Spectral Clustering with Jensen-type kernels and their multi-point extensions

Debarghya Ghoshdastidar, Ambedkar Dukkipati, Ajay P. Adsul, Aparna S. Vijayan
Department of Computer Science & Automation, Indian Institute of Science

Focus of this work

- Jensen-Shannon (JS) divergence defines a metric on probability spaces – suitable for defining kernel.
- Jensen-Tsallis (JT) divergence is a generalization of JS-divergence based on Tsallis entropy. JT kernels on probability measures are positive definite.

Q: Can we define JT kernels on Euclidean space? Then this would generalize the linear kernel.
- JT-divergence can measure distance among multiple probability distributions.

Contributions:

- Positive definite JT kernels on unit cube in Euclidean space, and n-point extension of JT kernel.
- Multi-point spectral clustering (MSC) method for clustering with model independent n-point kernels.
- Proof of reduced complexity for MSC for n-point extension of linear kernel.

Multi-point Jensen-Tsallis kernels

JT-kernel on \([0, 1]^d \times [0, 1]^d\):

\[
k_q(x, y) = \begin{cases} \frac{1}{q-1} \sum_{j=1}^d \left( (x_j^q + y_j^q) - (x_j^q y_j^q) \right) & \text{for } q \neq 1 \\ \frac{d}{q-1} \sum_{j=1}^d (x_j^q + y_j^q) \ln(x_j^q + y_j^q) - x_j^q \ln(x_j^q) - y_j^q \ln(y_j^q) & \text{for } q = 1. \end{cases}
\]

- For \(q = 2\), we retrieve linear kernel, \(k_2(x, y) = 2x^T y\).
- \(k_q\) is positive definite on \([0, 1]^d\) for all \(q \in [0, 2]\).

Multi-point JT-kernel

For any \(x_1, \ldots, x_n \in [0, 1]^d\),

\[
K_n(x_1, \ldots, x_n) = \frac{1}{n} \sum_{j=1}^d \left( \sum_{i=1}^n x_i^j \right)^{q-1} \left( \sum_{i=1}^n x_i^j \right) - \frac{1}{n} \sum_{i=1}^n \left( x_i^j \right)^{q-1} \left( x_i^j \right) 
\]

- For \(q = 1\), we achieve \(O(N^2)\) complexity for MSC with \(n\)-point linear kernel for all \(n\) since \(V\) can be computed as below.

MSC Algorithm

**Given:** Data vectors \(\{x_1, \ldots, x_N\} \in \mathbb{R}^d\), and \(n\)-point kernel function \(K_n : \mathbb{R}^d \times \cdots \times \mathbb{R}^d \to \mathbb{R}\).

1. From \(n\)-point similarities, compute affinity matrix \(V_{ij} = \sum_{k=1}^n K_n(x_i, x_j, \ldots, x_n)\).
2. Normalize affinity matrix \(V\) as \(Z = D^{-1/2} V D^{-1/2}\), where \(D\) is diagonal with \(D_{ii} = \sum_j V_{ij}\).
3. Compute \(u_1, \ldots, u_m\), top-\(m\) unit eigenvectors of \(Z\).
4. Normalize rows of \(U = [u_1, \ldots, u_m]\) to unit length.
5. Cluster the rows of \(U\) into \(m\) clusters using \(k\)-means, and partition \(\{x_1, \ldots, x_N\}\) accordingly.

- MSC algorithm generalizes approaches in (Govindu, ‘05; Chen & Lerman, ‘09) to model independent similarities.
- MSC can be analyzed along the lines of (Ng et al., ‘02).
- Complexity of MSC algorithm is \(O(N^{m+1})\) due to computation of \(V\).
- Approximate methods for computing \(V\) do not work well for model independent \(n\)-point similarities.

Experiments on UCI data sets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Gaussian ((\tau))</th>
<th>2-point JT ((\tau))</th>
<th>3-point JT ((\tau))</th>
<th>n-linear ((\tau))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast Cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.986 (0.5)</td>
<td>0.963 (2.00)</td>
<td>0.971 (1.0)</td>
<td>0.966 (6.12)</td>
<td></td>
</tr>
<tr>
<td>Iris</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.900 (0.15)</td>
<td>0.963 (1.25)</td>
<td>0.965 (1.0-1.25)</td>
<td>0.929 (4)</td>
<td></td>
</tr>
<tr>
<td>Mammograph</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.759 (0.3)</td>
<td>0.807 (2.0)</td>
<td>0.776 (1.5)</td>
<td>0.810 (4-12)</td>
<td></td>
</tr>
<tr>
<td>Wine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.604 (5.0)</td>
<td>0.541 (1.25)</td>
<td>0.569 (0.25)</td>
<td>0.561 (4)</td>
<td></td>
</tr>
</tbody>
</table>

Image segmentation results

<table>
<thead>
<tr>
<th>Performance on Single object images (Alpert et al., ’07)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
</tr>
<tr>
<td>F-score</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>Recall</td>
</tr>
</tbody>
</table>

Segmentation of sample images

Acknowledgement: Debarghya is supported by Google India PhD Fellowship in Statistical Learning Theory