Marcus Rose
Math 123 HW 5
Due 3/7/23

---

1. Let $\{x_i\}_{i=1}^n \subset \mathbb{R}^D$ be a discrete set on unique points. Recall that the DBSCAN algorithm depends on two parameters: ϵ and MinPts.

   (a) Describe the behavior of DBSCAN as $\epsilon \to +\infty$ and as $\epsilon \to +0$.

   Answer As $\epsilon \to \infty$, there will be a larger number of core points since the neighborhood size will be larger and it will hence cover all n points in one cluster, and DBSCAN would consider them core points and not border or noise points, which would result in a large degree of error. As $\epsilon \to 0$, the algorithm will become more sensitive to distance between points. In order to create clusters of the points, they will need to be closer to each other in order for the points to be considered clusters. They would only contain itself in the cluster.

   (b) Describe the behavior of DBSCAN as MinPts $\to +0$ and MinPts $\to 0+$.

   Answer As MinPts $\to \infty$, it is harder for points to be called core points and that no neighborhood is meaningful which can result in smaller clusters being merged to create larger clusters. All points described are points that are not defined neither as core nor border points.

   As MinPts $\to 0$, it is easier for the data to be core points, so the noise points would be included in the clusters. That is, there would be no border nor noise points.

2. Let $L = D - W \in \mathbb{R}^{n \times n}$ be the graph Laplacian for data with associated symmetric weight matrix W with $W_{ij} \in [0, 1]$ for all $i, j = 1, ..., n$.

   (a) Show L is positive semi-definite.

   Proof Recall the definition of the Laplacian L:

   $$L = D - W,$$

   where

   $$D_{ij} = \begin{cases} d_i & i = j \\ 0 & \text{otherwise} \end{cases}$$

   is the degree matrix and

   $$W_{ij} = \exp \frac{-\|x_i - -x_j\|^2}{\sigma^2}$$

   for some under specified parameter.

   Then for any vector $\mathbf{x}$,

   $$\mathbf{x}^T L \mathbf{x} = \mathbf{x}^T B B^T \mathbf{x} = (B^T \mathbf{x})^T (B^T \mathbf{x}) = \|B^T \mathbf{x}\|^2 \geq 0,$$

   where B is defined as the incidence matrix[1] defined as

[1] https://en.wikipedia.org/wiki/Incidence_matrix

$$B_{ij} = \begin{cases} -1 & \text{if edge } e_j \text{ leaves vertex } v_i \\ 1 & \text{if edge } e_j \text{ enters vertex } v_i \\ 0 & \text{otherwise} \end{cases}.$$

Because $\mathbf{x}^T L\mathbf{x} \geq 0$ for any $\mathbf{x}, L$ is positive semi-definite.

Note: This proof makes use of the oriented incidence matrix, which is a topic not directly discussed.

□

(b) Show L is not positive definite by proving 0 is an eigenvalue of L.

Proof We have just showed that L is positive semi-definite. Assume, by contradiction, that L is positive definite and that there exists an eigenvalue 0 of L. Because L is symmetric, there exists an orthonormal basis $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n \in \mathbb{R}^n$ with corresponding eigenvectors of L. That is, $\mathbf{x}_i\mathbf{x}_j^T = 0$ if $i \neq j$ and $\mathbf{x}_i\mathbf{x}_j^T = 1$ if $i = j$. Let $\lambda_{1,2}, ..., \lambda_n$ be corresponding eigenvalues of L such that

$$L\mathbf{x}_i = \lambda_i\mathbf{x}_i.$$

Because L is positive definite by assumption, $\mathbf{x}^T L\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$ and

$$\mathbf{x}_i L\mathbf{x}_i^T = \mathbf{x}_i(\lambda_i\mathbf{x}_i) =_i (\mathbf{x}_i^T\mathbf{x}_i) = \lambda_i > 0.$$

But if L is positive definite, then every λ_i must be strictly greater than 0. That is, there cannot exist a zero eigenvalue and its existence here falsifies the claim that L is positive definite.

□

3. Compute the graph Laplacian with $W_{ij} = \exp(-\|x_i - x_j\|_2^2/\sigma^2)$ on the image in Ncut_Data.mat using a range of σ. For each of these σ, use the second eigenvector (i.e. the eigenvector with second smallest eigenvalue) to segment the image by thresholding at 0. Discuss the results. Do they make sense? How do the results depend on σ?

