

Homework 2

1. Concept Questions

1. Please prove the first principle component direction v corresponds to the largest eigenvector of the sample covariance matrix.

Note: I followed the steps in lecture notes from module 4, page 13 to 19

We want to find the direction w such that the variance of the data along direction w is maximized.

We start with the definition of variance: $\frac{1}{m}(w^T x^i - w^T \mu)^2$

Using some linear algebra, the above expression becomes:

$$\begin{aligned} & w^T \left(\frac{1}{m} \sum_{i=1}^m (x^i - \mu)(x^i - \mu)^T \right) w \\ &= w^T C w, \text{ where } C = \frac{1}{m} \sum_{i=1}^m (x^i - \mu)(x^i - \mu)^T = \text{covariance matrix} \end{aligned}$$

Now, we want to maximize this expression with respect to w to get the direction with largest variance.

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

From optimization, we introduce the equivalent Lagrangian function where λ = Lagrangian multiplier that represents the cost to violating the constraint $1 - ||w|| \geq 0$. We want to find the largest w that has the smallest cost of violating the restraint.

$$\min_{\lambda} \max_w L(w, \lambda) = w^T C w + \lambda(1 - ||w||^2)$$

To maximize, we set $\frac{\partial L}{\partial w} = 0$:

$$0 = 2Cw - 2\lambda w$$

$Cw = \lambda w \rightarrow w$ must be the eigenvectors of C and the eigenvector that corresponds to the largest eigenvalue is the direction where the variance of the data is maximized.

2. What is the relationship between SVD and eigendecomposition? Please explain this mathematically, and touch on why this is relevant for PCA.

Eigendecomposition (ED) and SVD are two ways to factor a matrix into its eigenvectors and eigenvalues that can be used to compute reduced representations or principle components. The eigenvectors are used to create the projection into a new space (R^k).

In ED, the goal is to maximize the variance of data along a direction w by solving for a covariance matrix, C , such that $Cw = \lambda w$. There will be multiple eigenvectors (w^i) and eigenvalue (λ_i) pairs so C can be written as $C = U \Lambda U^T$. From here, PCA can be used to select the k largest eigenvalues and their corresponding eigenvectors to reduce dimensionality.

In SVD, solving for C is not necessary; matrix can be factored directly into the eigenvectors and eigenvalues such that $M = U \Sigma V^T$ where U and V are the eigenvectors and Σ are the eigenvalues. PCA can also be performed on M by selecting the largest k eigenvalues and their corresponding eigenvectors.

SVD is preferred over ED when performing PCA because it's computationally cheaper and can be performed on rectangular matrices. You can perform PCA directly after SVD instead of computing C first then performing PCA as in ED.

3. Explain the three key ideas in ISOMAP (for manifold learning and non-linear dimensionality reduction).

In ISOMAP, the idea is to reduce data to a lower k dimensional representation such that distance between data points is preserved. The steps are:

1. Find neighbours of each data point within ϵ radius or such that each data point has k neighbours. Using Euclidean distance, populate adjacency matrix, A .
2. Find the geodesic distance/shortest distance between pairs of points based on A and populate matrix, D .
3. Find the low dimensional representation that preserves distance information in D .

2. Eigenfaces and Simple Face Recognition






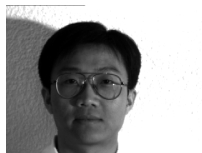
(a) Perform analysis on the Yale face dataset for Subject 1 and Subject 2, respectively, using all the images EXCEPT for the two pictures named subject01-test.gif and subject02-test.gif. Plot the first 6 eigenfaces for each subject. When visualizing, please reshape the eigenvectors into proper images. Please explain can you see any patterns in the top 6 eigenfaces?

Eigenface	Subject 1	Subject 2
1		
2		
3		
4		
5		
6		

Some Obeservations

1. The first two eigenfaces of both subjects closely mirror the characteristics of the sample photos that have the most shadows and contrast.

Eigenface	Subject	Sample
1		

Eigenface	Subject	Sample
2		
1		
2		

- Both subjects in the sample photos wear glasses in some of their photos - Subject 1 wears glasses in 1 out of 10 photos while Subject 2 wears them in 4 out of 9. However, glasses can only be detected in Subject 2's eigenfaces. If a test photo is presented of Subject 1 with glasses, image recognition have a hard time or may fail to identify the subject correctly.
- There is more contrast in the Subject 1's eigenfaces compared to Subject 2's. The shadows and highlights in the Subject 1's face are closer to black and white (especially in 3 to 6) and are more pronounced. This could be due to the difference in facial structure (eg - more pronounced cheekbones) and/or lighting.

