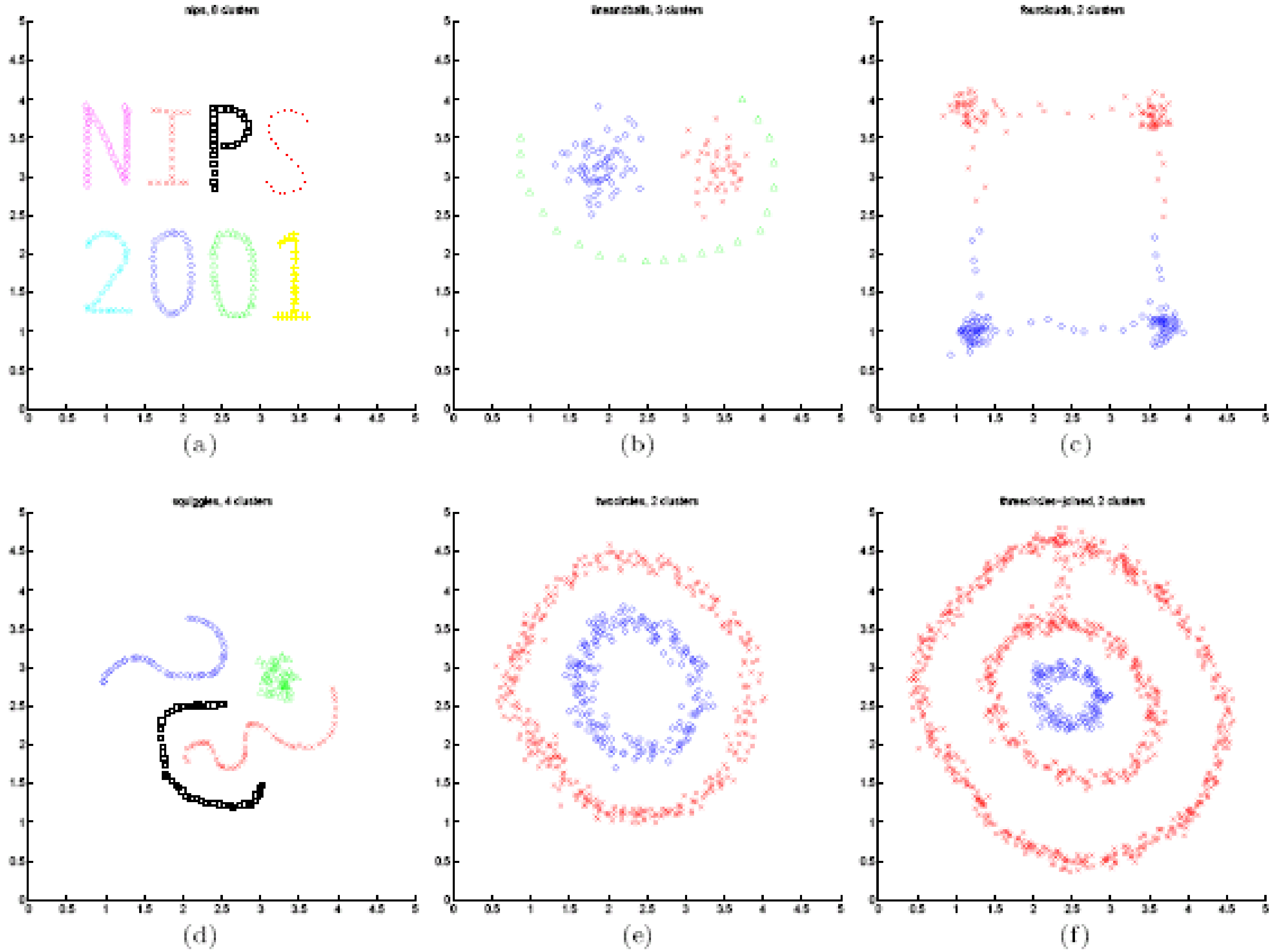
Free Parameters

- K-The number of clusters
  - Set by the user.
  - No automation explained.
- σ-Affinity parameter
  - Can be automatically found.
  - Search over σ and keep the tightest clusters

