Homework Assignment 2 – Math 118, Winter 2021

Dhruv Chakraborty

February 14, 2021

Problem 1

Write code that implements the alternating least squares algorithm for finding the non-negative matrix factorization. You can use the function `scipy.optimize.nnls` (in Python) or `lsqnonneg` (in MATLAB) to solve the least squares problems arising in each iteration of your algorithm. Use your code to find the NMF of A given below. Include your code and output as your answer.

$$\begin{bmatrix} 0.238 & 0.387 & 1.065 & 0.494 \\ 0.345 & 0.603 & 1.056 & 0.512 \\ 0.302 & 0.555 & 0.59 & 0.308 \\ 0.283 & 0.473 & 1.132 & 0.531 \end{bmatrix}$$

Solution:

```
import numpy as np
from scipy.optimize import nnls

A = np.array([[0.238, 0.387, 1.065, 0.494],
              [0.345, 0.603, 1.056, 0.512],
              [0.302, 0.555, 0.590, 0.308],
              [0.283, 0.473, 1.132, 0.531]])

def nmf(A, num_topics=2, tol=0.3):
    W, H = np.random.rand(A.shape[0], num_topics), np.random.rand(num_topics, A.shape[1])
    while True:
        for i in range(H.shape[0]):
            H[:,i] = nnls(W, A[:,i])[0]
        for i in range(W.shape[0]):
            W[i,:] = nnls(H.T, A[i,:])[0]
        error = np.linalg.norm(A-(W@H), ord='fro')
        if error <= tol:
            break
    return W, H

W, H = nmf(A)
```

Output:

$$W = \begin{bmatrix} 0.381 & 0.867 \\ 0.583 & 0.717 \\ 0.53 & 0.258 \\ 0.463 & 0.882 \end{bmatrix} \quad H = \begin{bmatrix} 0.503 & 0.941 & 0.549 & 0.671 \\ 0.034 & 0.0 & 0.955 & 0.36 \end{bmatrix} \quad A \approx WH = \begin{bmatrix} 0.221 & 0.359 & 1.038 & 0.568 \\ 0.318 & 0.549 & 1.005 & 0.65 \\ 0.276 & 0.499 & 0.538 & 0.449 \\ 0.263 & 0.435 & 1.096 & 0.628 \end{bmatrix}$$

Problem 2

Suppose that you run a small grocery store, and you wish to understand your customers better. Specifically, you'd like to *segment* your customers into a small number of groups with similar purchasing characteristics. So, you do the following:

1. You identify 6 main categories of product: Bread, Fresh fruit & Veggies, Cigarettes, Milk, Canned Food, Newspapers.
2. You select 5 customers and (legally and ethically!) track their purchasing habits for a month. Specifically you record the amount that they spend on each category of product in the month, and then normalize this amount to a number between 0 and 1.

This data is presented in matrix X. The first column corresponds to customer 1, the second to customer 2 and so on. The first row corresponds to Bread, the second to Fresh fruit & Veggies and so on. Also presented is the *non-negative matrix factorization* of X. That is, $X \approx WH$.

$$X = \begin{bmatrix} 0.735 & 0.179 & 0.079 & 0.698 & 0.307 \\ 0.093 & 0.655 & 0.655 & 0.436 & 0.604 \\ 0.686 & 0.782 & 0.547 & 0.427 & 0.471 \\ 0.138 & 0.223 & 0.573 & 0.846 & 0.164 \\ 0.427 & 0.018 & 0.488 & 0.193 & 0.819 \\ 0.78 & 0.69 & 0.354 & 0.221 & 0.184 \end{bmatrix} W = \begin{bmatrix} 0.162 & 0.449 & 0.696 \\ 0.784 & 0.0 & 0.0 \\ 0.632 & 0.599 & 0.0 \\ 0.413 & 0.0 & 0.648 \\ 0.62 & 0.0 & 0.0 \\ 0.336 & 0.766 & 0.0 \end{bmatrix} H = \begin{bmatrix} 0.268 & 0.557 & 0.896 & 0.538 & 0.854 \\ 0.94 & 0.595 & 0.0 & 0.0 & 0.0 \\ 0.226 & 0.0 & 0.0 & 0.917 & 0.0 \end{bmatrix}$$

1. Recall that the columns of W represent the categories, or segments, of customers. Interpret these segments based on the products they purchase the most.
2. With which segment is Customer 1 most strongly associated?
3. With which segment is Customer 3 most strongly associated?

