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) $v - 2w$

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.

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 $\{v_1, v_2, v_3\}$?

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?

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.

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.

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

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.

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?

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.

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