Homework 2 **ISYE 6740**

Shrikanth Mahale (smahale6)

1 Spectral clustering [50 points]

Question 1.1

1. (20 points) Consider an undirected graph with non-negative edge weights wij and graph Laplacian L. Suppose there are m connected components A1,A2,...,Am in the graph. Show that there are m eigenvectors of L corresponding to eigenvalue zero, and the indicator vectors of these components IA1,...,IAm span the zero eigenspace.

Solution 1.1

For graph *G* that is given, its adjacency matrix A contains an entry at A_{ij} if vertices *i* and *j* have an edge between them. The degree matrix *D* contains the degree of each vertex along its diagonal.

The graph Laplacian *L* of *G* is given by $D - A$.

For every vector $\vartheta \in \mathbb{R}^n$ we have:

$$f'L\vartheta = f'D\vartheta - f'A\vartheta = \sum_{i=1}^{n} d_i\vartheta_i^2 - \sum_{i,j=1}^{n} \vartheta_i\vartheta_j w_{ij}$$

$$= \frac{1}{2}\left(\sum_{i=1}^{n} d_i\vartheta_i^2 - 2\sum_{i,j=1}^{n} \vartheta_i\vartheta_j w_{ij} + \sum_{j=1}^{n} d_j\vartheta_j^2\right) = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}(\vartheta_i - \vartheta_j)^2$$

The symmetry of *L* follows directly from the symmetry of *A* and *D*. The positive semi-definiteness is a direct consequence of above derivation, which shows that $f'L\vartheta \geq 0$ for all $\vartheta \in \mathbb{R}^n$

Set of eigenvalues $(\lambda_0, \lambda_1, \lambda_2, \ldots\ldots \lambda_{m-1})$ can be defined, such that the vector (1,1,1 ..... ,1) satisfies $\vartheta_0 = 0$. Such that eigenvalue of the Laplacian matrix is $\lambda_0 = 0$.

A subgraph could be a connected component of an undirected graph such that there are no edges between vertices of the subgraph and vertices of the rest of the graph.

Now let's consider the case of k connected components. We assume that without the loss of generality the vertices are ordered according to the connected components they belong to. In this case, the adjacency matrix A has a block diagonal form, and the same is true for the matrix L:

$$L = \begin{pmatrix} L_1 & & & \\ & L_2 & & \\ & & \ddots & \\ & & & L_k \end{pmatrix}$$

Each of the blocks L_i is a proper graph Laplacian on its own, namely the Laplacian corresponding to the subgraph of the i-th connected component. If there are no edges between these connected components, all values exterior the component will be 0, so the matrix is block diagonal. Each block matrix within the Laplacian has eigenvalue 0.

Question 1.2

2. (30 points) Real data: political blogs dataset. We will study a political blog dataset first compiled for the paper Lada A. Adamic and Natalie Glance, "The political blogosphere and the 2004 US Election", in Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem (2005). The dataset nodes.txt contains a graph with n = 1490 vertices ("nodes") corresponding to political blogs. Each vertex has a 0-1 label (in the 3rd column) corresponding to the political orientation of that blog. We will consider this as the true label and try to reconstruct the true label from the graph using the spectral clustering on the graph. The dataset edges.txt contains edges between the vertices. Here we assume the number of clusters to be estimated is k = 2. Using spectral clustering to find the 2 clusters. Compare the clustering results with the true labels. What is the false classification rate (the percentage of nodes that are classified incorrectly)?

Solution 1.2

Code spectral_clustering.py attached separately

First, checking the true label, 2 data points were found with quotes missing, they were added manually. Spectral Clustering was performed with k=2, and a false classification rate of 0.508 was found.

---

2 PCA: Food consumption in European area [50 points]

Question 2.1

The data food-consumption.csv contains 16 countries in the European area and their consumption for 20 food items, such as tea, jam, coffee, yoghurt, and others. There are some missing data entries: you may remove the rows "Sweden", "Finland", and "Spain". The goal is to perform PCA analysis on the data, i.e., find a way to perform linear combinations of features across all 20 food-item consumptions, for each country. If we extract two principal components, that means we use two singular vectors that correspond to the largest singular values of the data matrix, in combining features. You will need to implement PCA by writing your own code.

1. (15 points) Write down the set-up of PCA for this setting. Explain how the data matrix is set-up in this case (e.g., each dimension of the matrix corresponds to what.) Explain in words how PCA is performed in this setting.

Solution 2.1

Principal component analysis (PCA) is a procedure of statistic that uses an orthogonal transformation to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of linearly uncorrelated variables called principal components. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors (each being a linear combination of the variables and containing n observations) are an uncorrelated orthogonal basis set. PCA is sensitive to the relative scaling of the original variables.

