

Homework 3 - Maths Foundation for Machine Learning

kipngeno koech - bkoech

October 16, 2024

Question 1: Eigen Values and Eigen Vectors

Consider the matrix $A = \begin{bmatrix} 9 & 1 & -1 \\ -1 & 11 & 1 \\ -2 & 2 & 10 \end{bmatrix}$.

1. Obtain the characteristic polynomial of A . (Hint: find the determinant of $A - \lambda I$)

$$A - \lambda I = \begin{bmatrix} 9 - \lambda & 1 & -1 \\ -1 & 11 - \lambda & 1 \\ -2 & 2 & 10 - \lambda \end{bmatrix}$$

2. From the characteristic polynomial, obtain the 3 eigenvalues of $\lambda_1, \lambda_2, \lambda_3$, such that $\lambda_1 < \lambda_2 < \lambda_3$.

$$\det(A - \lambda I) = 0$$

$$\det \begin{bmatrix} 9 - \lambda & 1 & -1 \\ -1 & 11 - \lambda & 1 \\ -2 & 2 & 10 - \lambda \end{bmatrix} = 0$$

$$(9 - \lambda)((11 - \lambda)(10 - \lambda) - (2 \times 1)) - 1((-1)(10 - \lambda) - (-2 \times 1)) - 1((-1)(2) - (-2)(11 - \lambda)) = 0$$

$$(9 - \lambda)((11 - \lambda)(10 - \lambda) - 2) - 1((-10 + \lambda) + 2) - 1(-2 + 22 - 2\lambda) = 0$$

$$(9 - \lambda)(110 - 21\lambda + \lambda^2 - 2) - 1(-8 + \lambda) - 1(20 - 2\lambda) = 0$$

$$(9 - \lambda)(108 - 21\lambda + \lambda^2) - 1(-8 + \lambda) - 1(20 - 2\lambda) = 0$$

$$9(108 - 21\lambda + \lambda^2) - \lambda(108 - 21\lambda + \lambda^2) + 8 - \lambda - 20 + 2\lambda = 0$$

$$972 - 189\lambda + 9\lambda^2 - 108\lambda + 21\lambda^2 - \lambda^3 + 8 - \lambda - 20 + 2\lambda = 0$$

$$-\lambda^3 + 30\lambda^2 - 296\lambda + 960 = 0$$

$$\lambda^3 - 30\lambda^2 + 296\lambda - 960 = 0$$

$$(\lambda - 10)(\lambda^2 - 20\lambda + 96) = 0$$

$$(\lambda - 10)(\lambda - 8)(\lambda - 12) = 0$$

$$\lambda_1 = 8, \lambda_2 = 10, \lambda_3 = 12$$

3. Using Gaussian elimination, find the eigenvectors of A corresponding to the eigenvalues $\lambda_1, \lambda_2, \lambda_3$.
for eigen value $\lambda_1 = 8$

$$A - 8I = \begin{bmatrix} 1 & 1 & -1 \\ -1 & 3 & 1 \\ -2 & 2 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & -1 \\ -1 & 3 & 1 \\ -2 & 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Augumented matrix:

$$\begin{aligned} & \begin{bmatrix} 1 & 1 & -1 & | & 0 \\ -1 & 3 & 1 & | & 0 \\ -2 & 2 & 2 & | & 0 \end{bmatrix} \xrightarrow{R_2=R_2+R_1} \begin{bmatrix} 1 & 1 & -1 & | & 0 \\ 0 & 4 & 0 & | & 0 \\ -2 & 2 & 2 & | & 0 \end{bmatrix} \\ & \xrightarrow{R_3=R_3+2R_1} \begin{bmatrix} 1 & 1 & -1 & | & 0 \\ 0 & 4 & 0 & | & 0 \\ 0 & 4 & 0 & | & 0 \end{bmatrix} \xrightarrow{R_3=R_3-R_2} \begin{bmatrix} 1 & 1 & -1 & | & 0 \\ 0 & 4 & 0 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix} \\ & \xrightarrow{R_2=\frac{1}{4}R_2} \begin{bmatrix} 1 & 1 & -1 & | & 0 \\ 0 & 1 & 0 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix} \xrightarrow{R_1=R_1-R_2} \begin{bmatrix} 1 & 0 & -1 & | & 0 \\ 0 & 1 & 0 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix} \end{aligned}$$

