Linear Algebra for ML

1 Foundations of Linear Algebra

1.1 Vectors

Vectors are the fundamental objects of linear algebra. You can think of them as points in a space or as arrows with direction and magnitude. The operations of addition and scalar multiplication are the two basic actions we can perform on them

1. Let $v = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$ and $w = \begin{bmatrix} 0 \\ 4 \\ -2 \end{bmatrix}$. Compute the following:
   (a) $v + w$
   (b) $4v$
   (c) $v - 2w$

Solution

Vector addition/subtraction and scalar multiplication are performed element-wise.
(a)

$$\begin{aligned} v + w &= \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 4 \\ -2 \end{bmatrix} \\ &= \begin{bmatrix} 3+0 \\ -1+4 \\ 2-2 \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix} \end{aligned}$$

(b)

$$\begin{aligned} 4v &= 4\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} 4 \cdot 3 \\ 4 \cdot (-1) \\ 4 \cdot 2 \end{bmatrix} \\ &= \begin{bmatrix} 12 \\ -4 \\ 8 \end{bmatrix} \end{aligned}$$

(c)

$$\begin{aligned} v - 2w &= \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} - 2\begin{bmatrix} 0 \\ 4 \\ -2 \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} - \begin{bmatrix} 0 \\ 8 \\ -4 \end{bmatrix} \\ &= \begin{bmatrix} 3-0 \\ -1-8 \\ 2-(-4) \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ -9 \\ 6 \end{bmatrix} \end{aligned}$$

1.2 Vector Spaces and Subspaces

A vector space is a collection of vectors that is “closed” under addition and scalar multiplication. A subspace is simply a vector space that lives inside a larger one, but it must still obey the same rules, crucially including the requirement that it contains the zero vector.

1. Consider the set $\mathcal{U}$ of all vectors in $\mathbb{R}^3$ of the form $\begin{bmatrix} a \\ b \\ a+b \end{bmatrix}$, where $a, b \in \mathbb{R}$. Is $\mathcal{U}$ a subspace of $\mathbb{R}^3$? Justify your answer by checking the three required properties.

Solution

To verify if $\mathcal{U}$ is a subspace of $\mathbb{R}^3$, we must check the three properties:

1. **Contains the zero vector:** Does $0 \in \mathcal{U}$? The zero vector in $\mathbb{R}^3$ is $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. We can obtain this vector by setting $a = 0$ and $b = 0$. Since $\begin{bmatrix} 0 \\ 0 \\ 0+0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, the zero vector is in $\mathcal{U}$. This property holds.

2. **Closed under vector addition:** Let $v_1 = \begin{bmatrix} a_1 \\ b_1 \\ a_1+b_1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} a_2 \\ b_2 \\ a_2+b_2 \end{bmatrix}$ be two arbitrary vectors in $\mathcal{U}$. Their sum is

$$v_1 + v_2 = \begin{bmatrix} a_1+a_2 \\ b_1+b_2 \\ (a_1+b_1)+(a_2+b_2) \end{bmatrix} = \begin{bmatrix} a_1+a_2 \\ b_1+b_2 \\ (a_1+a_2)+(b_1+b_2) \end{bmatrix}$$

Let $a' = a_1 + a_2$ and $b' = b_1 + b_2$. The resulting vector is $\begin{bmatrix} a' \\ b' \\ a'+b' \end{bmatrix}$, which is of the form required to be in $\mathcal{U}$. This property holds.

3. **Closed under scalar multiplication:** Let $v = \begin{bmatrix} a \\ b \\ a+b \end{bmatrix}$ be a vector in $\mathcal{U}$ and let $\gamma \in \mathbb{R}$ be a scalar. The product is:

$$\gamma v = \gamma \begin{bmatrix} a \\ b \\ a+b \end{bmatrix} = \begin{bmatrix} \gamma a \\ \gamma b \\ \gamma(a+b) \end{bmatrix} = \begin{bmatrix} \gamma a \\ \gamma b \\ \gamma a+\gamma b \end{bmatrix}$$

Let $a' = \gamma a$ and $b' = \gamma b$. The resulting vector is $\begin{bmatrix} a' \\ b' \\ a'+b' \end{bmatrix}$, which is also of the form required to be in

$\mathcal{U}$. This property holds.
Since all three properties are satisfied, $\mathcal{U}$ is a subspace of $\mathbb{R}^3$.

1.3 Linear Independence, Span, Basis, and Dimension

These four concepts are deeply intertwined. The span of a set of vectors is all the points you can "reach" using those vectors. A set of vectors is linearly independent if none of them are redundant. A **basis** is a set of linearly independent vectors that spans the entire space. The dimension is simply the number of vectors in any basis for that space.