(b) Now we will perform a simple face recognition task. Face recognition through PCA is proceeded as follows. Given the test image subject01-test.gif and subject02-test.gif, first downsize by a factor of 4 (as before), and vectorize each image. Take the top eigenfaces of Subject 1 and Subject 2, respectively. Then we calculate the projection residual of the 2 vectorized test images with the vectorized eigenfaces.

$$s_{ij} = \|(testimage)_j - (eigenface_i)(eigenface_i)^T(testimage)_j\|_2^2$$

Report all four scores: s_{ij} , $i = 1, 2$, $j = 1, 2$. Explain how to recognize the faces of the test images using these scores.

$$s_{1,1} = 6,091,271$$

$$s_{2,1} = 6,267,943$$

$$s_{1,2} = 4,503,195$$

$$s_{2,2} = 2,560,012$$

At test time, we only know that we have 2 sets of eigenfaces and 2 test images. In a true test, we wouldn't know which test image belongs which subject. To classify the images, we would need to compare the performance of each eigenface to the test subjects.

For example: I have two images, A and B and two sets of eigenfaces belonging to subject 1 and 2. We calculate the projection residual $s_{A,1}$, $s_{B,1}$, $s_{A,2}$, $s_{B,2}$.

If $s_{A,1} < s_{B,1}$, then it's likely that test image A belongs to Subject 1. Similarly, if $s_{B,2} < s_{A,2}$ then test image B belongs to Subject 2.

In the general case, if i =test image and j =eigenface set, you'd want to compare all s_{ij} where j is the same. The image with the smallest residual will belong to Subject j .

In this particular exercise, $s_{1,1} < s_{2,1}$ and $s_{2,2} < s_{1,2}$. This confirms that test image 1 belongs to Subject 1 and test image 2 belongs to Subject 2.

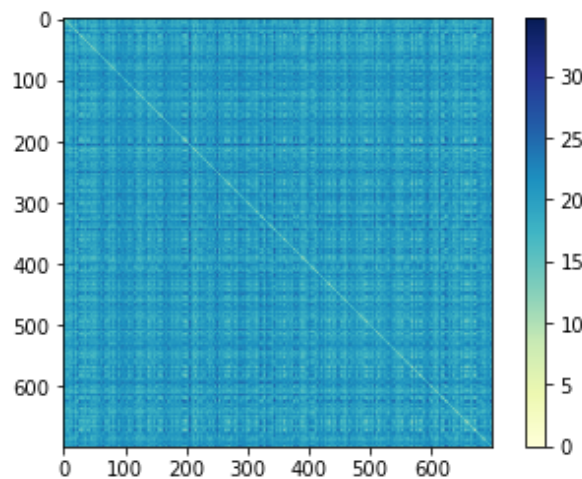
(c) Explain if eigenface facial recognition can work well, and discuss how it can be improved.

I think eigenface facial recognition could work well if you had a large database of labelled training images. In my earlier observations, Subject 1's eigenfaces didn't detect the subject's glasses because he only wears them 1 out of 10 training images. If there were more training images of the subject wearing glasses, they may have appeared in the lower eigenfaces.

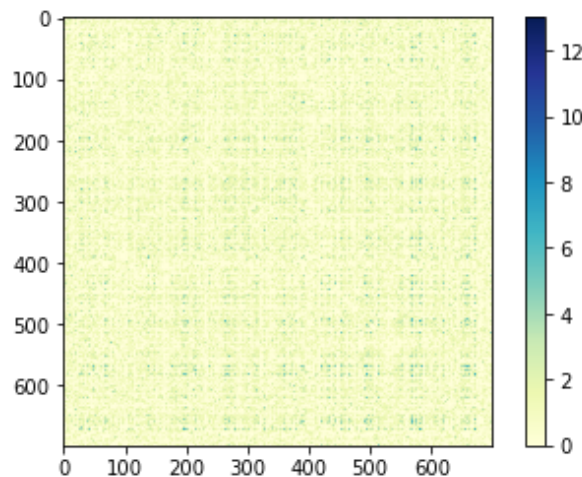
3. Order of Faces using ISOMAP

(a) Visualize the nearest neighbor graph (you can either show the adjacency matrix (e.g., as an image), or visualize the graph similar to the lecture slides using graph visualization packages such as Gephi (<https://gephi.org>) and illustrate a few images corresponds to nodes at different parts of the graph, e.g., mark them by hand or use software packages).

