Introduction to Linear Algebra

WiMLDS Accra and AIMS Mathematics Bootcamp

AIMS Ghana

African Institute for Mathematical Sciences
Ghana

April 9, 2024

Overview

1. Vectors & Vector Spaces
2. Matrix operations
3. Eigenvalues & Eigenvectors
4. Principal Component Analysis (PCA)
5. References

Introduction

Linear algebra is a branch of mathematics that is universally agreed to be a prerequisite to a deeper understanding of machine learning. Although linear algebra is a large field with many esoteric theories and findings. In this chapter, we will focus on just the good or relevant part that will be very useful in machine learning.

Vectors & Vector Spaces

AIMS Ghana

Vectors

- What is a vector?

Definition

- A vector is a tuple of one or more values called scalars. one can think of a vector as a list of numbers. In ML, a vector could represent a feature of a model or a column in a data set.
- Vectors are built from components which are ordinary numbers and it is always represented by a character such as v.

For example $\mathbf{v} = (v_1, v_2, v_3, ..., v_n)$.

- If each elements is in $\mathbb{R}$ then $\mathbf{v}$ is in $\mathbb{R}^n$.

Vector Arithmetic

Let $\mathbf{v} = (v_1, v_2, v_3, ..., v_n)$ and $\mathbf{u} = (u_1, u_2, u_3, ..., u_n)$ be two vectors and a be a scalar.

- **Addition:** $\mathbf{v} + \mathbf{u} = (v_1 + u_1, v_2 + u_2, v_3 + u_3, ..., v_n + u_n)$
- **Multiplication:** $\mathbf{v} * \mathbf{u} = (v_1 u_1, v_2 u_2, v_3 u_3, ..., v_n u_n)$.
- **Vector scalar multiplication:** $a * \mathbf{v} = (av_1, av_2, av_3 ..., av_n)$.
- **Component-wise Division:** $\mathbf{v}/\mathbf{u} = (v_1/u_1, v_2/u_2, v_3/u_3, ..., v_n/u_n)$.
- **Vector dot product:** $\mathbf{v}.\mathbf{u} = v_1 u_1 + v_2 u_2 + v_3 u_3 + ... + v_n + u_n$.

Vector Norm

In linear algebra, we have different way to compute the vector norm.
Let $\mathbf{v} = (v_1, v_2, v_3, ..., v_n)$ be a vector.

- **Vector L^1 Norm:** $||\mathbf{v}||_1 = |v_1| + |v_2| + |v_3| + ... + |v_n|$
- **Vector L^2 Norm:** $||\mathbf{v}||_2 = \sqrt{v_1^2 + v_2^2 + v_3^2 + ... + v_n^2}$
- **Vector Max Norm:** Max $(v_1, v_2, v_3, ..., v_n)$

Matrix operations

AIMS Ghana

Matrix definition

- A matrix is a two-dimensional array of scalars with one or more columns and one or more rows.

  A $n \times m$ matrix is more represented with notation

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & . & . & . & . & . & a_{1,m} \\ a_{2,1} & a_{2,2} & . & . & . & . & . & a_{2,m} \\ . & . & . & . & . & . & . & . \\ . & & . & & & & & . \\ . & & & . & & & & . \\ . & & & & & . & & . \\ a_{n,1} & a_{n,2} & . & . & . & . & . & a_{n,m} \end{pmatrix}$$

- So each element identified by two indices, such as $a_{i,j}$
- If A has shape of height m and width n with real-values then $A \in \mathbb{R}^{n \times m}$

Matrix arithmetic

Let $A = (a_{i,j})_{1\leq i\leq n, 1\leq j\leq m}$ and $B = (b_{i,j})_{1\leq i\leq n, 1\leq j\leq m}$ be two $n \times m$ matrix.

- **Addition:** $A + B = (a_{i,j} + b_{i,j})_{1\leq i\leq n, 1\leq j\leq m}$
- **Multiplication:** If $A = (a_{i,j})$ is of shape nxm and $B = (b_{i,j})$ is of shape mxp then matrix product AB is of shape nxp, where $(ab)_{i,j} = \sum_k a_{i,k} b_{k,j}$
- **Transpose:** If $A = (a_{i,j})$ is of shape nxm, then transpose of A is a matrix $A^T = (a_{j,i})$ of shape mxn.

$$A = \begin{pmatrix} 2 & 4 & 7 & 6 \\ 5 & 1 & 6 & 4 \\ 11 & 3 & 0 & 9 \end{pmatrix} \quad A^T = \begin{pmatrix} 2 & 5 & 11 \\ 4 & 1 & 3 \\ 7 & 6 & 0 \\ 6 & 4 & 9 \end{pmatrix} \tag{1}$$