The dataset "food-consumption.csv" has 16 rows of data for each country with pivoted for 21 food items like Tea, Apples, Oranges etc.

- For performing PCA, data needs to be imported to pandas data frame and then converted into a 16 x 20 NumPy matrix. Since countries Spain, Finland and Sweden have null values for certain food items, we will remove them from analysis as instructed. This will reduce the number of rows by 3 and remove the label for country. Finally, the matrix is of size 13 x 20
- We then define a vector with a mean value of each row of size 1 x 20.

- Before running the PCA on data, it needs to be scaled which standardizes all variables with same weight for PCA. This is done by subtracting each data point by its mean and then dividing by standard deviation.
- A covariance matrix of the standardized and scaled matrix is defined. a square matrix giving the covariance between each pair of elements of a given random vector. In the matrix diagonal there are variances, i.e., the covariance of each element with itself. Size 20 x 20
- For each covariance matrix, eigen values and eigen vectors is computed.
- Once computed, the eigen values and eigen vectors are sorted.
- The first 2 principal components are extracted. We create a matrix (W) with these two eigen vectors with shape 20 x 2 since there are two principal components used in this analysis.
- New data frame is created after country labels are assigned to the data.
- By using k-means as a metric to find any patterns, a scatter plot is created of the principal components. Then a second scatter plot is created with the principal components and a sub plot is added to create a loading chart with the weight/usage of the food items.

Question 2.2

The data food-consumption.csv contains 16 countries in the European area and their consumption for 20 food items, such as tea, jam, coffee, yoghurt, and others. There are some missing data entries: you may remove the rows "Sweden", "Finland", and "Spain". The goal is to perform PCA analysis on the data, i.e., find a way to perform linear combinations of features across all 20 food-item consumptions, for each country. If we extract two principal components, that means we use two singular vectors that correspond to the largest singular values of the data matrix, in combining features. You will need to implement PCA by writing your own code.

2. (15 points) Suppose we aim to find top k principal components. Write down the mathematical optimization problem involved for solving this problem. Explain the procedure to find the top k principal components in performing PCA.

Solution 2.2

The aim of PCA is dimensionality reduction of data by combining features that are correlated while retaining most of the information. In order to achieve that we need to find a direction to project the data such that it maximizes the variation as that is an important criterion to retain useful information.

→ Given m data points $\{x^1, x^2, x^3 \dots x^m\} \in R^n$
the mean being $u = \frac{1}{m} \sum_{i=1}^{m} x^i$

→ The normalized direction $w \in R^n$ where $\|w\| \leq 1$

→ Maximizing the variance of data along w.

$$\max_{w: \|w\| \leq 1} \frac{1}{m} \sum_{i=1}^{m} (w^T x^i - w^T u)^2$$

$$= \frac{1}{m} \sum_{i=1}^{m} (w^T x^i - w^T u)^2$$

$$= \frac{1}{m} \sum_{i=1}^{m} (w^T (x^i - u))^2$$

$$= w^T \left(\frac{1}{m} \sum_{i=1}^{m} (x^i - u)(x^i - u)^T \right) w$$

↑
Co-variance matrix. Taking x excluding w since it is not dependent on i

$$= \max_{w: \|w\| \leq 1} w^T C w$$

Lets suppose we have data with 2 dimensions.
so the covariance matrix C will be a 2x2 diagonal matrix

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

From the optimization problem $\max_{w: \|w\| \leq 1} w^T C w$

$$= \max_{w: \|w\| \leq 1} [w_1 \; w_2] \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = w_1^2 + 2w_2^2$$

$$= \max_{w: \|w\| \leq 1} w_1^2 + 2w_2^2$$

The quadratic closed form and can be easily solved by using the eigen vectors of C

Provided $C \in R^n$, find a vector $w \in R^n$ and $\|w\| = 1$ such that $Cw = \lambda w$

