Marcus Rose
Math 123 HW 6
Due 3/14/23

---

1. Let $L = D - W$ be the unnormalized graph Laplacian associated to a graph $G = (V, W)$ on points $\{x\}_i^n$ with symmetric weight matrix W and diagonal degree matrix D. Let $\{C\bar{C}\}$ be any partition of $\{x\}_i^n$, and let

$$f_c = \begin{cases} \sqrt{\frac{\text{vol}(\bar{C})}{\text{vol}(C)}} & x \in C \\ \sqrt{\frac{\text{vol}(C)}{\text{vol}(\bar{C})}} & x \in \bar{C} \end{cases}.$$

(a) Prove $\langle Df^c, \mathbb{1} \rangle = 0$.

Proof

$$\langle Df^c, \mathbb{1} \rangle = \sum_{x_i \in C} (Df^c) + \sum_{x_i \in \bar{C}} (Df^c) = \sum_{x_i \in C, x_j \in V} W_{ij} \sqrt{\frac{\text{vol}(\bar{C})}{\text{vol}(C)}} - \sum_{x_i \in \bar{C}, x_j \in V} W_{ij} \sqrt{\frac{\text{vol}(C)}{\text{vol}(\bar{C})}}.$$

Note that these are linear terms, so the f^c cam be removed and the sum further simplifies,

$$\sqrt{\frac{\text{vol}(\bar{C})}{\text{vol}(C)}} \sum_{x_i \in C, x_j \in V} W_{ij} - \sqrt{\frac{\text{vol}(C)}{\text{vol}(\bar{C})}} \sum_{x_i \in \bar{C}, x_j \in V} W_{ij} = \sqrt{\frac{\text{vol}(\bar{C})}{\text{vol}(C)}} \text{vol}(C) - \sqrt{\frac{\text{vol}(C)}{\text{vol}(\bar{C})}} \text{vol}(\bar{C}) = 0$$

□

(b) Prove $(f^c)^T D f^C = \text{vol}(V)$.

Proof

$$(f^c)^T D f^c = \sqrt{\frac{\text{vol}(\bar{C})}{\text{vol}(C)}} \sqrt{\frac{\text{vol}C}{\text{vol}(\bar{C})}} \text{vol}(C) - \sqrt{\frac{\text{vol}C}{\text{vol}(\bar{C})}} \text{vol}(C) \sqrt{\frac{\text{vol}(\bar{C})}{\text{vol}(C)}} \text{vol}(\bar{C})$$

Note that the above simplifies nicely to

$$\text{vol}(\bar{C}) + \text{vol}(C) = \text{vol}(V) = \text{Total Volume}.$$

□

(c) Prove that $(f^C)^T L f^C = \text{vol}(V)\text{Ncut}(C, \bar{C})$.

Proof

$$(f^c)^T L f^c = (f^c)^T (D - W) f^c = (f^c)^T D f^c - (f^c)^T W f^c = \sum_i^n d_i f_i^{c2} - \sum_i^n f_i^c f_j^c W_{ij}.$$

Note that if we were to expand this binomial, this becomes

$$\frac{\sum_i d_i f_i^{c^2} - 2 \sum_{i,j} f_i^c f_j^c W_{ij} + \sum_j d_j f_j^{c^2}}{2} = \frac{\sum W_{ij} (f_i^c - f_j^c)^2}{2}.$$

Plugging in f^c as given, we arrive at the Ncut function such that the above is equal to

$$\text{Ncut}(C, \bar{C})(\frac{\text{vol}(\bar{C})}{\text{vol}(C)} + \frac{\text{vol}(C)}{\text{vol}(\bar{C})} + 2) = \text{Ncut}(C, \bar{C})\text{vol}(V).$$

□

2. Recall that one construction of the weight matrix for a graph on data $\{x_i\}_i^n$ is to use the Gaussian kernel $W_{ij} = \exp(-\|x_i - x_j\|_2^2/\sigma^2), i \neq j$ and $W_{ij} = 0, i = j$ j for some choice of $\sigma > 0$.

   Note: This question can be answered through the lens of eigenvectors, as the eigenvalues of the graph Laplacian give us insight into the optimal number of clusters.

   (a) What happens to the resulting Laplacian matrix L as $\sigma \to 0^+$?

   Answer As $\sigma \to 0^+$, note that all of the values in the Laplacian L also tend to 0. Basically, you have a 0 eigenvalue for corresponding nonzero eigenvector.

