CMPSCI 791BB: Advanced ML: Laplacian Learning

Sridhar Mahadevan
Overview of lecture

- Introduction to representation theory
  - Finding a good basis – a billion $ problem!
- Introduction to spectral graph theory
  - Spectrum and eigenspace of graphs
  - Different operators and some of their properties
- Laplacian Least Squares Approximation
  - Some MATLAB demos
How to find a good basis?

Any function on this graph is a vector in $\mathbb{R}^7$

The question we want to ask is how to construct representations for all functions on this graph.

Solution 1: use the *unit basis*
Solution 2: use polynomials or RBFs

Neither of these exploit geometry
Graph Adjacency w.r.t. unit bases

Adjacency Matrix =

```
0 1 1 0 0 0 0
1 0 0 1 0 0 0
1 0 0 1 1 0 0
0 1 1 0 0 1 1
0 0 1 0 0 1 0
0 0 0 1 1 0 0
0 0 0 1 0 0 0
```
Another Representation of the Adjacency Matrix

\[ \Lambda = \begin{bmatrix}
2.5243 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1.0000 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.7923 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -0.7923 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1.0000 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -2.5243 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -2.5243
\end{bmatrix} \]

What is the underlying basis, and why is it better?
New Bases formed by diagonalizing the adjacency matrix

$$V = \begin{pmatrix} -0.3213 & 0.5000 & -0.3831 & 0.3831 & 0.5000 & 0.3213 \\ -0.3419 & 0.5000 & 0.0900 & 0.0900 & -0.5000 & -0.3419 \\ -0.4692 & -0.0000 & -0.3935 & -0.3935 & -0.0000 & -0.4692 \\ -0.5418 & 0.0000 & 0.4544 & -0.4544 & -0.0000 & 0.5418 \\ -0.3213 & -0.5000 & -0.3831 & 0.3831 & -0.5000 & 0.3213 \\ -0.3419 & -0.5000 & 0.0900 & 0.0900 & 0.5000 & -0.3419 \\ -0.2146 & -0.0000 & 0.5735 & 0.5735 & 0.0000 & -0.2146 \end{pmatrix}$$

This is also an orthonormal basis set, since $\langle V_i, V_j \rangle = 0$
Spectral Theorem

- From basic linear algebra, we know that since the adjacency matrix $A$ is symmetric, we can use the spectral theorem

\[ A = V \Lambda V^T \]

- $V$ is a matrix of orthonormal eigenvectors, $\Lambda$ is a diagonal matrix of eigenvalues

- Eigenvectors satisfy the following property:

\[ A \mathbf{x} = \lambda \mathbf{x} \]
Spectra and Eigenspace of Graphs

- There has been several decades of research on understanding the spectra of graphs.
  - If two graphs have the same spectrum, are they isomorphic?
  - Does the spectrum reveal structural properties of the graph?
- Eigenvalues can be distinct or repeated.
- Each eigenvalue is associated with an eigenvector.
- The set of eigenvectors associated with an eigenvalue form its eigenspace.
Spectra of Different Graphs

$K_5$
Spectra of Different Graphs

$C_5$
Spectra of Different Graphs

Hypercube

Graphs with different spectra are visualized, showing the distribution of eigenvalues.
Inner Product Spaces

- We will now introduce the notion of inner product spaces.
- An inner product space is a vector space associated with an inner product (e.g., $\mathbb{R}^n$).
- The set of all functions $\Phi$ on a graph $G = (V, E)$ forms an inner product space, where the inner product is defined as:
  $$<f, g> = \sum_i f(i) g(i)$$
- An operator $O$ on an inner product space of functions is a mapping $O: \Phi \rightarrow \Phi$. 
Adjacency Operator

- Let us now revisit the adjacency matrix and treat it as an operator.
- What is its effect on functions on the graph?
- It is easy to see that

\[ A f(i) = \sum_{j \sim i} f(j) \]
A Simple Theorem on Graph Spectra

- Let $G$ be a $k$-regular graph (i.e., every vertex has degree $k$).
- What can we say about its spectra and eigenspaces?
- Consider the unit vector $i = [1, 1, \ldots, 1]$.
- Is this an eigenvector of a $k$-regular graph?
- If so, what is the associated eigenvalue?

**Theorem:** For any $k$-regular graph, ____ is an eigenvector with associated eigenvalue ____
Combinatorial Graph Laplacian

Laplacian Matrix =

\[
\begin{bmatrix}
2 & -1 & -1 & 0 & 0 & 0 & 0 \\
-1 & 2 & 0 & -1 & 0 & 0 & 0 \\
-1 & 0 & 3 & -1 & -1 & 0 & 0 \\
0 & -1 & -1 & 4 & 0 & -1 & -1 \\
0 & 0 & -1 & 0 & 2 & -1 & 0 \\
0 & 0 & 0 & -1 & -1 & 2 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 1 \\
\end{bmatrix}
\]