$$x_1 - x_3 = 0 \Rightarrow x_1 = x_3$$

$$x_2 = 0$$

Let $x_3 = 1$, then $x_1 = 1$ and $x_2 = 0$. Therefore, the eigenvector corresponding to $\lambda_1 = 8$ is $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.
for eigen value $\lambda_2 = 10$

$$A - 10I = \begin{bmatrix} -1 & 1 & -1 \\ -1 & 1 & 1 \\ -2 & 2 & 0 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 1 & -1 \\ -1 & 1 & 1 \\ -2 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Augumented matrix:

$$\begin{bmatrix} -1 & 1 & -1 & | & 0 \\ -1 & 1 & 1 & | & 0 \\ -2 & 2 & 0 & | & 0 \end{bmatrix} \xrightarrow{R_1 = -R_1} \begin{bmatrix} 1 & -1 & 1 & | & 0 \\ -1 & 1 & 1 & | & 0 \\ -2 & 2 & 0 & | & 0 \end{bmatrix}$$

$$\xrightarrow{R_2 = R_2 + R_1} \begin{bmatrix} 1 & -1 & 1 & | & 0 \\ 0 & 0 & 2 & | & 0 \\ -2 & 2 & 0 & | & 0 \end{bmatrix} \xrightarrow{R_3 = R_3 + 2R_1} \begin{bmatrix} 1 & -1 & 1 & | & 0 \\ 0 & 0 & 2 & | & 0 \\ 0 & 0 & 2 & | & 0 \end{bmatrix}$$

$$\xrightarrow{R_3 = R_3 - R_2} \begin{bmatrix} 1 & -1 & 1 & | & 0 \\ 0 & 0 & 2 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix} \xrightarrow{R_2 = \frac{1}{2}R_2} \begin{bmatrix} 1 & -1 & 1 & | & 0 \\ 0 & 0 & 1 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}$$

$$\xrightarrow{R_1 = R_1 + R_2} \begin{bmatrix} 1 & -1 & 0 & | & 0 \\ 0 & 0 & 1 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}$$

$$x_1 - x_2 = 0, x_3 = 0 \Rightarrow x_1 = x_2$$

Let $x_2 = 1$, then $x_1 = 1$ and $x_3 = 0$. Therefore, the eigenvector corresponding to $\lambda_2 = 10$ is $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$.
for eigen value $\lambda_3 = 12$

$$A - 12I = \begin{bmatrix} -3 & 1 & -1 \\ -1 & -1 & 1 \\ -2 & 2 & -2 \end{bmatrix}$$

$$\begin{bmatrix} -3 & 1 & -1 \\ -1 & -1 & 1 \\ -2 & 2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -3 & 1 & -1 & | & 0 \\ -1 & -1 & 1 & | & 0 \\ -2 & 2 & -2 & | & 0 \end{bmatrix} \xrightarrow{R_1 = -\frac{1}{3}R_1} \begin{bmatrix} 1 & -\frac{1}{3} & \frac{1}{3} & | & 0 \\ -1 & -1 & 1 & | & 0 \\ -2 & 2 & -2 & | & 0 \end{bmatrix}$$