1. Consider the vectors $v_1 = \begin{bmatrix}1\\0\\1\end{bmatrix}$, $v_2 = \begin{bmatrix}0\\1\\1\end{bmatrix}$, and $v_3 = \begin{bmatrix}1\\1\\2\end{bmatrix}$ in $\mathbb{R}^3$.
   (a) Are these vectors linearly independent? Justify your answer.
   (b) What is the dimension of the subspace spanned by $\{v_1, v_2, v_3\}$?

Solution

(a) To check for linear independence, we need to see if the only solution to $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = 0$ is $\alpha_1 = \alpha_2 = \alpha_3 = 0$. The intuition is that no vector can be written as a combination of the others.
Let's observe the relationship between the vectors. We can see that $v_3 = v_1 + v_2$:

$$\begin{bmatrix}1\\1\\2\end{bmatrix} = \begin{bmatrix}1\\0\\1\end{bmatrix} + \begin{bmatrix}0\\1\\1\end{bmatrix}$$

Since v_3 is a linear combination of v_1 and v_2, the set is linearly dependent. We can write the zero vector as a non-trivial combination: $1 \cdot v_1 + 1 \cdot v_2 - 1 \cdot v_3 = 0$.

(b) The span of a set of vectors is the set of all their possible linear combinations. Since v_3 is redundant (it's already in the span of v_1 and v_2), the subspace spanned by $\{v_1, v_2, v_3\}$ is the same as the subspace spanned by just $\{v_1, v_2\}$. The vectors v_1 and v_2 are linearly independent because neither is a scalar multiple of the other. Therefore, $\{v_1, v_2\}$ forms a basis for the subspace. The dimension of a vector space is the number of vectors in its basis. Since the basis has two vectors, the dimension of the subspace spanned by $\{v_1, v_2, v_3\}$ is 2.

2 Matrices and Inner Products

2.1 Matrix Operations

Matrices are powerful objects for representing data and transformations. Just like with vectors, we can perform basic operations like addition and scalar multiplication. Matrix multiplication is more complex but is fundamental to composing linear transformations. **Remember the key rule: for AB to be defined, the number of columns in A must equal the number of rows in B.**

1. Let $A = \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ 3 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 1 \\ -2 & 2 \end{bmatrix}$.

   (a) What are the dimensions of A and B?

   (b) Compute $A^\top$.

   (c) Compute the matrix product AB.

   (d) Can you compute BA? Why or why not?

Solution

(a) The dimension of a matrix is its number of rows and columns. A has 3 rows and 2 columns, so its dimensions are 3×2. B has 2 rows and 2 columns, so its dimensions are 2×2.

(b) The transpose of a matrix is found by swapping its rows and columns, so $(A^\top)_{ij} = A_{ji}$.

$$A^\top = \begin{bmatrix} 1 & 0 & 3 \\ 2 & -1 & 0 \end{bmatrix}$$

(c) The product of an $m \times n$ matrix and an $n \times p$ matrix is an $m \times p$ matrix. Here, A is 3×2 and B is 2×2, so the product AB will be a 3×2 matrix. The element $(AB)_{ij}$ is the inner product of the i-th row of A and the j-th column of B.

$$AB = \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ -2 & 2 \end{bmatrix} = \begin{bmatrix} (1)(4) + (2)(-2) & (1)(1) + (2)(2) \\ (0)(4) + (-1)(-2) & (0)(1) + (-1)(2) \\ (3)(4) + (0)(-2) & (3)(1) + (0)(2) \end{bmatrix} = \begin{bmatrix} 0 & 5 \\ 2 & -2 \\ 12 & 3 \end{bmatrix}$$

(d) To compute BA, the number of columns in B (2) must equal the number of rows in A (3). Since $2 \neq 3$, the product BA is undefined. This illustrates that matrix multiplication is not commutative, i.e., in general, $AB \neq BA$.

2.2 Inner Products and Norms

The inner product (or dot product) measures the "alignment" of two vectors. It's positive if they point in similar directions, zero if they're orthogonal, and negative if they point in opposite directions. A norm, on the other hand, is a generalization of the concept of "length" or "magnitude" for a vector.

1. Let $x = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$ and $y = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$.

   (a) Compute the inner product $x^\top y$. What does the result imply about the geometric relationship between x and y?

   (b) Compute the ℓ^1-norm, ℓ^2-norm, and ℓ^∞-norm of the vector x.

Solution

(a) The inner product is $x^\top y = \sum_{j=1}^{3} x_j y_j$.

$$x^\top y = (1)(4) + (-2)(2) + (3)(0) = 4 - 4 + 0 = 0$$

Since the inner product is 0, the vectors x and y are orthogonal. Geometrically, this means they are perpendicular to each other.

