Limits of Spectral Clustering

Presented By: Ashish Shrivastava

University of Maryland, College Park

Oct 04, 2011

Outline

1. Introduction/Motivation
2. "Goodness" aka Consistenecy
3. Theorems
   - Theorem Statements
4. Sketch of proof for Theorem 2
   - Step 1
   - Step 2
   - Step 3
   - Step 4
   - Step 5
5. Conclusion

Outline

1. Introduction/Motivation
2. "Goodness" aka Consistenecy
3. Theorems
   - Theorem Statements
4. Sketch of proof for Theorem 2
   - Step 1
   - Step 2
   - Step 3
   - Step 4
   - Step 5
5. Conclusion

Introduction/Motivation

- How to measure the "goodness" of the clustering algorithms?

Introduction/Motivation

- How to measure the "goodness" of the clustering algorithms?
  - Unnormalized Clustering Algorithm:

$$L_n = D_n - K_n$$

where $[K_n]_{(i,j)} = s(X_i, X_j)$ and D_n is a diagonal matrix with $[D_n]_{(i,i)} = \sum_{j=1}^{n} s(X_i, X_j)$.

Introduction/Motivation

- How to measure the "goodness" of the clustering algorithms?
  - Unnormalized Clustering Algorithm:

$$L_n = D_n - K_n$$

    where $[K_n]_{(i,j)} = s(X_i, X_j)$ and D_n is a diagonal matrix with $[D_n]_{(i,i)} = \sum_{j=1}^{n} s(X_i, X_j)$.
  - Normalized Clustering Algorithm

$$L'_n = D_n^{-1/2} L_n D_n^{-1/2}$$

Outline

1. Introduction/Motivation
2. "Goodness" aka Consistenecy
3. Theorems
   - Theorem Statements
4. Sketch of proof for Theorem 2
   - Step 1
   - Step 2
   - Step 3
   - Step 4
   - Step 5
5. Conclusion

"Goodness" aka Consistenecy

- Consistenecy:

"Goodness" aka Consistenecy

- Consistenecy:
  - Well defined partitions given sufficiently many datapoints.

"Goodness" aka Consistenecy

- Consistenecy:
  - Well defined partitions given sufficiently many datapoints.
  - Focus on limit behavior for sample size tending to infitnity (Convergence).

Outline

1. Introduction/Motivation
2. "Goodness" aka Consistenecy
3. Theorems
   - Theorem Statements
4. Sketch of proof for Theorem 2
   - Step 1
   - Step 2
   - Step 3
   - Step 4
   - Step 5
5. Conclusion

Theorems

- Theorem 1: Convergence of normalized spectral clustering.

Theorems

- Theorem 1: Convergence of normalized spectral clustering.
- Theorem 2: Convergence of unnormalized spectral clustering.

Theorems

- Theorem 1: Convergence of normalized spectral clustering.
- Theorem 2: Convergence of unnormalized spectral clustering.
- Theorem 3: Structure of the limit operators.

Theorem 1:
Assumptions:

- General assumptions
- first r eigenvalues of the limit operator U' have multiplicity 1.

Claim: The first r eigenvalues of U', and the corresponding eigenvectors converge almost surely.
Notations: U' is the limit operator of L'_n and $d \in C(\mathcal{X})$ is "degree function" defined later.

Theorem 2:

Assumptions:

- General assumptions
- first r eigenvalues of the limit operator U have multiplicity 1, and are not element of $range(d)$

Claim: The first r eigenvalues of U', and the corresponding eigenvectors converge almost surely.

Notations: U is the limit operator of L_n.

Theorem 3:
Let $\mathcal{X} = \cup_{i=1}^{k} \mathcal{X}_i$ be a partition of the data space. Let $L_{ij,n}$ be the sub-matrices of L_n, $U_{ij} : \mathcal{C}(\mathcal{X}_j) \to \mathcal{C}(\mathcal{X}_i)$ the restrictions of U corresponding to the set $\mathcal{X}_i$ and $\mathcal{X}_j$, and $U'_{ij,n}$ and U'_{ij} the analogous quantities for the normalized case. Then under the general assumptions, $\frac{1}{n} L_{ij,n}$ converges compactly to U_{ij} a.s. and $L'_{ij,n}$ converges compactly to U'_{ij} a.s.

Outline

1. Introduction/Motivation
2. "Goodness" aka Consistenecy
3. Theorems
   - Theorem Statements
4. Sketch of proof for Theorem 2
   - Step 1
   - Step 2
   - Step 3
   - Step 4
   - Step 5
5. Conclusion

Construction of the operators U_n on $C(\mathcal{X})$

- Since dimension of L_n increases as n increases, an operator U_n is constructed and its convergence is showed.

$$U_n f(x) := d_n(x) f(x) - \int s(x,y) f(y) dP_n(y)$$

where, $d_n(x) := \int s(x,y) dP_n(y)$ and $f \in C(\mathcal{X})$.

Construction of the operators U_n on $C(\mathcal{X})$

- Since dimension of L_n increases as n increases, an operator U_n is constructed and its convergence is showed.

$$U_n f(x) := d_n(x) f(x) - \int s(x,y) f(y) dP_n(y)$$

where, $d_n(x) := \int s(x,y) dP_n(y)$ and $f \in C(\mathcal{X})$.

