**1.**

**Research about the Spectral Clustering method, and answer the following questions:**

**a. In which cases might it be more useful to apply?**

- Spectral clustering is particularly useful in situations where the groupings are not linearly separable or have irregular shapes :

some cases are:

**Non-spherical clusters:** Unlike algorithms like k-means, which tend to find spherical clusters, spectral clustering can identify clusters with non-convex or irregular shapes.

**Clusters with varying densities:** Spectral clustering can detect clusters that have different point densities, which can be challenging for distance-based methods like k-means.

**Non-linear dimensionality reduction:** Spectral clustering can also be used for non-linear dimensionality reduction, as the eigenvectors of the Laplacian matrix represent a projection of the data into a lower-dimensional space.

**Image segmentation:** Spectral clustering can be used to segment images based on the similarity of pixels, taking into account both color intensity and spatial location of the pixels. This can help identify regions of interest in images with complex or irregular shapes.

**Text analysis and natural language processing (NLP):** Spectral clustering can be applied in text analysis to group documents or words based on their semantic similarity. This can be useful for identifying common themes, organizing large collections of documents, or improving search and recommendation systems.

**b. What are the mathematical fundamentals of it?**

The mathematical foundations of spectral clustering are based on graph theory, linear algebra, and spectral analysis.

Graph: In the context of spectral clustering, each node corresponds to a data point, and the edges represent the similarity between the data points.

Similarity measure: To construct the graph, a similarity (or distance) measure between data points is needed. There are various similarity measures, such as Euclidean distance, Manhattan distance, or Gaussian kernel. The choice of similarity measure will depend on the specific problem and the nature of the data.

Affinity matrix: The affinity matrix is a square matrix that stores the similarities between all pairs of data points. The elements of the affinity matrix (A) represent the similarity between data points i and j, with A(i, j) = 0 if the points are not connected in the graph.

Laplacian matrix: The Laplacian matrix is derived from the affinity matrix and is a central concept in spectral graph analysis. There are several ways to compute the Laplacian matrix, but the normalized Laplacian is the most common. It is calculated as L = D^(-1/2) \* A \* D^(-1/2), where D is a diagonal matrix with elements D(i, i) being the sum of row i in the affinity matrix.

Eigenvectors and eigenvalues: Eigenvectors and eigenvalues are fundamental concepts in linear algebra and are used in spectral analysis. In the context of spectral clustering, the eigenvectors and eigenvalues of the Laplacian matrix are calculated. The eigenvectors associated with the k smallest eigenvalues are used to transform the data into a lower-dimensional space.

Projection into a lower-dimensional space: The k selected eigenvectors are used to form a feature matrix (Y) that projects the data into a lower-dimensional space. This allows identifying the structure of the data and simplifying the clustering problem.

Clustering in the transformed space: Finally, a clustering algorithm, such as k-means, is applied in the transformed space using the feature matrix Y. This allows identifying clusters in the original data based on the structure revealed by the spectral analysis.

**What is the algorithm to compute it?**

**Construct the similarity graph:** Calculate the similarity between data points using a chosen similarity measure. Commonly used measures are the Gaussian (RBF) kernel or the k-nearest neighbors (k-NN) approach. Then, construct a graph with data points as nodes and edges weighted by their similarity.

**Compute the affinity matrix (A):** Create an affinity matrix, a square matrix whose elements A(i, j) represent the similarity between data points i and j, derived from the similarity graph.

Calculate the degree matrix (D): Compute the degree matrix, a diagonal matrix where each element D(i, i) is the sum of the ith row in the affinity matrix.

**Compute the Laplacian matrix (L):** Calculate the Laplacian matrix, which can be computed using different methods, such as the unnormalized Laplacian (L = D - A) or the normalized Laplacian (L = D^(-1/2) \* A \* D^(-1/2)).

**Calculate eigenvectors and eigenvalues:** Compute the eigenvectors and eigenvalues of the Laplacian matrix (L). Sort the eigenvalues in ascending order, and select the k eigenvectors corresponding to the k smallest eigenvalues.

**Form the feature matrix (Y):** Create the feature matrix (Y) by stacking the selected k eigenvectors column-wise. Each row in this matrix now represents a data point in the lower-dimensional space.

Cluster in the lower-dimensional space: Apply a clustering algorithm, such as k-means, to the feature matrix (Y) to partition the data points into k clusters.

**Assign original data points to clusters:** Finally, assign the original data points to the clusters formed in the lower-dimensional space. Each data point is assigned to the cluster corresponding to the transformed point in the feature matrix (Y).

Principio del formulario

**d. Does it hold any relation to some of the concepts previously mentioned in class? Which, and how?**

the similarities are as follows: - Spectral clustering uses measures of similarity such as k-means and k-medoids to identify how similar or distant the data is. (euclidean, manhatha...) the use of eigenvalues and eigenvectors to transform data into a lower-dimensional space.