Solution:

1. The three segments of customers may be interpreted as follows:
   (a) Customers belonging to the first category buy all or most of their groceries from our store. We can see this by the high values in rows 2 and 5, which correspond to fresh fruits & veggies and canned food respectively. It should be noted however that it appears customers belonging to this category do not spend much on bread.
   (b) The second segment seems to denote customers that make additional trips to purchase newspapers and/or cigarettes. These customers sometimes also add bread to these purchases.
   (c) The last category represents customers that only purchase the staple items of bread and milk at the store.
2. Customer 1 is most strongly associated with the category (b) as defined above.
3. Customer 3 is most strongly associated with category (a). In fact, they fully belong to this category.

Problem 3

Let $\mathcal{G}$ be a $2 \times 2 \times 4$ tensor with slices:

$$\mathcal{G}(:,:,1) = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \quad \mathcal{G}(:,:,2) = \begin{bmatrix} -2 & 1 \\ 1 & 2 \end{bmatrix} \quad \mathcal{G}(:,:,3) = \begin{bmatrix} 3 & 5 \\ -11 & 2 \end{bmatrix} \quad \mathcal{G}(:,:,4) = \begin{bmatrix} -2 & -1 \\ 7 & 1.5 \end{bmatrix}$$

Compute $\text{unfold}_1(\mathcal{G}), \text{unfold}_2(\mathcal{G})$ and $\text{unfold}_3(\mathcal{G})$.

Solution:

$$\text{unfold}_1(\mathcal{G}) = \begin{bmatrix} 1 & 2 & -2 & 1 & 3 & 5 & -2 & -1 \\ -2 & 1 & 1 & 2 & -11 & 2 & 7 & 1.5 \end{bmatrix} \in \mathbb{R}^{2\times(2\times 4)}$$

$$\text{unfold}_2(\mathcal{G}) = \begin{bmatrix} 1 & -2 & -2 & 1 & 3 & -11 & -2 & 7 \\ 2 & 1 & 1 & 2 & 5 & 2 & -1 & 1.5 \end{bmatrix} \in \mathbb{R}^{2\times(2\times 4)}$$

$$\text{unfold}_3(\mathcal{G}) = \begin{bmatrix} 1 & 2 & -2 & 1 \\ -2 & 1 & 1 & 2 \\ 3 & 5 & -11 & 2 \\ -2 & -1 & 7 & 1.5 \end{bmatrix} \in \mathbb{R}^{4\times(2\times 2)}$$

Problem 4

Consider the matrices:

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \end{bmatrix} \quad B = \begin{bmatrix} -2 & 2 \\ 3 & -3 \end{bmatrix} \quad \text{and } C = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Compute $[[A, B, C]]$. Express your answer by giving the slices of the resulting tensor.

Solution: Let $\mathcal{G} = [[A, B, C]] \in \mathbb{R}^{2\times 2\times 2}$. Then,

$$\begin{aligned} \mathcal{G} &= \sum_{i=1}^{2} \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i \\ &= \begin{pmatrix} 1 \\ 2 \end{pmatrix} \otimes \begin{pmatrix} -2 \\ 3 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \end{pmatrix} \otimes \begin{pmatrix} 2 \\ -3 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ -1 \end{pmatrix} \end{aligned}$$

where $(\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c})_{ijk} = \mathbf{a}_i \mathbf{b}_j \mathbf{c}_k$. So,

$$\mathcal{G}(:, :, 1) = \begin{bmatrix} -2 & 3 \\ -4 & 6 \end{bmatrix} \text{ and } \mathcal{G}(:, :, 2) = \begin{bmatrix} -8 & 12 \\ -10 & 15 \end{bmatrix}.$$

Problem 5

The following questions are related to Spectral Clustering. The notation is as established in Lecture 9. Recall that $(\mathbf{I}_S)_i = \begin{cases} \sqrt{\frac{|S^c|}{|S|}} & \text{if } v_i \in S \\ -\sqrt{\frac{|S|}{|S^c|}} & \text{if } v_i \notin S \end{cases}$ for any $S \subset V$.

1. Show that if $S \neq \emptyset, V$ then $\|\mathbf{I}_S\|_2 = \sqrt{n}$.

2. Suppose that G has two connected components: $V = C_1 \cup C_2$. Show that $L\mathbf{I}_{C_1} = 0$. Conclude that Spectral clustering will successfully find connected components of a graph.