- Similarily, the limit operator is defined as below:

$$Uf(x) := d(x) f(x) - \int s(x,y) f(y) dP(y)$$

where, $d(x) := \int s(x,y) dP(y)$

Relations between $\sigma(\frac{1}{n}L_n)$ and $\sigma(U_n)$

- The spectrum of U_n consists of $range(d_n)$, plus some isolated eigenvalues with finite multiplicity. The same holds for U and $range(d)$.

Relations between $\sigma(\frac{1}{n}L_n)$ and $\sigma(U_n)$

- The spectrum of U_n consists of $range(d_n)$, plus some isolated eigenvalues with finite multiplicity. The same holds for U and $range(d)$.
- If $f \in C(\mathcal{X})$ is an eigenfunction of U_n with arbitrary eigenvalue λ, then the vector $v \in \mathbb{R}$ with $v_i = f(X_i)$ is an eigenvector of the matrix $\frac{1}{n}L_n$ with eigenvalue λ.

Relations between $\sigma(\frac{1}{n}L_n)$ and $\sigma(U_n)$

- The spectrum of U_n consists of $range(d_n)$, plus some isolated eigenvalues with finite multiplicity. The same holds for U and $range(d)$.
- If $f \in C(\mathcal{X})$ is an eigenfunction of U_n with arbitrary eigenvalue λ, then the vector $v \in \mathbb{R}$ with $v_i = f(X_i)$ is an eigenvector of the matrix $\frac{1}{n}L_n$ with eigenvalue λ.
- If v is an eigenvector of the matrix $\frac{1}{n}L_n$ with eigenvalue $\lambda \notin range(d_n)$, then the function $f(x) = \frac{1}{n}(\sum_j s(x, X_j)v_j)/(d_n(x) - \lambda)$ is the unique eigenfunction of U_n with eigen value λ satisfying $f(X_i) = v_i$.

Compact Convergence of U_n to U a.s.

- **Compact convergence:**
  A sequence of operator S_n converges compactly to S if it converges pointwise and if for every sequence $(x_n)_n$ in unit ball B, the sequence $(S - S_n)x_n$ is relatively compact.

Compact Convergence of U_n to U a.s.

- **Compact convergence:**
  A sequence of operator S_n converges compactly to S if it converges pointwise and if for every sequence $(x_n)_n$ in unit ball B, the sequence $(S - S_n)x_n$ is relatively compact.
- $$sup_{x\in\mathcal{X}}\left|\int s(x,y)dP_n(y) - \int s(x,y)dP(y)\right| \to 0 \qquad \text{almost surely}$$

Compact Convergence of U_n to U a.s.

- **Compact convergence:**
  A sequence of operator S_n converges compactly to S if it converges pointwise and if for every sequence $(x_n)_n$ in unit ball B, the sequence $(S - S_n)x_n$ is relatively compact.
- $$sup_{x\in\mathcal{X}} \left| \int s(x,y)dP_n(y) - \int s(x,y)dP(y) \right| \to 0 \qquad \text{almost surely}$$
- Remembner,
  $$U_n f(x) := d_n(x)f(x) - \int s(x,y)f(y)dP_n(y)$$

Convergence of eigenfunctions of U_n to those of U

Perturbation theory says,

- Compact convergence of operators implies the converegence of isolated eigenvalues.
- Corresponding eigen vectors converge upto a change of sign (if multiplicity is 1).
- For multiplicity larger than 1, but finite, corresponding eigenspaces converge.

Convergence of unnormalized spectral clustering

The above 4 steps imply:
if λ denotes the j-th eigenvalue of of U with eigenfunction $f \in C(\mathcal{X})$ and λ_n the j-th eigenvalue of $\frac{1}{n}L_n$ with eigenvector $v_n = (v_{n,1}, \ldots, v_{n,n})'$, then there exists a sequence of signs $a_i \in \{-1, +1\}$ such that $sup_{i=1,\ldots,n}|a_i v_{n,i} - f(X_i)| \to 0$ a.s.

Outline

1. Introduction/Motivation
2. "Goodness" aka Consistenecy
3. Theorems
   - Theorem Statements
4. Sketch of proof for Theorem 2
   - Step 1
   - Step 2
   - Step 3
   - Step 4
   - Step 5
5. Conclusion

Conclusion

- Normalized spectral clustering always converges to a limit partitions of the whole data space.
- Convergence of unnormalized spectral clustering can be guaranteed under strong additional assumption that the first eigenvalues of the Laplacian do not fall inside the range of the degree function.
- Consistenecy results are a basic sanity check for behavior of statistical learning algorithms. Algorithms that do not converge, can not be expected to exhibit reliable results on finite samples.

References

Ulrike Von Luxburg, Oliver Bousquet and Mikhail Belkin.
Limits of Spectral Clustering.
NIPS, 2004.

Ulrike Von Luxburg, Oliver Bousquet and Mikhail Belkin.
On the Convergence of Spectral Clustering on Random Samples: The Normalized Case.
COLT, 2004.