COMSATS University Islamabad
Department of Mahematics

Some Mathematical Tools For Data Science

Haider Rasool Qadri
CIIT/SP20-BSM-014/ISB

Supervisor
Dr. Salman Amin Malik

January 22, 2024

Layout

1. Outlines
    - Eigenvalues and Eigenvectors
    - Symmetric Positive Definite Matrices
    - Singular Value Decomposition (SVD)
    - Image Compression Using SVD
    - Mean, Variance and Covariance
    - Principle Component Analysis

2. Conclusions

Eigenvalues and Eigenvectors

- Eigenvalues and Eigenvectors
  - In Linear Algebra, an eigenvector or characteristic vector for a square matrices is a non-zero vector that does not change its direction under the associated linear transformation. Mathematically,

$$\mathrm{Ax} = \lambda \mathrm{x} \qquad \begin{array}{c} \mathrm{x} \text{ eigenvector of } \mathrm{A} \\ \lambda \text{ eigenvalue of } \mathrm{A} \end{array}$$

  - The matrix $(\mathrm{A} - \lambda \mathrm{I_n})$ is called the characteristic matrix of A where $\mathrm{I_n}$ is Identity matrix.
  - The equation $\det(\mathrm{A} - \lambda \mathrm{I_n}) = 0$ is called characteristic equation of A and the roots of this equation are called eigenvalues of the matrix A.

Continued...

- Properties
  - Eigenvectors with distinct Eigenvalues are linearly Independent.
  - The sum of the eigenvalues of a matrix is equal to the trace of the matrix (sum of diagonal elements).
  $$\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i$$
  - The product of the eigenvalues is equal to the determinant of the matrix.
  $$\det(A) = \prod_{i=1}^{n} \lambda_i$$

Computations of Eigenvalues and Eigenvectors

- We summarize the computational approach for determining the eigenvalues and eigenvectors as a two step procedure :

Step 1 To find the Eigenvalues of A :

Compute the roots of the characteristic equation $\det\left(\mathrm{A} - \lambda \mathrm{I_n}\right) = 0$

Step 2 To find an Eigenvector corresponding to an eigenvalue λ:

Compute a non-trivial solution to the homogeneous linear system $(\mathrm{A} - \lambda \mathrm{I_n})\mathrm{x} = 0$

Symmetric Matrices

Symmetric matrices are square matrices with the property $S = S^T$ and they deserve all the attention they get. Looking at their eigenvalues and eigenvectors, you see why they are special :

- All n eigenvalues λ of a symmetric matrix S are real numbers.
- The n eigenvectors q can be chosen orthogonal (perpendicular to each other).
- The identity matrix $S = I$ is an extreme case.

Continued...

- With repeated eigenvalues like $\lambda_1 = \lambda_2 = 1$, we have a choice of eigenvectors. We can choose them to be orthogonal. And we can rescale them to be unit vectors. Then those eigenvectors $q_1, \ldots, q_n$ are not just orthogonal they are orthonormal.
- We write Q instead of X for the eigenvector matrix of S, to emphasize that these eigenvectors are orthonormal: $Q^T = Q^{-1}$.
- The eigenvector matrix for S has $Q^T Q = I$.
- This eigenvector matrix is an orthogonal matrix.
- Spectral Theorem :

$$\boxed{\text{Every real symmetric matrix has the form } S = Q \wedge Q^T}$$

Symmetric Positive Definite Matrices

- We are working with real symmetric matrices $S = S^T$. Some of those symmetric matrices (not all) have a further powerful property that puts them at the center of applied mathematics.

$$\boxed{\text{A positive definite matrix has all positive eigenvalues}}$$

- May I bring forward the most important idea about positive definite matrices? This approach doesn't involve eigenvalues, but it turns out to be a perfect test for $\lambda > 0$. This is good definition of positive definite matrices: the energy test.

$$\boxed{\text{S is positive definite if the energy } x^T S x \text{ is positive for all vectors } x \neq 0}$$

Of course $S = I$ is positive definite: All $\lambda_i = 1$. The energy is $x^T I x = x^T x$, positive if $x \neq 0$.