(b)
- The ℓ^1-norm is the sum of the absolute values of the elements: $\|x\|_1 = \sum |x_j| = |1| + |-2| + |3| = 1 + 2 + 3 = 6$.
- The ℓ^2-norm (Euclidean norm) is the square root of the sum of the squared elements: $\|x\|_2 = \sqrt{\sum_{j=1}^{3} x_j^2} = \sqrt{1^2 + (-2)^2 + 3^2} = \sqrt{1 + 4 + 9} = \sqrt{14}$.
- The ℓ^∞-norm is the maximum absolute value of the elements: $\|x\|_\infty = \max_{1 \le j \le 3} |x_j| = \max(|1|, |-2|, |3|) = 3$.

3 The Four Fundamental Subspaces

Every matrix A defines four fundamental subspaces that tell the complete story of the linear transformation it represents. The column space is the "output space" which is the set of all possible vectors you can get from Ax. The nullspace contains all input vectors x that get mapped to zero. The row space and left nullspace are just the column space and nullspace of $A^\top$. The Rank-Nullity Theorem provides a connection between the dimensions of the column space and the nullspace.

1. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$.
   (a) Find a basis for the column space, $\mathrm{col}(A)$, and determine its dimension (the rank of A).
   (b) Find a basis for the nullspace, $\mathrm{null}(A)$, and determine its dimension.
   (c) Verify that the Rank-Nullity Theorem holds for this matrix.

Solution

(a) The column space is the span of the columns of A. The columns are $a_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $a_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, and $a_3 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$. We can see that $a_2 = 2a_1$ and $a_3 = 3a_1$. Thus, the second and third columns are linearly dependent on the first. A basis for the column space consists of the linearly independent columns. A basis for $\mathrm{col}(A)$ is $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$. The dimension of the column space is the rank of the matrix. Since the basis has one vector, $\mathrm{rank}(A) = \dim(\mathrm{col}(A)) = 1$.

(b) The nullspace consists of all vectors x such that $Ax = 0$. We need to solve the system:

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

This gives us the single equation $x_1 + 2x_2 + 3x_3 = 0$. We can express x_1 in terms of x_2 and x_3, which are free variables. Let $x_2 = \alpha$ and $x_3 = \beta$. Then $x_1 = -2\alpha - 3\beta$. The vectors in the nullspace look like:

$$x = \begin{bmatrix} -2\alpha - 3\beta \\ \alpha \\ \beta \end{bmatrix} = \alpha \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

The two vectors span the nullspace and are linearly independent. A basis for $\mathrm{null}(A)$ is $\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}$. The dimension of the nullspace is the number of vectors in its basis, so $\dim(\mathrm{null}(A)) = 2$.

(c) The Rank-Nullity Theorem states that for an $m \times n$ matrix, $\mathrm{rank}(A) + \dim(\mathrm{null}(A)) = n$. Here, A is a 2×3 matrix, so $n = 3$. We found $\mathrm{rank}(A) = 1$ and $\dim(\mathrm{null}(A)) = 2$. Plugging these values in: $1 + 2 = 3$. The theorem holds.

4 Least Squares

Often, a system of equations $Ax = y$ has no solution. This happens when the vector y is not in the column space of A. We can't find an x that works perfectly, but we can find the next best thing: an x^* that makes Ax^* as close as possible to y. This is the essence of the least squares problem. The solution involves projecting y onto the column space of A.

1. Consider the system of linear equations:

$$\begin{aligned} x_1 + x_2 &= 4 \\ -x_1 + x_2 &= 0 \\ x_2 &= 3 \end{aligned}$$

   (a) Write this system in the form $Ax = y$.

   (b) Show that this system has no solution.

   (c) Find the least squares solution $x^* = (A^\top A)^{-1} A^\top y$. You are given that $(A^\top A)^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$.

Solution

(a) The system can be written in matrix form as:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 4 \\ 0 \\ 3 \end{bmatrix}}_{y}$$

(b) From the last equation, we must have $x_2 = 3$. Substituting this into the second equation gives $-x_1 + 3 = 0 \implies x_1 = 3$. Substituting both into the first equation gives $3 + 3 = 6$, but the equation requires the sum to be 4. Since $6 \neq 4$, there is a contradiction, and the system has no solution. Geometrically, the vector y is not in the column space of A.

(c) We need to find the least squares solution that minimizes $\|Ax - y\|_2^2$. The formula is $x^* = (A^\top A)^{-1} A^\top y$. First, let's compute $A^\top y$:

$$A^\top y = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} (1)(4) + (-1)(0) + (0)(3) \\ (1)(4) + (1)(0) + (1)(3) \end{bmatrix} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$$

Finally, we assemble the solution:

$$x^* = (A^\top A)^{-1} A^\top y = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \end{bmatrix} = \begin{bmatrix} (1/2)(4) + (0)(7) \\ (0)(4) + (1/3)(7) \end{bmatrix} = \begin{bmatrix} 2 \\ 7/3 \end{bmatrix}$$

The least squares solution is $x^* = \begin{bmatrix} 2 \\ 7/3 \end{bmatrix}$.