Below is the adjacency matrix with no epsilon adjustments.



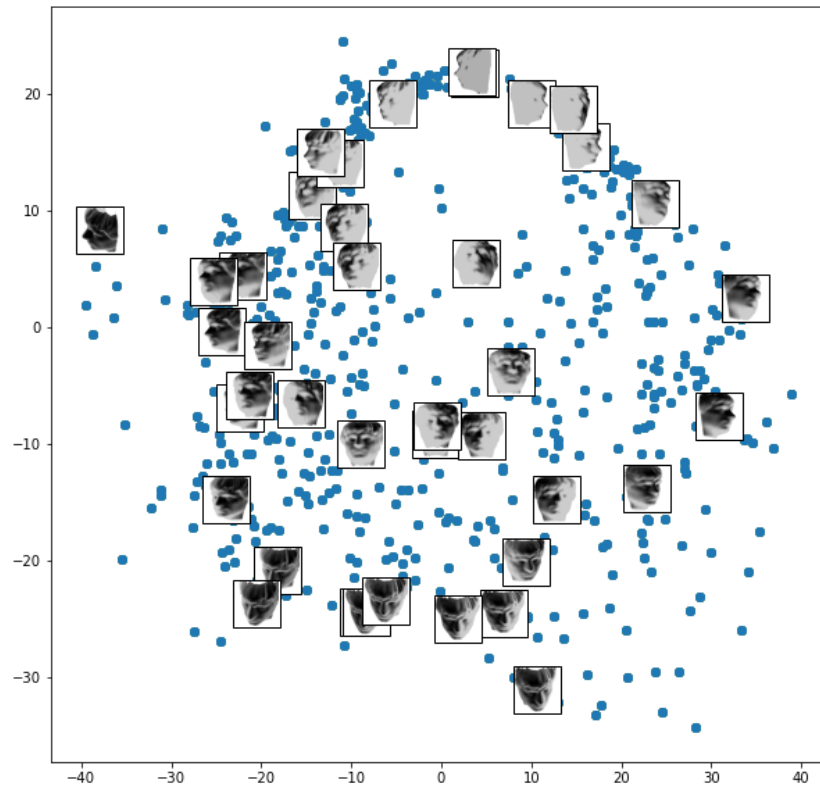
Here is the same adjacency matrix using $\epsilon = 13$.



I tuned ϵ by running the ISOMAP algorithm with multiple values and chose the value that closely produced the expected upside down Mickey Mouse shape. See Q3_output/Part_B folder for output images.

(b) Implement the ISOMAP algorithm yourself to obtain a two-dimensional low-dimensional embedding. Plot the embeddings using a scatter plot, similar to the plots in lecture slides. Find a few images in the embedding space and show what these images look like and specify the face locations on the scatter plot. Comment on do you see any visual similarity among them and their arrangement, similar to what you seen in the paper?

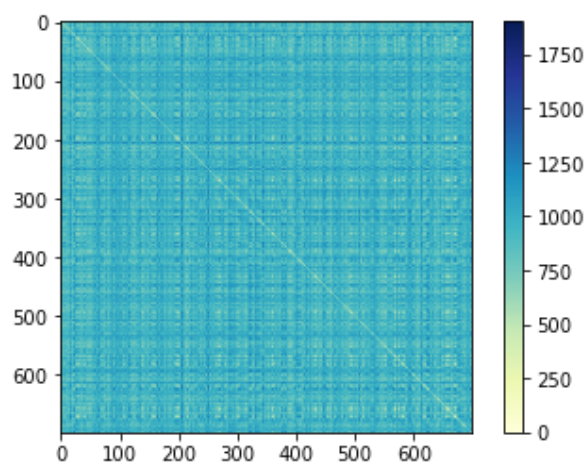
Images are arranged to correspond to the direction the face is pointing. Starting from the bottom center of the graph, the faces are looking down. As you move clockwise the the left, the faces also turn to face the left - this pattern continues as you move clockwise through the graph. The images in the center correspond to images where the face is looking forward. This is similar to the findings in the paper and in the professor's lecture.