Answer First note the original image with scaled colors:

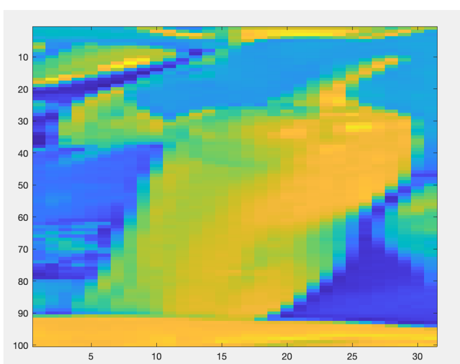
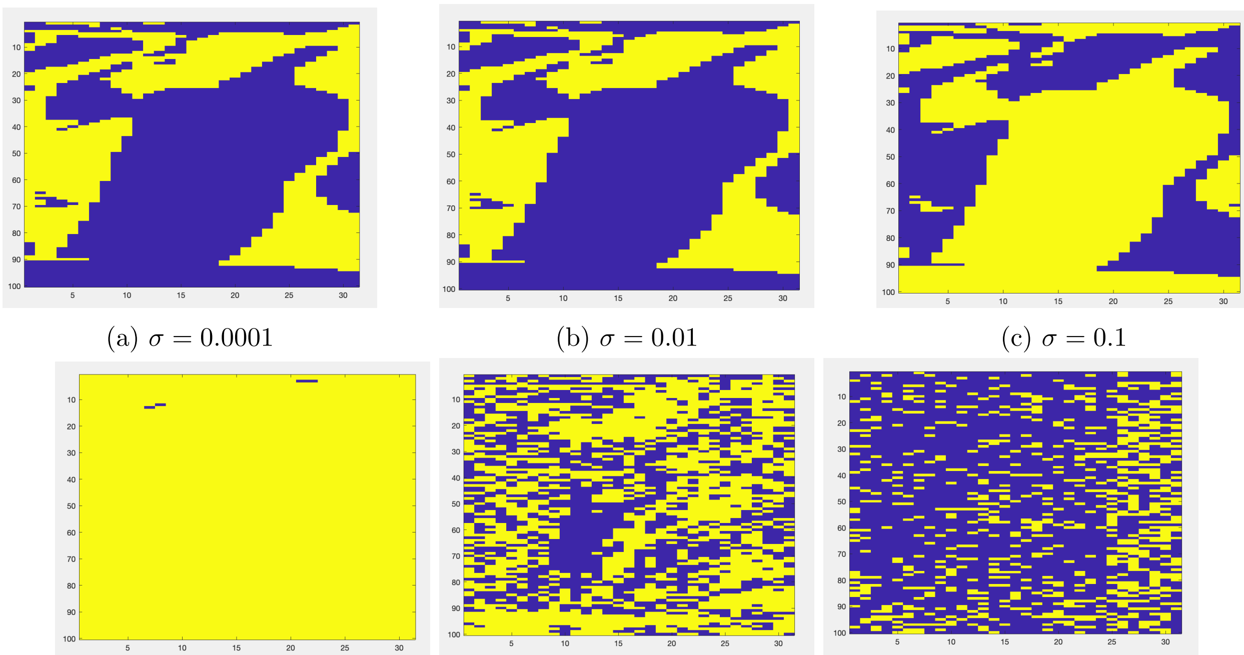


Figure 1: Original image, scaled colors

Using a range of σ over the interval $[0.0001, 1000]$, the following was the output for each sigma value.



(a) $\sigma = 0.0001$ (b) $\sigma = 0.01$ (c) $\sigma = 0.1$

(d) $\sigma = 1$ (e) $\sigma = 100$ (f) $\sigma = 1000$

Figure 2: The outputs for each σ value for the Ncut image

Larger values of sigma resulted in a more scattered result, but it would appear that the smaller σ values present something more akin to the original image. We can clearly see the main figure in the lighter colors that is shown in Figure 1. Strangely, this structure was preserved up until $\sigma = 1$. This made a little sense, as a $\sigma = 1$ value complete removes the denominator from the graph Laplacian.

```
sigma = [ 0.0001, 0.001, 0.01,1, 10, 10000, 10000];

% Populate the graph Laplacian with the specified values
for s = sigma
    for i = 1:size(X,1)
        for j = 1:size(X,1)
            W(i,j)=exp(-(norm(X(i,:)-X(j,:))^2/s^2));
        end
    end

    % Set diagonal of W to 0
    W(logical(eye(size(X,1)))) = 0;

    % Diagonal matrix
    D = diag(sum(W, 2));

    % Laplacian
    L = D - W;

    % Eigendecomp
    [eigenvectors, eigenvalues] = eig(L);
    eigenvalues = diag(eigenvalues);

```

```
    % Find second eigenvector
    [~, index] = sort(eigenvalues, 'ascend');
    eigenvalues=eigenvalues(index);
    eigenvectors=eigenvectors(:,index);
    disp(['First 5 eigenvalues (sigma = ' num2str(s) ')']);
    disp(eigenvalues(1:5));

    % Segment the image by thresholding t at 0
    t = 0;
    eigenvector_prime = double(eigenvectors(:,2) < t);

    figure;
    imagesc(reshape(eigenvector_prime, [100,31]));
    colormap();

end
```

Listing 1: Code for the problem