Example

- Suppose S is symmetric positive definite 2 by 2 matrix. Apply the test:
  $S = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ I will choose an example with $a = c = 5$ and $b = 4$.
  This matrix S has $\lambda = 9$ and $\lambda = 1$.
  $$\text{Energy } E = x^T S x \qquad \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 5x^2 + 8xy + 5y^2 > 0$$
  The graph of that energy function E(x,y) is a bowl opening upwards. The bottom point of the bowl has energy $E = 0$ when $x = y = 0$.

Singular Value Decomposition

- If A is not square then $Ax = \lambda x$ is impossible and eigenvectors fail (left side in R^m, right side in R^n). We need an idea that succeeds for every matrix.

$$\boxed{\text{The singular Value Decomposition of A is} \quad A = U\sum V^T}$$

where,

$$\begin{gathered} V \text{ contains orthonormal eigenvectors of } A^TA \\ U \text{ contains orthonormal eigenvectors of } AA^T \\ \sigma_1^2 \text{ to } \sigma_n^2 \text{ are the nonzero eigenvalues of both } A^TA \text{ and } AA^T \end{gathered}$$

- The column-row multiplication of $U\sum$ times V^T separates A into n pieces of rank 1:

$$\text{Pieces of the SVD} \qquad \boxed{A = U\sum V^T = \boldsymbol{\sigma}_1 u_1 v_1^T + \ldots + \boldsymbol{\sigma}_n u_n v_n^T}$$

Important Fact For Data Science

- Why is the SVD so important?

  Like the other factorizations $A = LU$ and $S = Q \wedge Q^T$, it separates the matrix into rank one pieces. A special property of the SVD is that those pieces come in order of important.

$$A = \sigma_1 u_1 v_1^T + ... + \sigma_n u_n v_n^T$$

- The first piece $\sigma_1 u_1 v_1^T$ is the closet rank one matrix of A.

Continued...

- More than that is true: The sum of the first k pieces is best possible for rank k.

$$A_k = \sigma_1 u_1 v_1^T + \ldots + \sigma_k u_k v_k^T \text{ is the best rank k approximation to A}$$

- Eckart and Young gave a proof in 1936 (using the Frobenius norm for matrices).

$$\text{Eckart-Young} \qquad \boxed{\text{If B has rank k then } ||A - A_k|| \leq ||A - B||}$$

Steps For Singular Value Decomposition

Step 1 Calculate AA^T and A^TA

Step 2 Calculate Singular values and Σ

Step 3 Calculate U using AA^T

Step 4 Calculate V using A^TA

Step 5 The Complete SVD

$$A = u_1\sigma_1 v_1^T + u_2\sigma_2 v_2^T + \dots + u_n\sigma_n v_n^T$$

The sum of n rank 1 pieces

Example of Singular Value Decomposition

Find the matrices U,$\sum$,V for $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$

With rank 2, this A has two positive singular values σ_1 and σ_2. Begin with A^TA and AA^T:

$$A^TA = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \qquad AA^T = \begin{bmatrix} 9 & 12 \\ 12 & 41 \end{bmatrix}$$

Those have the same trace (50) and the same eigenvalues $\sigma_1^2 = 45$ and $\sigma_2^2 = 5$. The square roots are $\sigma_1 = \sqrt{45}$ and $\sigma_2 = \sqrt{5}$. Then $\sigma_1\sigma_2 = 15$ and this is the determinant of A.
A key step is to find the eigenvectors of A^TA (with eigenvalues 45 and 5):

$$\begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 45 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 5 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Then v_1 and v_2 are those orthogonal eigenvectors rescaled to length 1. Divide by $\sqrt{2}$.

Right singular vectors $v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

Continued...

