July 16, 2014

- Approximate Ncut for k = 2:

$$\min_{A} Ncut(A, \overline{A})$$

- Given a subset $A \subset V$, define $x = (x_1, \ldots, x_n)$

$$x_i = \begin{cases} \sqrt{vol(\overline{A}) \,/\, vol(A)} & if \;\; v_i \in A \\ -\sqrt{vol(A) \,/\, vol(\overline{A})} & if \;\; v_i \in \overline{A} \end{cases}$$

- Then one can prove that

$$x^T L x = vol(V) \cdot Ncut(A, \overline{A})$$

$$\sum_{i=1}^{n} d_i x_i = 0 \qquad i.e. \qquad (Dx)^T 1 = 0$$

$$x^T D x = vol(V)$$

- The min problem becomes:

$$\min_{A} x^T L x$$

subject to x_i as defined above, $Dx \perp \mathbf{1}$, and $x^T D x = vol(V)$.

- Relaxed min problem:

$$\min_{x \in R^n} x^T L x$$

subject to $Dx \perp \mathbf{1}$, and $x^T D x = vol(V)$.

- Define $y = D^{1/2} x$. The relaxed problem becomes:

$$\min_{x \in R^n} y^T D^{1/2} L D^{1/2} y \equiv \min_{x \in R^n} y^T \hat{L} y$$

subject to $y \perp D^{1/2}$ **1**, and $||y||^2 = vol(V)$.

- Solution of the relaxed min problem: the eigenvector corresponding to the second smallest eigenvalue of L^.

<u>K-means clustering</u>

- Clustering for k = 2. From the solution vector x (or y), we need to find a partition. For example,

$$\begin{cases} v_i \in A & \text{if } x_i \geq 0 \ (\text{or } y_i \geq 0) \\ v_i \in \overline{A} & \text{if } x_i < 0 \ (\text{or } y_i < 0) \end{cases}$$

- It does not work for k > 2.

- K-mean clustering: Given a set of n data points $\{p_j\}$, find partitions $A_1, A_2, \ldots, A_k$ which solve the min problem

$$\min_{\{A_i\}} \sum_{i=1}^{k} \sum_{p \in A_i} \| p - \mu_i \|_2^2$$

1. Start with an initial guess for the k means $\{\mu_i\}$.
2. Assign p to A_i if p is closest to μ_i.
3. Update $\{\mu_i\}$ using the new partitions $\{A_i\}$.
4. Repeat (1)-(3).

- For the case k = 2. Consider $\{x_i\}$ as n points in R. Apply the k-means algorithm to cluster the points into 2 groups.

Unnormalized spectral clustering

- Construct the weighted adjacency matrix W.
- Compute the unnormalized L.
- Compute the first k eigenvectors $q_1, \ldots, q_k$ of L.
- Consider $Q_k = [\, q_1, \ldots, q_k \,]$. Let $p_i \in R^k$ be the vector of row(i) of Q_k.
- For the n points $\{p_i\}$ in R^k, apply the k-means algorithm to cluster them into k groups: $\{A_1, \ldots, A_k\}$.

Normalized spectral clustering

- Change L to L^.
- Change p_i:

$$\overline{p}_i = \frac{p_i}{\|p_i\|}$$

How to define W?

- It is problem dependent.

- For image segmentation, W has a similar nonzero structure as the 2D Laplacian, if we only consider 4 neighbours.

- If we also include the 4 neighbours at the corners, there will be 8 neighbours. Then W has (at most) 8 nonzeros per row.

- In general, more neighbours $\Rightarrow$ more nonzeros in W.

- In practice, only consider a small number of neighbours. E.g. neighbours that are within graph distance 1 or 2.

- w_{ij} measures the similarity between pixel i and pixel j:
  - distance between pixel i and pixel j.
  - intensity difference between pixel i and pixel j.

- For example

$$w_{ij} = e^{-\frac{\left\|x_i - x_j\right\|^2}{\sigma_{dist}^2}} e^{-\frac{\left\|I_i - I_j\right\|^2}{\sigma_{\text{int}}^2}}$$