$$\xrightarrow{R_2 = R_2 + R_1} \begin{bmatrix} 1 & -\frac{1}{3} & \frac{1}{3} & | & 0 \\ 0 & -\frac{4}{3} & \frac{4}{3} & | & 0 \\ -2 & 2 & -2 & | & 0 \end{bmatrix} \xrightarrow{R_3 = R_3 + 2R_1} \begin{bmatrix} 1 & -\frac{1}{3} & \frac{1}{3} & | & 0 \\ 0 & -\frac{4}{3} & \frac{4}{3} & | & 0 \\ 0 & \frac{4}{3} & -\frac{4}{3} & | & 0 \end{bmatrix}$$

$$\xrightarrow{R_3 = R_3 + R_2} \begin{bmatrix} 1 & -\frac{1}{3} & \frac{1}{3} & | & 0 \\ 0 & -\frac{4}{3} & \frac{4}{3} & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix} \xrightarrow{R_2 = -\frac{3}{4}R_2} \begin{bmatrix} 1 & -\frac{1}{3} & \frac{1}{3} & | & 0 \\ 0 & 1 & -1 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}$$

$$\xrightarrow{R_1 = R_1 + \frac{1}{3}R_2} \begin{bmatrix} 1 & 0 & 0 & | & 0 \\ 0 & 1 & -1 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}$$

$$x_1 = 0, x_2 = x_3$$

Let $x_2 = 1$, then $x_1 = 0$ and $x_3 = 1$. Therefore, the eigenvector corresponding to $\lambda_3 = 12$ is $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

4. (in the notebook)
5. Compare the eigen vectors computed using `np.linalg.eig` and that obtained using Gaussian elimination in (3) above. Are they the same? If they are the same, explain why this is the case. Otherwise, explain why they are different eigen vectors using `np.linalg.eig`:

$$\begin{pmatrix} 9.00258517e-16 & -7.07106781e-01 & 7.07106781e-01 \\ -7.07106781e-01 & -3.36518470e-16 & -7.07106781e-01 \\ -7.07106781e-01 & -7.07106781e-01 & -5.62025848e-15 \end{pmatrix}$$

eigen vectors using Gaussian elimination:

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

The eigen vectors are different. This is because the eigen vectors obtained using `np.linalg.eig` normalizes the eigen vectors to have a norm of 1. The eigen vectors obtained using Gaussian elimination are not normalized.

6. Now, given a diagonal matrix D , whose diagonal entries are the eigenvalues of A , and a matrix P whose columns are the eigenvectors of A , use the eigen values you obtained in (2) and the eigenvectors you obtained in (3) to show that A as PDP^{-1} .

$$D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 12 \end{bmatrix}, P = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}, A = \begin{bmatrix} 9 & 1 & -1 \\ -1 & 11 & 1 \\ -2 & 2 & 10 \end{bmatrix}$$

$$PDP^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 8 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 12 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}^{-1}$$

$$P^{-1} = \begin{pmatrix} 0.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 \\ -0.5 & 0.5 & 0.5 \end{pmatrix}$$

$$PDP^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 8 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 12 \end{bmatrix} \begin{bmatrix} 0.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 \\ -0.5 & 0.5 & 0.5 \end{bmatrix}$$

$$DP^{-1} = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 12 \end{bmatrix} \begin{bmatrix} 0.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 \\ -0.5 & 0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} 4 & -4 & 4 \\ 5 & 5 & -5 \\ -6 & 6 & 6 \end{bmatrix}$$

$$P(DP^{-1}) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & -4 & 4 \\ 5 & 5 & -5 \\ -6 & 6 & 6 \end{bmatrix} = \begin{bmatrix} 9 & 1 & -1 \\ -1 & 11 & 1 \\ -2 & 2 & 10 \end{bmatrix}$$

Therefore, $A = PDP^{-1}$.