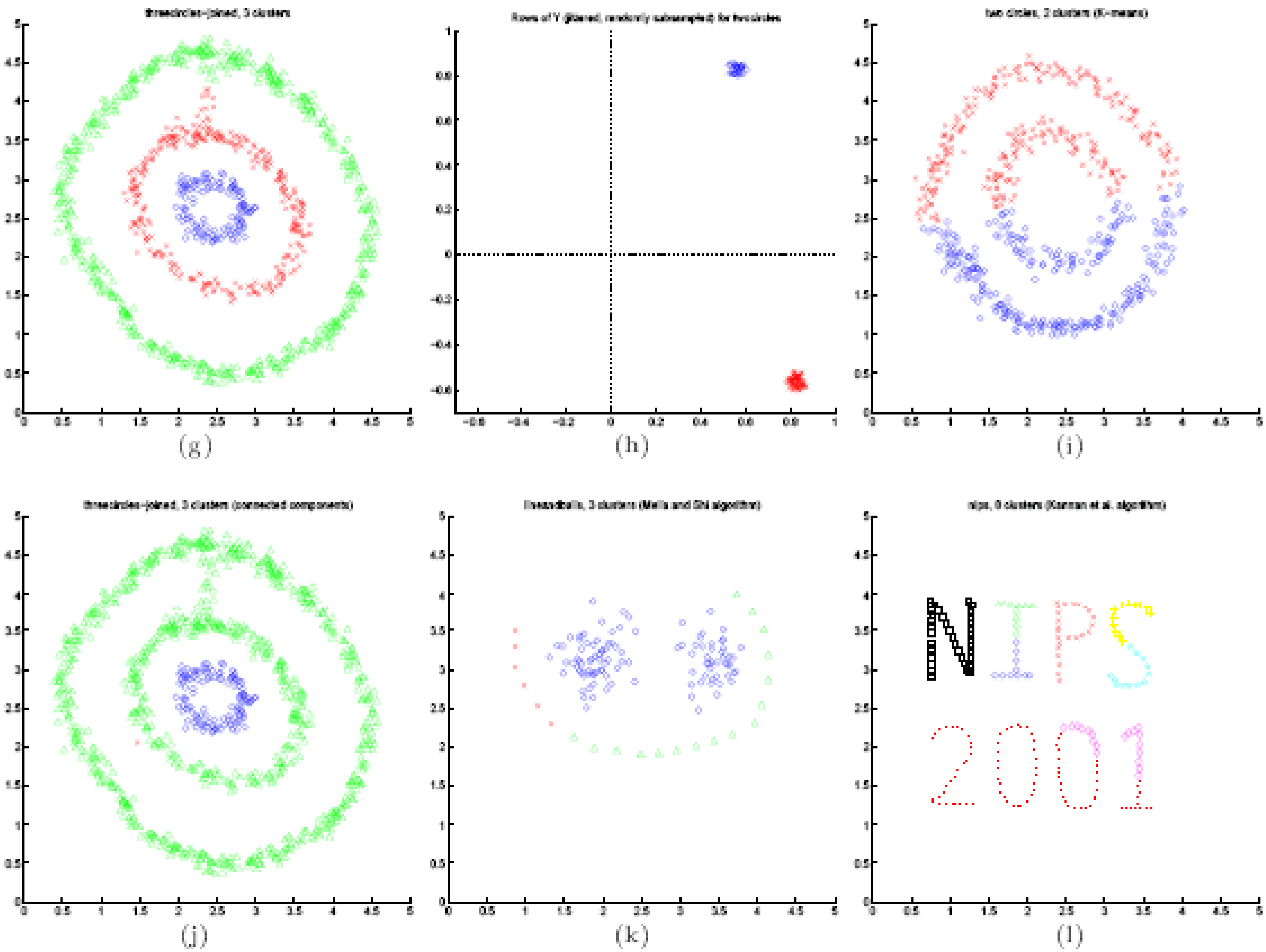
Code

```
function labels = SpectralCluster(data,k)
   sig = 1;
   sigSq2 = 2*sig*sig;

   s = size(data);
   affinity = zeros(s(1));

   a = reshape(data,1,s(1),s(2));
   a2 = repmat(a,[s(1) 1 1]);
   b = reshape(data,s(1),1,s(2));
   b2 = repmat(b,[1 s(1) 1]);

   diff = a2-b2;
   diff2 = diff.*diff;
   diffSum = sum(diff2,3);
   affinity = exp(-diffSum/sigSq2);
   affinity(eye(s(1)) == 1) = 0;
   L = affinity./(sum(affinity,2)*ones(1,s(1)));

   [V,EigVals] = eig(L);
   EV = V(:,1:k);
   EVN = zeros(s(1),1); for i=1:s(1) EVN(i) = norm(EV(i,:)); end;
   EVN2 = EV./repmat(EVN,1,k);
   labels = kmeans(EVN2,k);
return
```