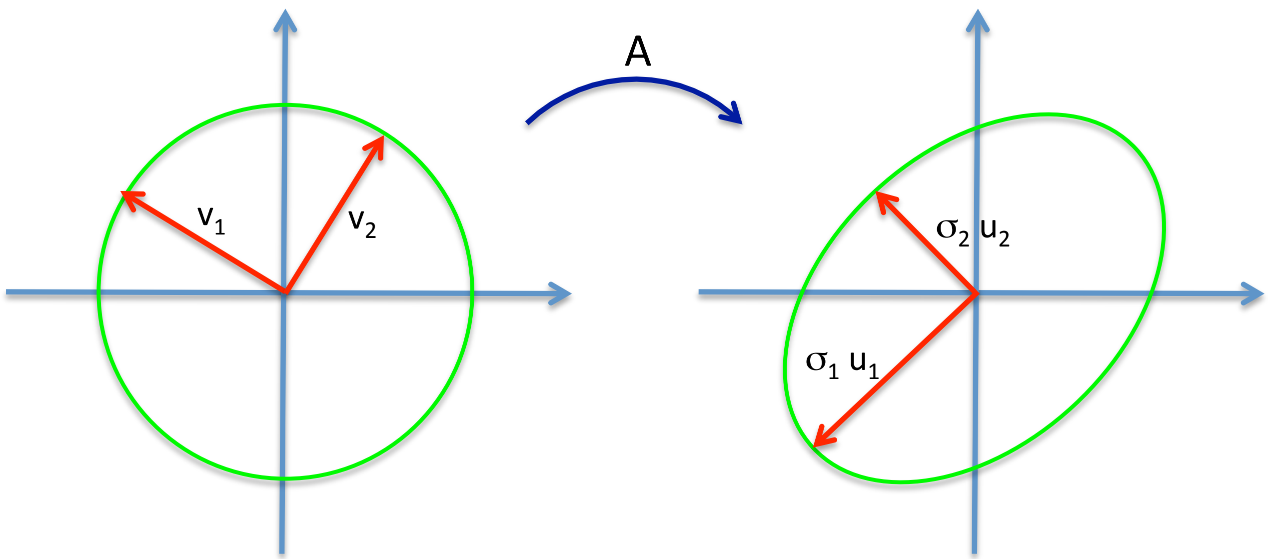
where pixel i is located at x_i with intensity I_i and pixel j is located at x_j with intensity I_j. Here $x_i = (r, c)$, if pixel i is at row r and column c.

Singular Value Decomposition

A geometric observation

The image of the unit circle sphere under any m×n matrix is a hyperellipse.

e.g.



Let S be the unit sphere in R^n. The image AS is an ellipse in R^m.

- The n singular values of A are the lengths of the n principal semiaxes of AS: $\sigma_1, \sigma_2, \ldots, \sigma_n$. Convention: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$.

- The n left singular vectors of A are the unit vectors $\{ u_1, \ldots, u_n \}$ in the direction of the principal semiaxes.

- The n right singular vectors of A are the unit vectors $\{ v_1, \ldots, v_n \} \in$ S such that $A\, v_j = \sigma_j\, u_j$.

Reduced SVD

$$A\, v_j \;=\; \sigma_j\, u_j \qquad\qquad j = 1, 2, \ldots, n$$

i.e.

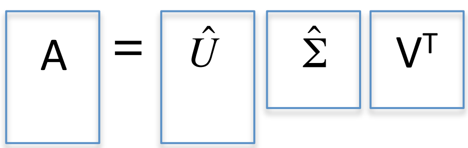
$$\begin{bmatrix} & A & \\ & & \\ & & \end{bmatrix}_{m\times n} \underset{V}{\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}_{n\times n}} = \underset{\hat{U}}{\begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix}_{m\times n}} \underset{\hat{\Sigma}}{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}_{n\times n}}$$

$$\therefore \qquad AV \;=\; \hat{U}\hat{\Sigma}$$

where $\hat{\Sigma}$ = diag, $\hat{U}$ and V have orthonormal columns.

Equivalently,

$$A \;=\; \hat{U}\hat{\Sigma}V^T \qquad\qquad \text{reduced SVD}$$

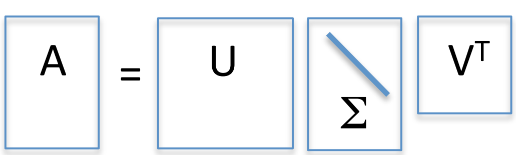


Full SVD

- Extend $\hat{U} \rightarrow$ U = orthogonal
- Accordingly, $\hat{\Sigma} \rightarrow \Sigma = \begin{bmatrix} \hat{\Sigma} \\ 0 \end{bmatrix} \begin{matrix} \}n \\ \}m-n \end{matrix}$

  Then

$$A \;=\; U\,\Sigma\,V^T \qquad\qquad \Sigma = \text{diag},\;\; U, V = \text{orthogonal}$$



Theorem: Every matrix $A \in R^{m \times n}$ has a singular value decomposition. The singular values are unique. If A is square and σ_j are distinct, then the left and right singular vectors are unique (modulo sign).

SVD vs Eigendecomposition

- They both diagonalize a matrix A. SVD uses 2 bases (left and right singular vectors). Eigendecomposition uses 1 base (eigenvectors).

- SVD uses orthonormal vectors. Eigenvectors are not orthonormal in general.

- Not all matrices have an eigendecomposition. All matrices have a singular value decomposition.

Matrix properties of SVD

Let $A \in R^{m \times n}$, $p = \min(m,n)$, r = number of nonzero singular values of A.

Theorem: $\text{rank}(A) = r$

Pf: The rank of a diag matrix = number of nonzero diag entries. Since

$$A = U \Sigma V^T$$

U, V orth. $\Rightarrow$ $\text{rank}(A) = \text{rank}(\Sigma)$.