7. Calculate the determinant of A, $\det(A)$:

$$\begin{aligned} \det(A) &= 9((11 \times 10) - (2 \times 1)) - 1((-1 \times 10) - (-2 \times 1)) - 1((-1 \times 2) - (11 \times -2)) \\ &= 9(110 - 2) - 1(-10 + 2) - 1(-2 + 22) = 9(108) - 1(-8) - 1(20) = 972 + 8 - 20 = \mathbf{960} \end{aligned}$$

8. Calculate the trace of A, $\text{tr}(A)$:

$$\text{tr}(A) = 9 + 11 + 10 = \mathbf{30}$$

9. confirm the following holds:

$$\det(A) = \lambda_1 \lambda_2 \lambda_3$$

$$960 = 8 \times 10 \times 12 = 960$$

$$\text{tr}(A) = \lambda_1 + \lambda_2 + \lambda_3$$

$$30 = 8 + 10 + 12 = 30$$

Question 2: Image Compression using SVD

5. 5. Assuming 1 byte of memory is required for one element of the image matrix, determine the top k rank-1 matrices required to represent the image if a compression ratio of 13:1 is desired. What is the value of k ? Add up the rank-1 matrices for the top K singular values and visualize your results. Remember to show your workings

$$\text{Compression ratio} = \frac{\text{Original size}}{\text{Compressed size}} = 13$$

Original size:

$$\text{Original size} = \text{Number of elements in the image} = 902 \times 602 = \mathbf{543004} \text{ bytes}$$

$$\text{Compressed size} = \text{Number of elements in the top } k \text{ rank-1 matrices} = k \times 902 + k \times 602 = 1504k \text{ bytes}$$

$$\frac{543004}{1504k} = 13 = \frac{543004}{13} = 1504k$$

$$k = \frac{543004}{13 \times 1504} = \mathbf{28}$$

Question 3: Linear Regression

In linear regression, SVD can help by decomposing the design matrix, identifying dependencies between predictor variables, and enabling regularization techniques like ridge regression. By reducing the rank or eliminating singular values, SVD assists in stabilizing regression models and improving their predictive accuracy. In this question you will solve a linear regression problem using SVD. Complete the cells in the provided starter code and answer the following:

1. What is the size of the dataset?

Size of the dataset = **489**

2. How many singular values were obtained during the calculation of the co-efficients? Why is this number exact - neither more nor less?

Question 4: PCA & Eigenfaces

capture the maximum variance in the data of high dimensional vectors. An image can be seen as a high dimensional vector. For example, a 100x100 image is a 10000 dimensional vector. PCA can be used to reduce this dimensionality by projecting images onto a lower dimensional subspace, given a dataset of images. 2 Eigenfaces is an application of PCA to obtain a lower dimensional subspace onto which face images can be projected, such that face images can be reconstructed by linearly combining the basis vectors of the lower dimensional subspace. The basis vectors of this lower dimensional subspace are called the eigenfaces. Eigenfaces can be applied in face recognition. [You can read more about PCA in chapter 10 of the textbook and study this Wikipedia article on Eigenface] In this exercise, you will go through the process of implementing a face recognition system using eigenfaces. You will use faces from the olivetti faces dataset. A starter notebook has been provided for you in the handout.

1. In the starter code, the olivetti faces dataset has been loaded, using a scikit-learn function. A database of faces, a list of test faces, and a list of faces of people not in the database have been created.

- (a) How many face images are in the database?

Number of face images in the database = **400**

- (b) How many images are in the list of test faces?

Number of face images in the test faces = **36**

- (c) How many images are in the list of faces of people not in the database?

Number of face images in the faces of people not in the database = **40**

- (d) What are the dimensions of each face image?

Dimensions of each face image = **64 x 64**

2. Create a matrix T containing all faces in the database, such that each row corresponds to the pixels of a person's face. Note that the height of this matrix should be equal to the number of faces in the database. What are the dimensions of T ?

Dimensions of matrix T = **400 x 4096** *N/B : which dataset use? (64 * 64 = 4096)*

3. Obtain and visualize the mean face

Mean face = **in the notebook**

4. Obtain the covariance matrix, S, of the matrix T

Covariance matrix, S = **in the notebook**

- 5.