Left singular vectors $u_i = \frac{Av_i}{\sigma_i}$

Now compute Av_1 and Av_2 which will be $\sigma_1 u_1 = \sqrt{45} u_1$ and $\sigma_2 u_2 = \sqrt{5} u_2$

$$Av_1 = \frac{3}{\sqrt{2}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \sqrt{45}\, \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \sigma_1 u_1$$

$$Av_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \sqrt{5}\, \frac{1}{\sqrt{10}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \sigma_2 u_2$$

The division by $\sqrt{10}$ makes u_1 and u_2 orthonormal. Then $\sigma_1 = \sqrt{45}$ and $= \sigma_2 = \sqrt{5}$ as expected. The Singular Value Decomposition of A is U times $\sum$ times V^T.

$$\boxed{U = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix} \quad \sum = \begin{bmatrix} \sqrt{45} & \\ & \sqrt{5} \end{bmatrix} \quad V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}$$

The matrix $A = U \sum V^T$ splits into two rank one matrices, columns times rows, with $\sqrt{2}\sqrt{10} = \sqrt{20}$.

$$\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T = \frac{\sqrt{45}}{\sqrt{20}} \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} + \frac{\sqrt{5}}{\sqrt{20}} \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = A$$

Every matrix is a sum of rank one matrices with orthogonal u′s and orthogonal v′s.

Image Compression

Image compression is a process applied to a graphics file to minimize its size in bytes without degrading image quality below an acceptable threshold.

- Compression Ratio(C_R):

$$C_R = \frac{\text{umcompressed image file size}}{\text{compressed image file size}}$$

- Total storage for $A_k = u_1\sigma_1 v_1^T + u_2\sigma_2 v_2^T + \dots + u_k\sigma_k v_k^T$ wil be:

$$A_k = k(m+n+1)$$

Image Compression Using SVD

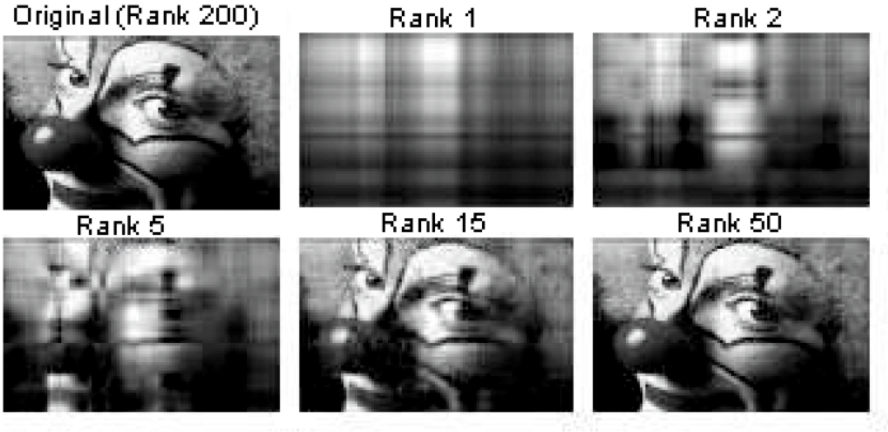

Figure 1: Image Compression

Continued...

- k represents the number of eigenvalues used in the reconstruction of the compressed image.
- Smaller the value of k, more is the compression ratio but image quality deteriorates.
- As the value of k increases, the image quality improves but more storage space is required to store the compressed image.
- When k is equal to the rank of the matrix (200 here), the reconstruction image is almost same as the original one.