   (b) What happens to the resulting Laplacian matrix L as $\sigma \to \infty$?

   Answer As $\sigma \to \infty$, the graph Laplacian L approaches a matrix filled with negative ones except with $n - 1$ along the diagonal.

   In other words, $L\mathbb{1} = 0$. Here, you do have eigenvalue 0, but $Lx = \lambda x - [\sum x_i ... \sum x_i]^T$ implies that there exists eigenvalue n that corresponds to the vector that is this difference. In this sense, the eigenvalues of the graph Laplacian can help us find the optimal nubmers of clusters k.

3. Using MATLAB,

   (a) Run spectral clustering on this data, using a sparse Laplacian with different numbers of nearest neighbors and $k = 6$ clusters. How do the results compare to the ground truth data?

   Running spectral clustering on the data, we can see that there is some difference between the ground truth and the clusters. There are tradeoffs between more clusters and the accuracy of the data with the ground truth. If k gets too large, the model becomes too generalized and doesn't really learn the data in both train and test sets (aka underfitting). The opposite is overfitting for k too small.

```
% Compute the adjacency matrix using k-nearest neighbors
W = zeros(size(D));
for i = 1:size(D, 1)
    [~, idx] = sort(D(i,:));
    W(i,idx(1:k+1)) = 1;
    W(idx(1:k+1),i) = 1;
end

% Construct the sparse Laplacian matrix
D = diag(sum(W));
L = D - W;

% Compute the eigenvectors of the Laplacian
```

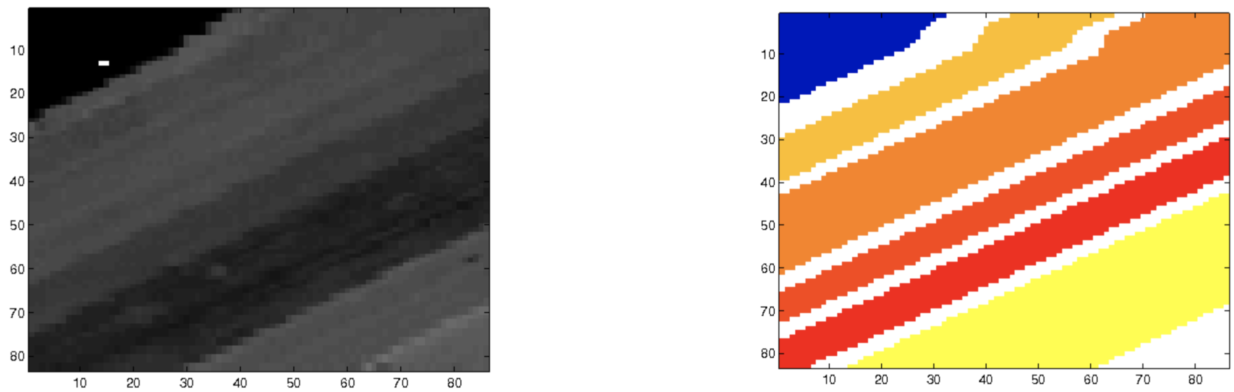
```
[V, ~] = eigs(L, 7, 'sm');

% Normalize eigenvectors
V = bsxfun(@rdivide, V, sqrt(sum(V.^2, 2)));

% Cluster using k-means
idx = kmeans(V(:,2:7), 6);

% Visualize
scatter(data(:,1), data(:,2), 10, idx, 'filled');
```

Listing 1: Code for the problem



(a) A single band of the salinasA.mat data

(b) The groundtruth

Figure 1: Comparing ground truth to a band of the data

(b) Plot the nearest 10 eigenvalues of the data for different choices of σ. What does the eigengap estimate as the number of clusters for these choices of σ?

The first nonzero eigenvalue is the spectral gap. The spectral gap provides an idea regarding the density of the graph. If this graph was densely connected (few disconnected nodes), then the spectral, eigen gap would increase. **Here, we see as the denseness of the Laplacian increases, the eigengap decreases.**

(c) Compare the projections onto the first three principle components with the first three Laplacian eigenvectors by plotting both sets in different figures using 'scatter3'. How do the representations differ qualitatively?

Qualitatively, there are similarities between the datasets, however, projecting onto the first 3 principal components does not provide enough information to give a good understanding of the data. Hence, this could be an example of overfitting, as the data has been dimensionally reduced too far for a fundamental understanding of the data to be obtained.