DD2434 - Advanced Machine Learning - Assignment # 3

Altay Dikme

January 22, 2021

Contents

1 Spectral Graph Analysis

1.1

We let $G = (V, E)$ be an undirected d-regular graph, let A be the adjacency graph of G and let $L = I - \frac{1}{d}A$ be the normalized Laplacian of G. We want to prove that for any vector $\mathbf{x} \in \mathbb{R}^{|V|}$ the following holds true.

$$\mathbf{x^T} L\mathbf{x} = \frac{1}{d} \sum_{(u,v)\in E} (x_u - x_v)^2 \tag{1}$$

Since we know that the graph is undirected, the adjacency matrix is symmetric. Furthermore since we have a d-regular graph means that each vertex has d neighbors. We can write out:

$$\mathbf{x^T} L\mathbf{x} = \mathbf{x^T}\left(\mathbf{I} - \frac{1}{d}\mathbf{A}\right)\mathbf{x} = \mathbf{x^T I x} - \frac{1}{d}\mathbf{x^T A x} \tag{2}$$

We can write this in index-notation:

$$\mathbf{x^T I x} - \frac{1}{d}\mathbf{x^T A x} = \sum_u x_u x_u - \frac{1}{d}\sum_{u,v} x_u x_v A_{uv} \tag{3}$$

We know that $A_{uv} = 1$ if $u, v \in E$ and $A_{uv} = 0$ otherwise. Therefore we can write:

$$\sum_u x_u x_u - \frac{1}{d}\sum_{u,v} x_u x_v A_{uv} = \sum_u x_u^2 - \frac{1}{d}\sum_{u,v} x_u x_v \tag{4}$$

Since we know that for the indicies u, v where $A_{uv} = 0$ there will be no contribution to the last sum, and otherwise the contribution will be $x_u x_v$. Furthermore we can expand by using the trick of adding one half $\sum_v x_v^2$ writing:

$$\sum_u x_u^2 - \frac{1}{d}\sum_{u,v} x_u x_u = \frac{1}{2}\left[\sum_u x_u^2 + \sum_v x_v^2 - \frac{2}{d}\sum_{u,v} x_u x_u\right] = \frac{1}{2}\left[\sum_{u,v} x_u^2 + x_v^2 - \frac{2}{d} x_u x_u\right] \tag{5}$$

Since we have a d-regular graph this means as stated earlier that each vertex will have d neighbors. In our first two sums we see that every x_u^2 will show up d times. In order to not double count then we must introduce a factor $1/d$ in front of them giving us.

$$\frac{1}{2}\left[\sum_{u,v} x_u^2 + x_v^2 - \frac{2}{d} x_u x_u\right] = \frac{1}{2}\left[\frac{1}{d}\sum_{u,v} x_u^2 + x_v^2 - 2x_u x_u\right] = \frac{1}{2d}\sum_{u,v}(x_u - x_v)^2 \tag{6}$$

We see that this will count over all combinations of vertices, and thus we will have both u, v and v, u pairs which are the same thing. Thus we are double counting once again. If we change the notation then such that we go through each node once we have:

$$\frac{1}{2d}\sum_{u,v}(x_u - x_v)^2 = \frac{1}{d}\sum_{(u,v)\in E}(x_u - x_v)^2 \tag{7}$$

Which is what we were looking for, i.e that:

$$\mathbf{x^T} L\mathbf{x} = \frac{1}{d} \sum_{(u,v)\in E} (x_u - x_v)^2 \tag{8}$$

$\square$

1.2

Now we want to show that the normalized Laplacian given earlier is a positive semidefinite matrix. This follows readily from the previous subquestion. We have shown that

$$\mathbf{x^T}L\mathbf{x} = \frac{1}{d}\sum_{(u,v)\in E}(x_u - x_v)^2$$

Remember that if we have $\mathbf{x^T}L\mathbf{x} \geq 0$ then L is positive semidefinite. We can see that Eq. 8 is never negative and only zero when u, v are the same, i.e we are at the same node. Thus the normalized Laplacian is a positive semidefinite matrix. □

1.3

We assume that we find a non-trivial vector $\mathbf{x}_*$ that minimizes the expression $\mathbf{x}^T L\mathbf{x}$. We want to begin by explaining what non-trivial means, and then explain how $\mathbf{x}_*$ can be used as an embedding of the vertices of the graph into the real line.

Basically when we minimize the expression we have

$$\min_{\mathbf{x}} \mathbf{x^T}L\mathbf{x} \tag{9}$$

with the constraint that $\mathbf{x^T}\mathbf{x} = \mathbf{1}$. Using this we can form a Lagrangian giving us:

$$\min_{\mathbf{x}} \mathbf{x^T}L\mathbf{x} - \lambda\mathbf{x^T}\mathbf{x} \tag{10}$$

Deriving w.r.t $\mathbf{x}$ and setting equal to zero gives us.

$$(L - \lambda)\mathbf{x} = 0 \rightarrow L\mathbf{x} = \lambda\mathbf{x} \tag{11}$$

Which is just an eigenvalue problem. Therefore the non-trivial vector $\mathbf{x}_*$ that minimizes the expression is an eigenvector of the Laplacian L. The question is which eigenvector. Since we have a d-regular graph we know that the smallest eigenvalue and corresponding eigenvector will be $\lambda = 0$ and $\mathbf{x} = (1, \ldots, 1)$, which is a constant vector and not interesting. Thus this is a trivial case. We are looking for a *non-trivial* case and therefore look at the next smallest eigenvalue λ_2 with eigenvector which must by definition be orthogonal to the constant eigenvector $\mathbf{x}$ described earlier.

Thus the non-trivial vector $\mathbf{x}_*$ that minimizes the expression is the eigenvector corresponding to the *second* smallest eigenvalue λ_2 (Assuming it is non-zero. No disconnected graphs allowed in other words), i.e the Fiedler vector.

Next we want to explain how $\mathbf{x}_*$ can be used as an embedding of the vertices of the graph into the real line. Each element of the vector $\mathbf{x}_*$ simply corresponds to the coordinate x_u that corresponds to the vertex $U \in V$. Then what Eq. 8 tells us is that we are looking at the sum of the squared differences of the values that the vector $\mathbf{x}_*$ assigns to each vertex in our graph.

The reason why $\mathbf{x}_*$ provides a meaninful embedding can be explained by first remembering the constraint $\mathbf{x^T}\mathbf{x} = \mathbf{1}$ meaning that the sum of the squared components of $\mathbf{x}_*$ equals one. However $\mathbf{x}_*$ must also be orthogonal to the eigenvector corresponding to $\lambda_1 = 0$ ($\mathbf{x_1} = (1, \ldots, 1)$). From this we see that $\mathbf{x}_*^{\mathbf{T}}\mathbf{x_1} = \sum_i x_{*i} = 0$. Thus our vector $\mathbf{x}_*$ needs to have some sort of balance of having values on the negative side of the real line and some on the positive side so that the sum of the elements cancel out.

The minimization problem states that we want to minimize the sum of the squared differences of labels across all the edges (See Eq. 8 and 10). From this we see that we have small differences if the sign of both vertices are the same (negative and negative or positive and positive), and we have larger differences if we have a positive vertex and a negative vertex. Using this we can see that the nodes which are on the same side of the real line are closer to each other in the graph and further away from nodes on the other side of the real line. Thus the vector $\mathbf{x}_*$ can be used for partitioning the graph and is a meaningful embedding.