Matrix arithmetic

- **Matrix vector multiplication:** Let $A = (a_{i,j})$ be a matrix of shape nxm and v a column vector of m elements. then Av is a column vector of m elements such that $(Av)_j = \sum_j a_{i,j} v_j. \quad \forall i.$
- **Matrix determinant:**Given a matrix $A = (a_{i,j})$ of shape nxn, the determinant of A noted det$A$ or $|A|$ is compute as
  det$A$=$a_{1,1} detA_{1,1} + a_{1,2} detA_{1,2} + ... + (-1)^{1+n} a_{1,n} detA_{1,n} = \sum_j^n (-1)^{1+j} a_{1,j} det(A_{1,j}$

  Exercise: Find the determinant of A or $|A|$

$$A = \begin{pmatrix} 2 & 4 & 7 \\ 5 & 1 & 6 \\ 11 & 3 & 9 \end{pmatrix}$$

- **Matrix inverse:** Given a matrix A, find matrix B, such that $AB = I_n$ or $BA = I_n$.

Eigenvalues & Eigenvectors

AIMS Ghana

Eigenvalues

- This is an important property of matrices which has numerous applications, particularly in **Machine Learning**.
- For a given matrix **A**, there exists a vector **V** and a scalar λ such that:

$$\mathbf{AV} = \lambda \mathbf{V}$$

, where λ is called **eigenvalue** of **A** and **V** is called **eigenvector** of **A** associated with λ (**V**$\neq$**zero-vector**).

Note: - In this equation, λ and **V** are **unknown**.

- The matrix **A** should be a square matrix ($n \times n$ matrix).

Eigenvalues

- To solve the equation above we need first to find the **eigenvalues**.
- **Eigenvalues** are roots of the so-called **Characteristic Polynomial**:

$$\det(A - \lambda I_n) = 0 \tag{2}$$

, where $\mathbf{I_n}$ is the square identity matrix in the space of your matrix A.

Examples: - If n=2, $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

- If n=3, $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Exercise: Given $\mathbf{AV} = \lambda \mathbf{V}$, prove Eq (2).

Hint: You may use the fact that **V** should be a non-zero vector.

Eigenvectors

- **Eigenvectors** are obtained, since we have the eigenvalues, by solving:

$$\mathbf{AV} = \lambda \mathbf{V}$$

for each value of λ found by solving (1).

Example: Find the eigenvalues and the eigenvectors of the matrix A given by:

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Examples

- **<u>Eigenvalues:</u>**
  we have:

$$\det(A - \lambda I_2) = 0 \implies \mathit{det}\left(\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) = 0$$

$$\implies \mathit{det}\left(\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right) = 0$$

$$\implies \mathit{det}\begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = 0$$

$$\implies (2-\lambda)^2 - 1^2 = 0$$

$$\implies (1-\lambda)(3-\lambda) = 0$$

The eigenvalues are therefore λ_1=1 and λ_2=3.

Examples

- **Eigenvectors:**

  Let $V = \begin{pmatrix} a \\ b \end{pmatrix}$ be a vector in $\mathbb{R}^2$.

  - For λ=1, V is a eigenvector of A iff:

$$AV = V \Longrightarrow \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

$$\Longrightarrow \begin{cases} 2a + b = a \\ a + 2b = b \end{cases}$$

$$\Longrightarrow a = -b$$

Hence, taking **a=1** leads to $V_{\lambda=1} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Principal Component Analysis (PCA)

AIMS Ghana

PCA

Definition

- PCA is a dimensionality reduction technique used in machine learning and data analysis.
- It simplifies complex datasets by identifying patterns and relationships among variables.

PCA

Definition

- PCA is a dimensionality reduction technique used in machine learning and data analysis.
- It simplifies complex datasets by identifying patterns and relationships among variables.

Purpose in Machine Learning:

- Reduces dimensionality while preserving most of the variance in the data.
- Improves computational efficiency and reduces overfitting in machine learning models.