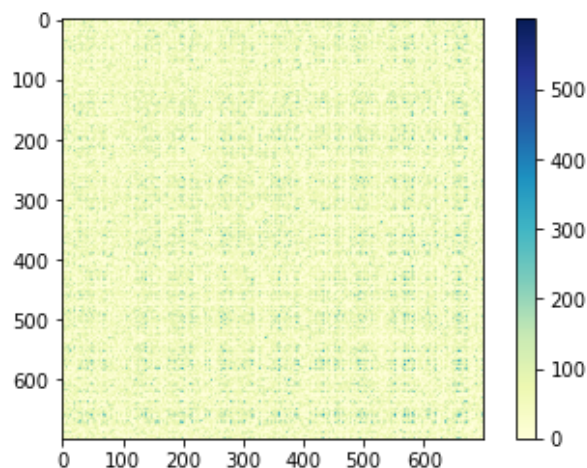
(c) Now choose l1 distance (or Manhattan distance) between images (recall the definition from “Clustering” lecture)). Repeat the steps above. Use ϵ -ISOMAP to obtain a $k = 2$ dimensional embedding. Present a plot of this embedding and specify the face locations on the scatter plot. Do you see any difference by choosing a different similarity measure by comparing results in Part (b) and Part (c)?

Adjacency Matrices

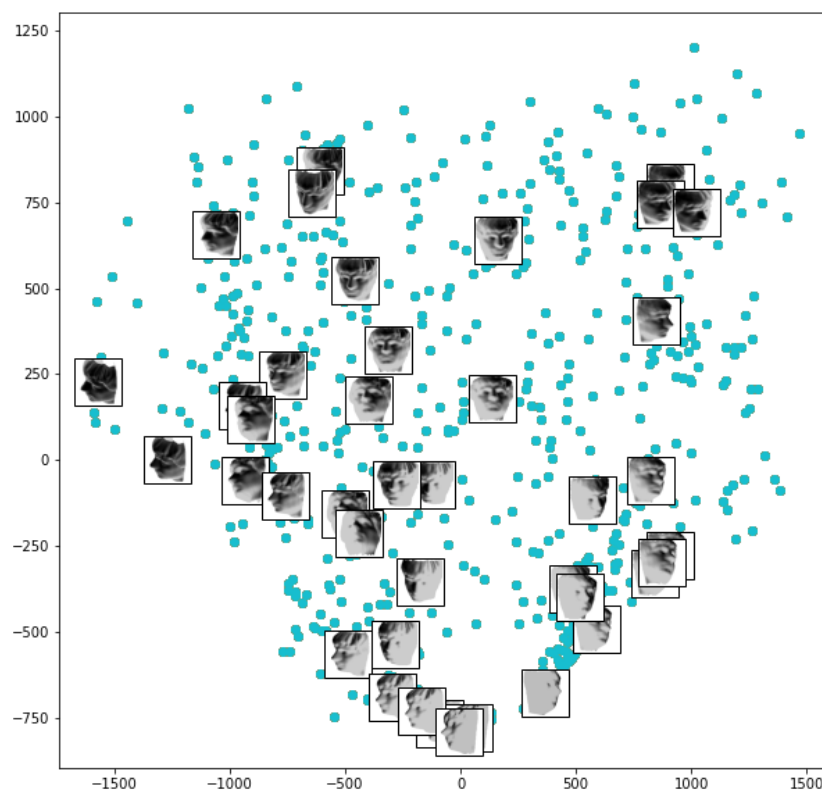
Adjacency matrix with no epsilon adjustments.



Adjacency matrix using $\epsilon = 600$. I tuned ϵ using the same procedure as Part A. The magnitude of L1 norms is much larger than the L2 norms.