Row sums

Negative of weights
### Laplacian Eigenvectors

#### Increasing eigenvalues

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>0.2018</th>
<th>-0.3873</th>
<th>-0.5000</th>
<th>-0.2980</th>
<th>0.5000</th>
<th>-0.2786</th>
<th>0.3780</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.2658</td>
<td>-0.0758</td>
<td>0.5000</td>
<td>0.5148</td>
<td>0.5000</td>
<td>-0.1252</td>
<td>0.3780</td>
</tr>
<tr>
<td></td>
<td>-0.4496</td>
<td>0.6332</td>
<td>0.0000</td>
<td>-0.4642</td>
<td>-0.0000</td>
<td>-0.1964</td>
<td>0.3780</td>
</tr>
<tr>
<td></td>
<td>0.7406</td>
<td>0.4963</td>
<td>0.0000</td>
<td>0.2106</td>
<td>-0.0000</td>
<td>0.1342</td>
<td>0.3780</td>
</tr>
<tr>
<td></td>
<td>0.2018</td>
<td>-0.3873</td>
<td>0.5000</td>
<td>-0.2980</td>
<td>-0.5000</td>
<td>-0.2786</td>
<td>0.3780</td>
</tr>
<tr>
<td></td>
<td>-0.2658</td>
<td>-0.0758</td>
<td>-0.5000</td>
<td>0.5148</td>
<td>-0.5000</td>
<td>-0.1252</td>
<td>0.3780</td>
</tr>
<tr>
<td></td>
<td>-0.1629</td>
<td>-0.2034</td>
<td>-0.0000</td>
<td>-0.1800</td>
<td>0.0000</td>
<td>0.8699</td>
<td>0.3780</td>
</tr>
</tbody>
</table>

#### Eigenvalue = 0
Spectral Analysis of Laplacian

Eigenfunctions

Eigenvalues
Simple Properties of the Laplacian

- The Laplacian is positive semi-definite

- The Laplacian for this graph is \([1 -1; -1 1]\)

- Note that \(x^T L x = (x_1 - x_2)^2\)

- Generalizing, we can write the Laplacian of any graph as the sum of the Laplacians of the same graph with all edges deleted, except for one.

- This implies that \(x^T L x = \sum (x_u - x_v)^2\)
Simple Properties of the Laplacian

- It is easy to see that $\lambda_1 = 0$ must be the smallest eigenvalue whose associated eigenvector is $\mathbf{1}$.
- The second eigenvalue of the Laplacian is called the Fiedler value, after the Czech mathematician who first studied the graph Laplacian (1973).
- Much of the work in spectral clustering and graph partitioning is based on Fiedler values.
- Graphs with low Fiedler values are easy to partition, whereas those with high Fiedler values have no bottlenecks (e.g., expander graphs).
Fiedler Value: Examples

Fiedler value = 4

Fiedler value = 0.58
Projection in Inner Product Spaces

Many problems in learning can be mapped into the framework of projections onto inner product spaces

\[ x = (A^T A)^{-1} b \]
## Polynomial Basis Functions

### One basis function applied to all vertices

<table>
<thead>
<tr>
<th></th>
<th>i⁰</th>
<th>i</th>
<th>i²</th>
<th>i³</th>
<th>i⁴</th>
<th>i⁵</th>
<th>i⁶</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>32</td>
<td>64</td>
<td>128</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>9</td>
<td>27</td>
<td>81</td>
<td>243</td>
<td>729</td>
<td>2187</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>1024</td>
<td>4096</td>
<td>16384</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>25</td>
<td>125</td>
<td>625</td>
<td>3125</td>
<td>15625</td>
<td>65536</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>36</td>
<td>216</td>
<td>1296</td>
<td>7776</td>
<td>46656</td>
<td>262144</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>49</td>
<td>343</td>
<td>2401</td>
<td>16807</td>
<td>117649</td>
<td></td>
</tr>
</tbody>
</table>

### All basis functions applied to one vertex

- **Polynomial Basis Functions**
Haar Basis

- Consider the set of all vectors in $\mathbb{R}^4$
- The unit basis vectors for this space is given by the vectors (shown transposed)
  - $[1,0,0,0], [0,1,0,0], [0,0,1,0], [0,0,0,1]$
- Can we design a better basis set?
  - Suppose we want to compress images, which are vectors in some high dimensional space, e.g. $\mathbb{R}^{10,000}$
- Here's the Haar wavelet basis set:
  - $[1, 1, 1, 1], [1, 1, -1, -1], [1, -1, 0, 0], [0, 0, 1, -1]$
  - These are four orthogonal vectors, so they must span the space
Compression using Haar Bases

- Represent [5, 5.2, -4, -4.1]
- Its representation in the Haar basis space is given by [0.525, 4.575, -0.1, 0.05]
- Let's reconstruct the original vector using just the largest coefficient.
- So, using just one coefficient, we are able to do a pretty good job.
Comparison of Polynomial and Laplacian Basis Representations

![Graph showing desired function and polynomial basis approximation.](image)

- **Desired Function**
  - Value vs. Number of basis functions
  - Mean squared error vs. Number of basis functions

**Graph Details:**
- **Y-axis (Value):** 0 to 8
- **X-axis (Number of basis functions):** 1 to 7
- **Legend:**
  - Blue line: Laplacian
  - Cyan line: Polynomial
Approximation on a Grid

Least-Squares Approximation using automatically learned Proto-Value Functions

Mean-Squared Error of Laplacian vs. Polynomial State Encoding

- LAPLACIAN
- POLYNOMIAL

Mean-Squared Error

Nonlinear Function Approximation

Target Function

Laplacian Least Squares Function Approximation

Polynomial Least Squares Function Approximation