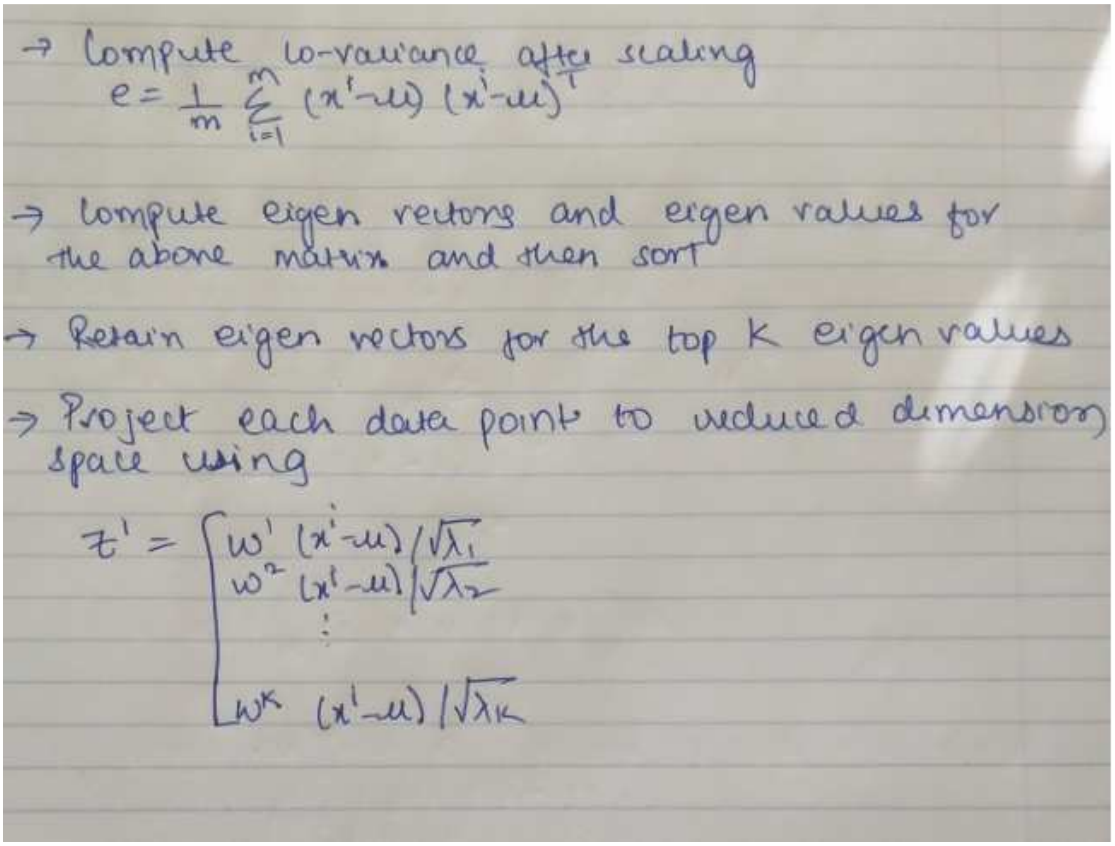
There can be multiple solutions of $w^1, w^2 \dots w^n$ with same or different $\lambda^1, \lambda^2 \dots \lambda^n$ where w^n are eigen vectors and λ^n are eigen values.

Eigen vectors are orthogonal such that
$w^{iT} w^i = 1$ and $w^T w^j = 0$
The variance of eigen vectors
$Cw = \lambda w$
Since variance is $w^T Cw = \lambda w^T w = \lambda$
Since $w^T w = 1$ or $\|w\| = 1$
λ - eigen value

So directions $w^1, w^2 \dots$ which have the largest variances
$\lambda^1, \lambda^2, \lambda^3 \dots$
λ^1 is the largest eigen value, λ^2 is the 2nd largest and so on

Finding the top K principal components
m data points $\{x^1, x^2 \dots x^m\} \in R^n$
Mean of data $u = \frac{1}{m} \sum_{i=1}^{m} x^i$

Standardize the data
$X = \sum_{i=1}^{m} \frac{x^i - u}{s_0^i}$



Question 2.3

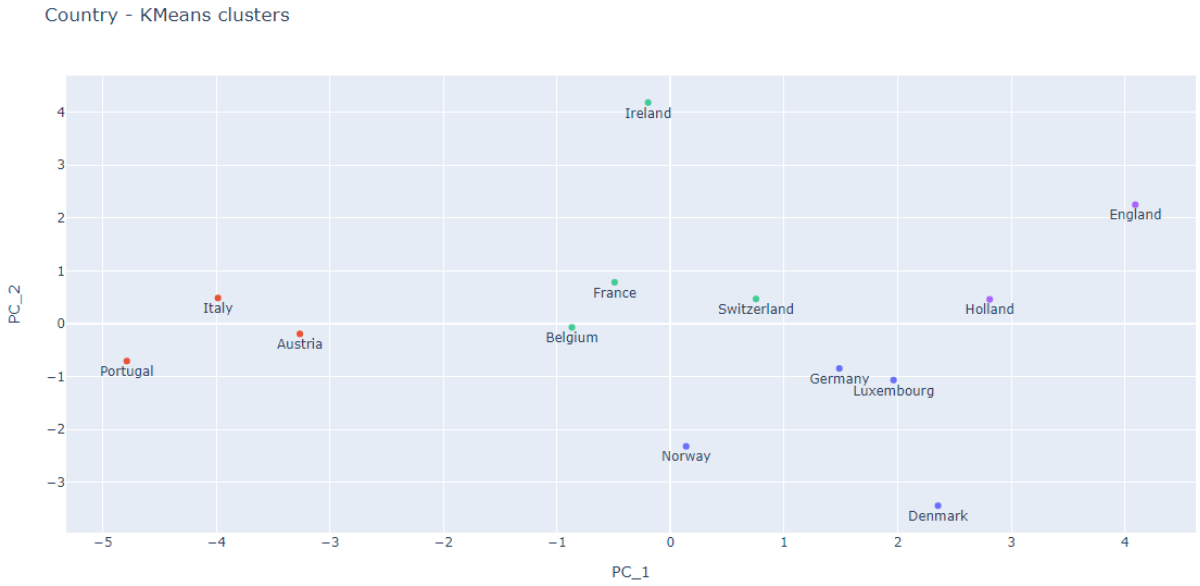
The data food-consumption.csv contains 16 countries in the European area and their consumption for 20 food items, such as tea, jam, coffee, yoghurt, and others. There are some missing data entries: you may remove the rows "Sweden", "Finland", and "Spain". The goal is to perform PCA analysis on the data, i.e., find a way to perform linear combinations of features across all 20 food-item consumptions, for each country. If we extract two principal components, that means we use two singular vectors that correspond to the largest singular values of the data matrix, in combining features. You will need to implement PCA by writing your own code.