3. It is a fact that $\mathbf{v}^\top L\mathbf{v} = \frac{1}{2}\sum_{i,j} A_{ij}(v_i - v_j)^2$ for any $\mathbf{v} \in \mathbb{R}^n$ *(You do not need to prove this!)*. Use this to show that:

$$\text{Rcut}(S) = \frac{1}{n^2}\mathbf{I}_S^\top L\mathbf{I}_S$$

Solution:

1. $||\mathbf{I}_S||_2 = \sqrt{\sum_{i=1}^{n}(\mathbf{I}_S)_i^2} = \sqrt{|S| \times \frac{|S^c|}{|S|} + |S^c| \times \frac{|S|}{|S^c|}} = \sqrt{|S^c| + |S|} = \sqrt{n}$ for $\emptyset \neq S \subsetneq V$.

2. Since G has two connected components such that $V = C_1 \cup C_2$, reorder its vertices so that we can write its adjacency matrix as $A = \begin{bmatrix} A_{C_1} & 0 \\ 0 & A_{C_2} \end{bmatrix}$ where A_{C_i} is the adjacency matrix of the subgraph represented by the connected component C_i. Note that $L = D - A$ so we also have $L = \begin{bmatrix} L_{C_1} & 0 \\ 0 & L_{C_2} \end{bmatrix}$. Moreover, reordering the vertices in this manner results in $\mathbf{I}_{C_1}^T = \left(\sqrt{\frac{|C_1|}{|C_2|}} \cdots \sqrt{\frac{|C_1|}{|C_2|}} \sqrt{\frac{|C_2|}{|C_1|}} \cdots \sqrt{\frac{|C_2|}{|C_1|}}\right)$. But then $L\mathbf{I}_{C_1} = \begin{bmatrix} L_{C_1} & 0 \\ 0 & L_{C_2} \end{bmatrix} \left(\sqrt{\frac{|C_1|}{|C_2|}} \cdots \sqrt{\frac{|C_1|}{|C_2|}} \sqrt{\frac{|C_2|}{|C_1|}} \cdots \sqrt{\frac{|C_2|}{|C_1|}}\right)^T = L_{C_1}(\frac{|C_1|}{|C_2|} \times \vec{1}) + L_{C_2}(\frac{|C_2|}{|C_1|} \times \vec{1})$. Finally, since $L1 = 0$, we have $L\mathbf{I}_{C_1} = 0$.

   Note that $L1 = 0$ also means that 0 is the smallest eigenvector of L. But $L\mathbf{I}_{C_1} = 0$ means that 0 is a repeated eigenvalue and in particular $\lambda_2 = 0$. Therefore, by definition of $\mathbf{I}_{C_1}$ (the second eigenvector of L), spectral clustering correctly finds the connected components of a graph.

3.

$$\begin{aligned}
\text{Rcut}(S) = \frac{e(S, S^c)}{|S||S^c|} &= \frac{1}{2|S||S^c|}\sum_{i,j} A_{ij}\frac{(\mathbf{I}_{S_i} - \mathbf{I}_{S_j})^2}{(\sqrt{|S^c|/|S|} + \sqrt{|S|/|S^c|})^2} \\
&= \frac{1}{2|S||S^c|}\sum_{i,j} A_{ij}\frac{(\mathbf{I}_{S_i} - \mathbf{I}_{S_j})^2}{(|S_c|^2 + |S|^2 + 2|S||S^c|)/|S||S^c|} \\
&= \frac{|S||S^c|}{2|S||S^c| \times (|S| + |S^c|)^2}\sum_{i,j} A_{ij}(\mathbf{I}_{S_i} - \mathbf{I}_{S_j})^2 \\
&= \frac{1}{2n^2}\sum_{i,j} A_{ij}(\mathbf{I}_{S_i} - \mathbf{I}_{S_j})^2 \\
&= \frac{1}{n^2}\mathbf{I}_S^\top L\mathbf{I}_S \text{ by the given fact.}
\end{aligned}$$

Problem 6

For this question you will use the provided notebook: labeled "Question on Spectral Clustering". Write code that implements the spectral clustering algorithm, for two clusters, in Python. Use the built in function `numpy.linalg.eig` to find the eigenvectors. Use your code to identify the clusters in the Mystery Graph provided (there are two of them). Include your code and a list of vertices in both clusters as your answer.