Illustration

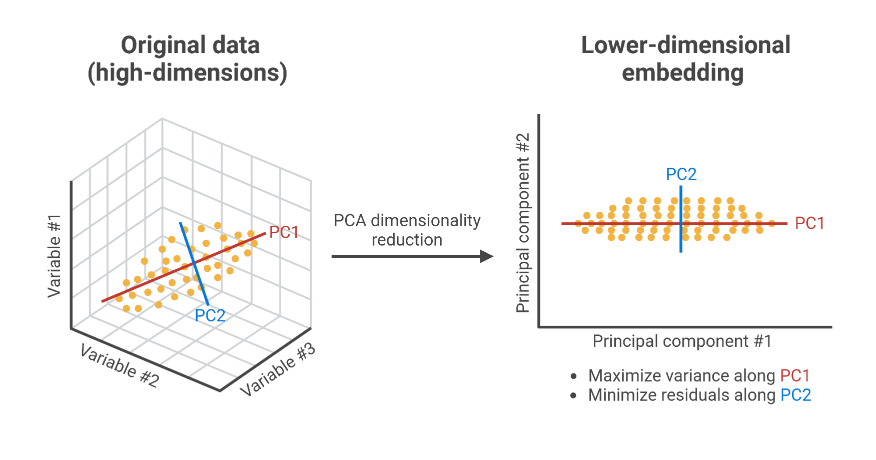

Figure: Principal Component Analysis (PCA) Transformation - Mina Nashed *BIO RENDER*

Step 1 – Standardization

- Standardize the range of initial variables for equal contribution.

$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}}$$

Suppose we have a dataset with two variables, x and y, and the following data points:

$$x : 3, 8, 5, 4$$

$$y : 6, 7, 2, 9$$

$$X = \begin{pmatrix} 3 & 6 \\ 8 & 7 \\ 5 & 2 \\ 4 & 9 \end{pmatrix}$$

We first calculate the mean and standard deviation of each variable:

$$\text{Mean of } x: \bar{x} = \frac{3+8+5+4}{4} = 5$$
$$\text{Mean of } y: \bar{y} = \frac{6+7+2+9}{4} = 6$$
$$\text{Standard deviation of } x: \sigma_x = \sqrt{\frac{(3-5)^2+(8-5)^2+(5-5)^2+(4-5)^2}{4}}$$
$$\text{Standard deviation of } y: \sigma_y = \sqrt{\frac{(6-6)^2+(7-6)^2+(2-6)^2+(9-6)^2}{4}}$$

After calculating the means and standard deviations, we standardize the data:

$$z_x = \frac{x-\bar{x}}{\sigma_x}$$
$$z_y = \frac{y-\bar{y}}{\sigma_y}$$

Step 2 - Covariance Matrix Computation

- Understand variable relationships through covariance.
- Identify correlations and redundant information.
- Covariance Matrix: Symmetric matrix with covariances of variable pairs.

$$\Sigma = \frac{1}{n-1}(X^T \cdot X)$$

$$\Sigma = \begin{pmatrix} \mathrm{cov}(z_x, z_x) & \mathrm{cov}(z_x, z_y) \\ \mathrm{cov}(z_y, z_x) & \mathrm{cov}(z_y, z_y) \end{pmatrix}$$

We calculate the covariance between z_x and z_y as well as the variances of z_x and z_y.

Step 3 - Compute Eigenvectors and Eigenvalues

- Determine principal components for variance capture.
- Eigenvectors and eigenvalues come in pairs, equal to data dimensions.
- Eigenvectors are axes with most variance, eigenvalues indicate variance amount.
- Rank eigenvectors by eigenvalues to obtain principal components.

We solve the equation $|\Sigma - \lambda I| = 0$ to find the eigenvalues, and then substitute the eigenvalues back into the equation $(\Sigma - \lambda I)\mathbf{v} = 0$ to find the corresponding eigenvectors.

Step 4 - Create Feature Vector

We form the feature vector by choosing the eigenvectors corresponding to the largest eigenvalues. This feature vector will be used to transform the data into a lower-dimensional space.

Step 5 - Recast the Data Along the Principal Components Axes

Finally, we use the feature vector to transform the original data onto the new axes defined by the principal components. This gives us the final PCA-transformed dataset.

Conclusion: Importance of Linear Algebra in Machine Learning

- Vectors, vector spaces, and matrix operations are fundamental mathematical concepts in machine learning.
- Vectors represent data points, features, and parameters, while vector spaces provide a framework for working with them.
- Matrix operations facilitate various computations, transformations, and optimizations in machine learning algorithms such as PCA.

References

References

Nicholson, W. K. (2020).
Linear algebra with applications.
Lyryx

Srihari, S. N. (n.d.).
Linear Algebra for Machine Learning [Slides]
University of Buffalo

Jaadi, Z. (2021).
A Step-by-Step Explanation of Principal Component Analysis (PCA)
Built In URL

Jolliffe, I. T. (2003).
Principal component analysis.
Technometrics (Vol. 2)

Bishop, C. M. (2006).
Pattern recognition and machine learning.
Springer google schola,(Vol. 2)

Thank you for your attention

AIMS Ghana

African Institute for Mathematical Sciences
Ghana

April 9, 2024