3. (10 points) Find the top two principal component vectors for the dataset and plot them (plot a value of the vector as a one-dimensional function). Describe do you see any pattern.

Solution 2.3

We can see 4 distinct groups after plotting a scatter on the k-means labels for patterns:
Group 1: Austria, Portugal and Italy
Group 2: Switzerland, Ireland, France and Belgium,
Group 2: Norway, Luxembourg, Germany and Denmark
Group 4: Holland and England

Each of these 4 groups and similar food items consumption



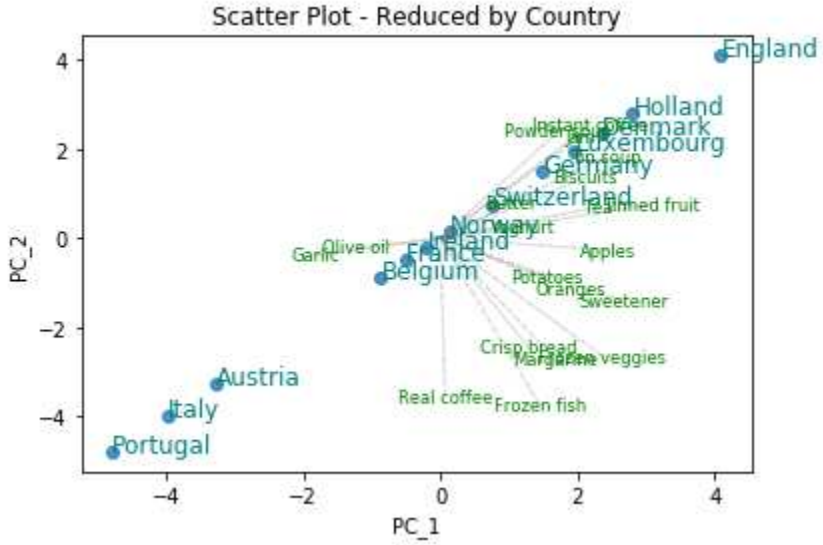
Question 2.4

The data food-consumption.csv contains 16 countries in the European area and their consumption for 20 food items, such as tea, jam, coffee, yoghurt, and others. There are some missing data entries: you may remove the rows “Sweden”, “Finland”, and “Spain”. The goal is to perform PCA analysis on the data, i.e., find a way to perform linear combinations of features across all 20 food-item consumptions, for each country. If we extract two principal components, that means we use two singular vectors that correspond to the largest singular values of the data matrix, in combining features. You will need to implement PCA by writing your own code.

4. (10 points) Now project each data point using the top two principal component vectors (thus now each data point will be represented using a two-dimensional vector). Draw a scatter plot of two-dimensional reduced representation for each country. What pattern can you observe?

Solution 2.4

The principal components have been plotted in a scatter plot. The food items have also been projected as a sub plot like a loading chart. The pattern seen is Olive and Garlic usage in Italy and Portugal, Tinned items in Switzerland and England, Frozen items in Norway and Denmark.


Scatter Plot - Reduced by Country
England
Holland
Denmark
Luxembourg
Germany
Switzerland
Norway
Ireland
France
Belgium
Austria
Italy
Portugal
Biscuits
Tinned fruit
Apples
Garlic
Olive oil
Potatoes
Oranges
Sweetener
Crisp bread
Real coffee
Frozen fish
PC_1
PC_2