Solution:

```
import numpy as np
import networkx as nx
import pickle

A = pickle.load(open('MysteryGraph.p','rb'))
G = nx.from_numpy_matrix(A)

def spectral_clustering(G):
    L = nx.laplacian_matrix(G).toarray()
    eig, eiv = np.linalg.eig(L)
    v2 = eiv[:,list(eig).index(sorted(eig)[1])]

    print('C1:')
    for i, val in enumerate(v2):
        if val < 0:
            print(i)

    print('C2:')
    for i, val in enumerate(v2):
        if val >= 0:
            print(i)

spectral_clustering(G)
```

- Vertices in $C1$: 1, 3, 5, 6, 8, 11, 13, 14, 16, 19, 20, 21, 22, 23, 25, 27, 29, 33, 34, 35.
- Vertices in $C2$: 0, 2, 4, 7, 9, 10, 12, 15, 17, 18, 24, 26, 28, 30, 31, 32, 36, 37, 38, 39.

Problem 9

Let $P = A^\top (D^{\text{out}})^{-1}$ denote the transition matrix for a directed graph G (as defined in Lecture 11). Prove that P is column stochastic.

Solution: $P = A^T (D^{\text{out}})^{-1} \implies$ the i-th column of $P, \mathbf{p}_i = \begin{pmatrix} \mathbf{a}_1 \mathbf{d}_{\text{out}_i}^{-1} \\ \vdots \\ \mathbf{a}_n \mathbf{d}_{\text{out}_i}^{-1} \end{pmatrix} = \frac{1}{d_{\text{out}_i}} \begin{pmatrix} \mathbf{a}_{1i} \\ \vdots \\ \mathbf{a}_{ni} \end{pmatrix}$. But then note that $\sum_j \mathbf{p}_{i_j} = \frac{1}{d_{\text{out}_i}} \sum_j \mathbf{a}_{ji} = \frac{1}{d_{\text{out}_i}} \times d_{\text{out}_i} = 1$ i.e. P is column stochastic.

Problem 10

In this question you will work through a single iteration of k-means for a simple data set. Let $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{10}\} \subset \mathbb{R}$ where:

$$\mathbf{x}_1 = -0.82, \quad \mathbf{x}_2 = -1.33 \quad \mathbf{x}_3 = -3.63, \quad \mathbf{x}_4 = -1.62, \quad \mathbf{x}_5 = -2.95$$
$$\mathbf{x}_6 = 0.95, \quad \mathbf{x}_7 = 0.53, \quad \mathbf{x}_8 = 1.79, \quad \mathbf{x}_9 = 0.95 \text{ and } \mathbf{x}_{10} = 2.28$$

We will take $k = 2$, that is we are seeking to partition $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$. Suppose that we initialize the cluster centers randomly as $\mu_1^{(0)} = -1.56$ and $\mu_2^{(0)} = 2.37$.

1. Compute the distances $\|\mathbf{x}_i - \mu_a^{(0)}\|_2$ for $i = 1, \ldots, 10$ and $a = 1, 2$. Write these as a 10×2 matrix.

2. Now, using these distances compute $\mathcal{X}_1^{(1)}$ and $\mathcal{X}_2^{(1)}$.

3. Finally, compute the new centroids $\mu_1^{(1)}$ and $\mu_2^{(1)}$.

Optional: Check your answer using Python.

Solution:

1. $\|\mathbf{x}_i - \mu_a^{(0)}\|_2 = \begin{bmatrix} 0.74 & 3.19 \\ 0.23 & 3.70 \\ 2.07 & 6.00 \\ 0.06 & 3.99 \\ 1.39 & 5.32 \\ 2.51 & 1.42 \\ 2.09 & 1.84 \\ 3.35 & 0.58 \\ 2.51 & 1.42 \\ 3.84 & 0.09 \end{bmatrix}$ where the i, j-th element is the distance of $\mathbf{x}_i$ from cluster center $\mu_j^{(0)}$.

2. Using the distances from above, we get $\mathcal{X}_1^{(1)} = \{1, 2, 3, 4, 5\}$ and $\mathcal{X}_2^{(1)} = \{6, 7, 8, 9, 10\}$.

3. We compute the updated centroid $\mu_i^{(1)}$ as the mean of the data points now in cluster i. Hence, we get $\mu_1^{(1)} = -2.07$ and $\mu_2^{(1)} = 1.3$.