Results with L1 Norm

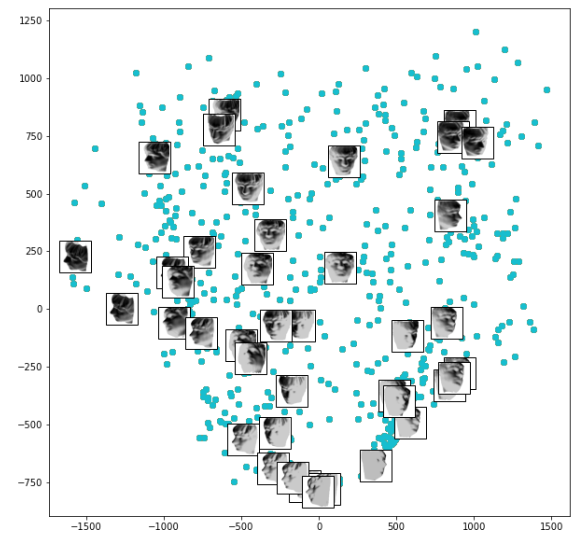
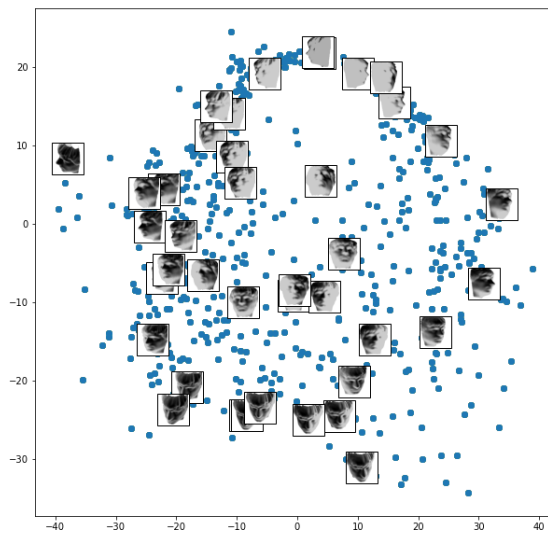


Compared to L2 Norm

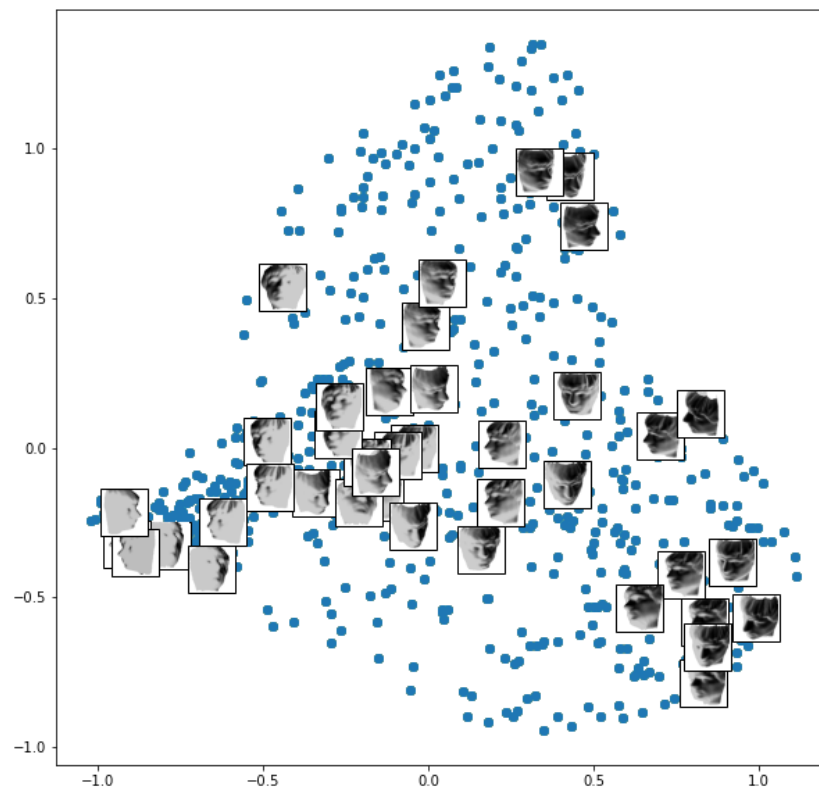
Similar to the L2 norm results, the L1 norm images are also arranged in a meaningful way but the overall shape is flipped, resulting in a normal Mickey Mouse shape. Starting at the top center image, the faces are pointing down and as you travel clockwise to the right, the direction of the face also turns right.

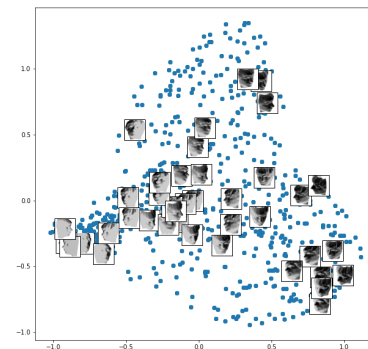
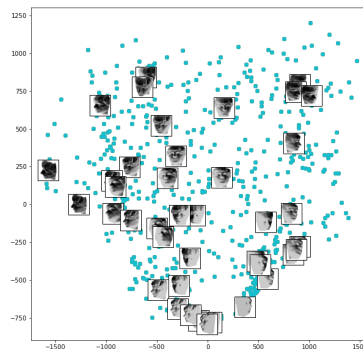
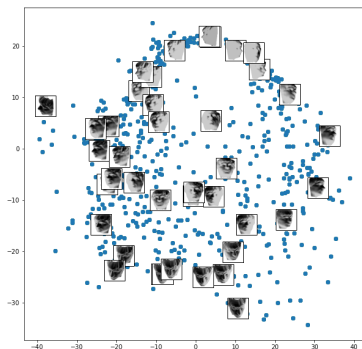
L2 Norm