5 Eigenvalues and Eigenvectors

For a given transformation matrix A, eigenvectors are special vectors that do not change direction when A is applied to them; they are only scaled. The scaling factor is the eigenvalue λ. Eigenvectors and eigenvalues reveal the intrinsic ”axes” of a linear transformation and are crucial for understanding matrix powers, diagonalization, and many algorithms like PCA (which you'll learn about in the Classical ML lecture).

1. Let $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ with eigenvalues $\lambda_1 = 3, \lambda_2 = 4$ and eigenvectors $v_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
   (a) Define what an eigenvalue and its corresponding eigenvector are for a square matrix A.
   (b) Since A is symmetric, what special properties do its eigenvalues and eigenvectors have according to the Spectral Theorem? Verify these properties.

Solution

(a) An eigenvector of a square matrix A is a non-zero vector x such that for some scalar λ, the equation $Ax = \lambda x$ holds. The scalar λ is called the eigenvalue corresponding to the eigenvector x.
(b) The matrix A is symmetric because $A = A^\top$.The Spectral Theorem states that for a symmetric matrix:
- The eigenvalues are always real. (Our eigenvalues are 2 and 4, which are real).
- Eigenspaces corresponding to different eigenvalues are orthogonal. We can check this with our eigenvectors: $v_1^\top v_2 = (-1)(1) + (1)(1) = -1 + 1 = 0$. They are indeed orthogonal.
- The matrix is orthonormally diagonalizable. This means there exists an orthonormal matrix U whose columns are the eigenvectors such that $A = U\Lambda U^\top$.

6 Singular Value Decomposition (SVD)

SVD is arguably the most important matrix decomposition. It extends the ideas of diagonalization to *any* matrix, not just square ones. It decomposes a matrix A into three simpler matrices: a rotation (U), a scaling (Σ), and another rotation ($V^\top$). Its most powerful application is finding the best low-rank approximation of a matrix, which is the key to data compression and noise reduction.

1. A full SVD of a matrix $A \in \mathbb{R}^{m\times n}$ with rank r is given by $A = U\Sigma V^\top$.
   (a) Describe the properties of the matrices U, Σ, and V. What are their dimensions?
   (b) How are the singular values in Σ related to the eigenvalues of $A^\top A$ and $AA^\top$?
   (c) The dyadic (or outer product) form of the SVD is $A = \sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^\top$. Explain how you would use this form to find the best rank-1 approximation of A. Which term in the sum would you use and why?

Solution

(a) The matrices in the SVD decomposition $A = U\Sigma V^\top$ for an $m \times n$ matrix A have the following properties:
- U: An $m \times m$ orthonormal matrix. Its columns (u_i) are called the left singular vectors.
- Σ: An $m \times n$ rectangular diagonal matrix. Its diagonal entries σ_i are the singular values of A, and they are non-negative and ordered such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. All other entries are zero.
- V: An $n \times n$ orthonormal matrix. Its columns (v_i) are called the right singular vectors. Note that the decomposition uses its transpose, $V^\top$.

(b) The singular values squared (σ_i^2) are the non-zero eigenvalues of both $A^\top A$ and $AA^\top$.
- The columns of V (right singular vectors) are the orthonormal eigenvectors of $A^\top A$.
- The columns of U (left singular vectors) are the orthonormal eigenvectors of $AA^\top$.

(c) The Eckart-Young-Mirsky theorem states that the best rank-k approximation to a matrix A (in terms of both Frobenius and spectral norm) is found by truncating the SVD sum after the k-th term. The terms in the sum are ordered by the magnitude of the singular values ($\sigma_1 \geq \sigma_2 \geq \ldots$). The first term, $\sigma_1 u_1 v_1^\top$, corresponds to the largest singular value and thus captures the most significant "component" or "direction" of the data in the matrix. To find the best rank-1 approximation of A, denoted A_1, you would use only the first term of the sum:

$$A_1 = \sigma_1 u_1 v_1^\top$$

This matrix has rank 1 and minimizes the distances $\|A - B\|_2$ and $\|A - B\|_F$ over all possible rank-1 matrices B.

7 Introduction to Python (again) and NumPy

In this problem, you will work through a Python + NumPy tutorial (link). NumPy is a foundational library that you will use extensively during the ML portion of CRUP, so be sure to take your time and develop a solid understanding of it.

Solution

See link.

8 What to Submit

Please submit a single .zip file that includes your written answers and completed notebook. Upload the .zip file to the corresponding homework assignment on Gradescope.

Contributors

- Patrick Mendoza