Mean and Variance

- Mean is the average of the given numbers and is calculated by dividing the sum of given numbers by the total number of data points.

$$\bar{x} = \frac{\Sigma x_i}{N}$$

- Variance is a measure of how far a set of data is spread out from their mean value.

$$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{N}$$

  - A high variance indicates that the values are more spread out.
  - covariance is always ≥ 0

Covariance

- Covariance is a measure of the relationship between two random variables and to what extent, they change together.

$$\mathrm{Covariance} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{N}$$

- There are two types of Covariance.
  - Positive Covariance
  - Negative Covariance

Principle Component Analysis (PCA)

- Principal component analysis, or PCA, is a statistical technique used for dimensionality reduction and data visualization of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

- Principal Component Analysis (PCA) uses the largest σ's connected to the first u's and v's to understand the information in a matrix of data.

$$A_k = \boldsymbol{\sigma}_1 u_1 v_1^T + \ldots + \boldsymbol{\sigma}_k u_k v_k^T \qquad \text{with rank } (A_k) = k.$$

- A_k solves a matrix optimization problem. This puts the SVD at the center of data science.

- Principal Component Analysis is based on matrix approximation by A_k.

Steps For Principle Component Analysis

Step 1 Standardize the data
The aim of this step is to standardize the range of the continuous initial variables so that each one of them contributes equally to the analysis. Mathematically, this can be done by subtracting the mean for each value of each variable.

Step 2 Calculate the Covariance matrix
The aim of this step is to understand how the variables of the input data set are varying from the mean with respect to each other, or in other words, to see if there is any relationship between them.

- Covariance matrix is symmetric matrix and can be determined by :

$$\mathrm{S} = \frac{\mathrm{AA}^{\mathrm{T}}}{\mathrm{n}-1}$$

Step 3 Find the Eigenvalues and Eigenvectors of the Covariance matrix
Eigenvectors and eigenvalues are the linear algebra concepts that we need to compute from the covariance matrix in order to determine the principal components of the data.

Principle Component Analysis Example

Let's suppose that our data set is 2-dimensional with 2 variables x,y and that the eigenvectors and eigenvalues of the covariance matrix are as follows:

$$\begin{bmatrix} 0.67 \\ 0.73 \end{bmatrix} \qquad \lambda_1 = 1.28$$

$$\begin{bmatrix} -0.73 \\ 0.67 \end{bmatrix} \qquad \lambda_2 = 0.04$$

If we rank the eigenvalues in descending order, we get $\lambda_1 > \lambda_2$, which means that the eigenvector that corresponds to the first principal component (PC1) is v_1 and the one that corresponds to the second principal component (PC2) is v_2.

Continued...

After having the principal components, to compute the percentage of variance (information) accounted for by each component, we divide the eigenvalue of each component by the sum of eigenvalues. If we apply this on the example above, we find that PC1 and PC2 carry the following percent of the variance of the data :

$$\text{PC1} = \frac{1.28}{1.28+0.04} * 100 = 97\%$$

$$\text{PC2} = \frac{0.04}{1.28+0.04} * 100 = 03\%$$

Graphical Representation of 3D to 2D Reduction

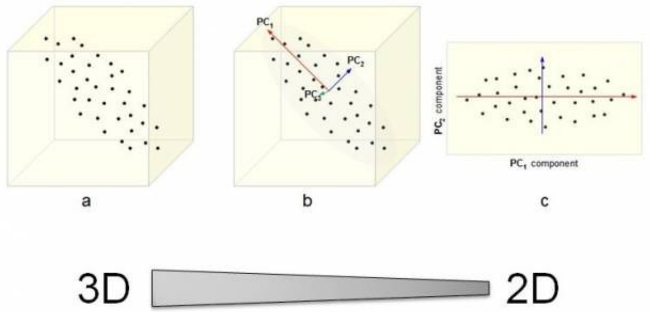

Figure 2: Reducing 3D data to 2D by PCA

Conclusions

- Singular Value Decomposition allow us to arrange the portions of the matrix in order of importance. The most important singular values will produce the most important unit eigenvectors.
- We can loose large portions of our matrix without losing quality.
- SVD's applications in the world of image compression are very useful.
- Principle Component Analysis an important application of SVD helps us to reduce our problem from higher dimensions to lower dimension so we can easily analyse the data.

Thank You For Your Attention.