L1 Norm



(d) Perform PCA (you can now use your implementation written in Question 1) on the images and project them into the top 2 principal components. Again show them on a scatter plot. Explain whether or you see a more meaningful projection using ISOMAP than PCA.





I think the ISOMAP representations provide a more meaningful arrangement compared to PCA. PCA appears to arrange the images based on the saturation of color. For example: in the left most point of the graph, there are two images that are facing opposite directions. The images in this corner are light and have very little black shading. Other clusters can be identified by looking at the level of shading in the images.

4. PCA: Food Consumption in European Countries

(a) For this problem of performing PCA on countries by treating each country's food consumption as their "feature" vectors, explain how the data matrix is set-up in this case (e.g., the columns and the rows of the matrix correspond to what).

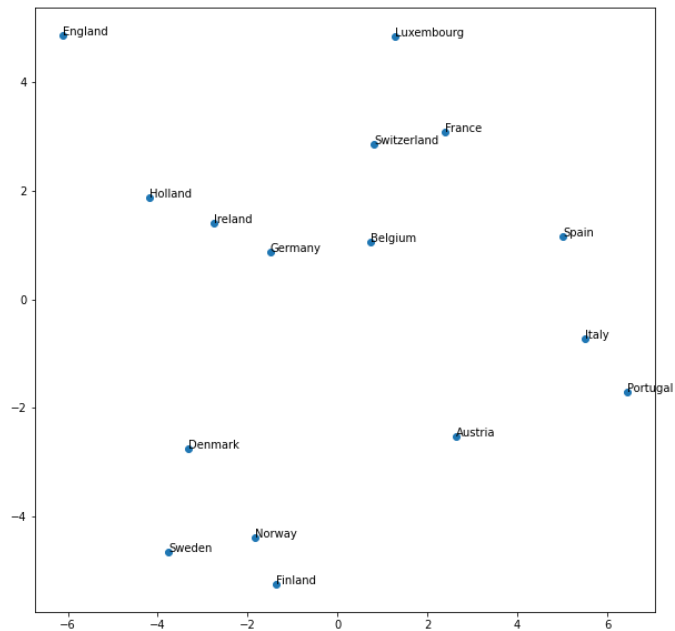
The rows of the food consumption data correspond to countries and the columns correspond to the food items. The matrix entries correspond to each countries' level of consumption of each food item.

Now extract the first two principal components for each data point (thus, this means we will represent each data point using a two-dimensional vector). Draw a scatter plot of two-dimensional representations of the countries using their two principal components. Mark the countries on the lot (you can do this by hand if you want). Please explain any pattern you observe in the scatter plot.

Clusters of countries can be identified in the scatterplot:

1. Scandinavia - Denmark, Sweden, Norway, Finland
2. Western Europe - Luxembourg, France, Switzerland, Belgium
3. Southern Europe - Spain, Italy, Portugal
4. Northern Europe - England, Holland, Ireland, Germany

The countries within each group are geographically close to each other and likely have similar diets.



(b) Now, we will perform PCA analysis on the data by treating country consumptions as “feature” vectors for each food item. In other words, we will now find weight vectors to combine country consumptions for each food item to perform PCA another way. Project data to obtain their two principle components (thus, again each data point – for each food item – can be represented using a two-dimensional vector). Draw a scatter plot of food items. Mark the food items on the plot (you can do this by hand if you do not want). Please explain any pattern you observe in the scatter plot.

Foods on the right side of the graph (power soup, instant coffee, crisp bread, frozen veggies, etc) are mostly non-perishables or food that lasts long if stored in the fridge. The left side of the graph (real coffee, orange, apples etc) are fresh foods in comparison. Garlic and olive oil look like outliers at the top of the graph - these are items that would be considered staples in some countries' cuisines (ex - Italy, Spain) but not in others (ex - Ireland, Sweden).