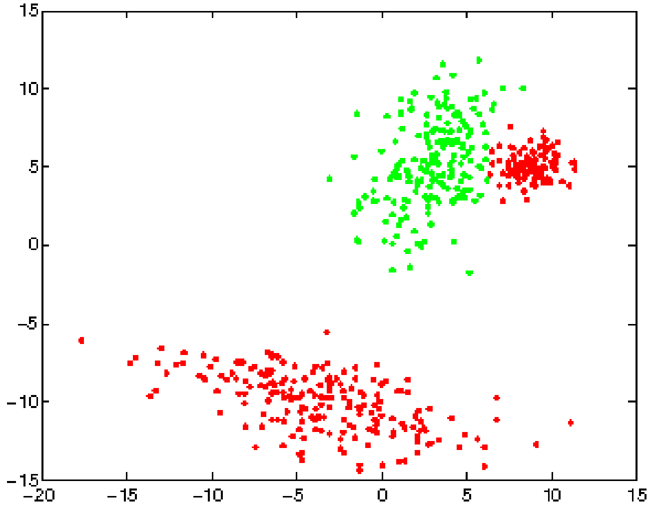
Experiments

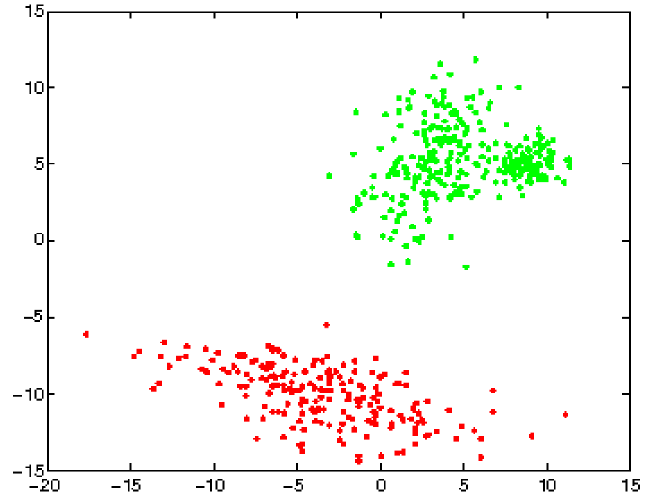


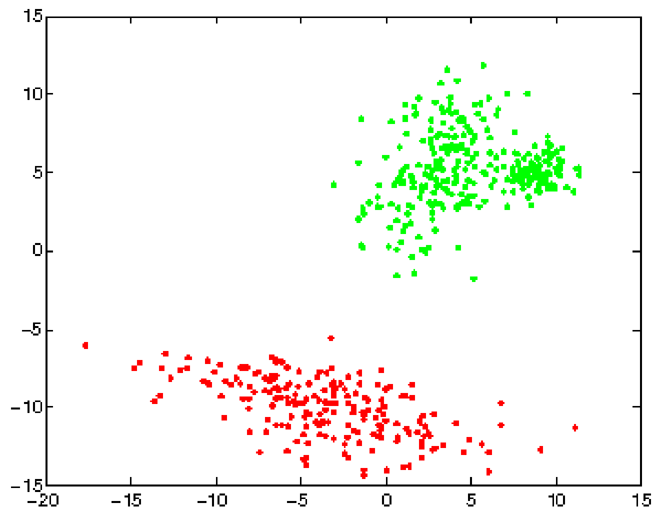
Experiments



My Own Results







More Results