



Deep Learning

CS60010

Abir Das

Computer Science and Engineering Department
Indian Institute of Technology Kharagpur

<http://cse.iitkgp.ac.in/~adas/>



Agenda

- To brush up basics of Linear Algebra.



Resources

- "Deep Learning", I. Goodfellow, Y. Bengio, A. Courville. (Chapter 2)



Scalars, Vectors, Matrices and Tensors

- **Scalars:** They are single numbers. Denoted mostly as lowercase variable names.
 - x, y, z
- **Vectors:** Vectors are array of numbers. Typically denoted as boldface lowercase variable names. Individual components are treated as scalars.
 - $\mathbf{x} = [x_1, x_2, \dots, x_d]^T, \mathbf{y} = [y_1, y_2, \dots, y_d]^T, \mathbf{z} = [z_1, z_2, \dots, z_d]^T$
- **Matrices:** Matrices are 2-D array of numbers. Typically denoted as boldface uppercase variable names.
 - $\mathbf{A} = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} B_{1,1} & \cdots & B_{1,n} \\ \vdots & \ddots & \vdots \\ B_{m,1} & \cdots & B_{m,n} \end{bmatrix}$
- **Tensors:** Arrays with more than 2 dimensions are generally called Tensors.



Matrix Operations

- **Transpose:** Transpose of a matrix is the mirror image of the matrix across the diagonal line, called the main diagonal of the matrix.
 - The transpose of a matrix A is denoted as A^T , where $(A^T)_{i,j} = A_{j,i}$
- **Addition:** Matrices can be added as long as they have the same shape, by adding their corresponding elements.
 - $C = A + B$, where $C_{i,j} = A_{i,j} + B_{i,j}$
- **Multiplication:** In order for the product of the two matrices A and B to be defined, A must have the same number of columns as that of the rows of B . If A is of shape $m \times n$ and B is of shape $n \times p$ then C is of shape $m \times p$, the product operation $C = AB$ is defined by,

$$C_{i,j} = \sum_k A_{i,k} B_{k,j}$$

Matrix Operations

- **Elementwise or Hadamard Product:** It's a matrix containing the product of the individual elements. It is denoted as $A \odot B$
- **Dot Product:** The dot product between two vectors x and y of the same dimensionality is the matrix product $x^T y$
- Matrix product is not commutative ($AB = BA$ does not always hold). However, the dot product between two vectors is commutative, i.e., $x^T y = y^T x$
- Let us consider a system of linear equations as follows,
$$\begin{aligned}A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,n}x_n &= b_1 \\A_{2,1}x_1 + A_{2,2}x_2 + \cdots + A_{2,n}x_n &= b_2 \\&\vdots \\A_{m,1}x_1 + A_{m,2}x_2 + \cdots + A_{m,n}x_n &= b_m\end{aligned}$$



System of Linear Equations

- $$\begin{aligned} A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,n}x_n &= b_1 \\ A_{2,1}x_1 + A_{2,2}x_2 + \cdots + A_{2,n}x_n &= b_2 \\ &\vdots \\ A_{m,1}x_1 + A_{m,2}x_2 + \cdots + A_{m,n}x_n &= b_m \end{aligned}$$

- We can write these as,

- $$\begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

- $$Ax = b$$



System of Linear Equations

$$\mathbf{Ax} = \mathbf{b}$$

- $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$
- There can be 3 possibilities
 - $m = n$ and $\det(\mathbf{A}) \neq 0$, the solution is unique, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. When is $\det(\mathbf{A}) = 0$?
 - $m < n$ - underdetermined problem (No. of equations < No. of variables). Infinitely many solutions. What can be a meaningful solution?
 - $m > n$ - overdetermined problem (No. of equations > No. of variables). No solution. What can be a meaningful solution?
 - We need to be familiar with the concept of norms for this.

Eigenvalues and Eigenvectors

- Suppose A is a matrix. The question is – does there exist any vector x for A so that the operation Ax gives a vector which is nothing but a stretched (and not rotated) version of the vector x ? i.e., $Ax = \lambda x$, or $(\lambda I - A)x = 0$.
- For non-trivial solution $\det(\lambda I - A) = |\lambda I - A| = 0$
- If $A \in \mathbb{R}^{n \times n}$, then $|\lambda I - A| = 0$ will be a n^{th} order equation. That means you can have n solutions of λ - such λ 's are called eigenvalues (real or complex conjugate). The corresponding vector x 's are the eigenvectors.
- Remember that eigenvectors are not unique. This is because if x is an eigenvector, then ax is also the same eigenvector (as it satisfies $Aax = \lambda ax$). So, we are satisfied with the direction of the eigenvectors only.



Standard Results on Eigenvalues and Eigenvectors

- If $\lambda_1, \lambda_2, \dots, \lambda_n$ are eigenvalues of A , then for any positive integer m , $\lambda_1^m, \lambda_2^m, \dots, \lambda_n^m$ are eigenvalues of A^m . The significance is that if you have eigenvalues of A , you don't have to compute the eigenvalues of A^m .
- If A is a non-singular or invertible matrix with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, then $\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_n^{-1}$ are eigenvalues of A^{-1} .
- For triangular matrix (upper, lower or diagonal), the eigenvalues are the diagonal elements itself.
- If a square matrix $A \in \mathbb{R}^{n \times n}$, is symmetric then all its eigenvalues are real and it has n linearly independent eigenvectors. The reverse is also true – i.e., if a square matrix $A \in \mathbb{R}^{n \times n}$ has n real eigenvalues and n real orthogonal eigenvectors, then the matrix is symmetric.



Standard Results on Eigenvalues and Eigenvectors

- A matrix $A \in \mathbb{R}^{n \times n}$ is positive definite if $\forall x \neq 0 \in \mathbb{R}^n, x^T A x > 0$. It is positive semi-definite if $x^T A x \geq 0$
- A matrix $A \in \mathbb{R}^{n \times n}$ is negative definite if $\forall x \neq 0 \in \mathbb{R}^n, x^T A x < 0$. It is negative semi-definite if $x^T A x \leq 0$
- If A is positive definite, $\lambda_i > 0 \forall i$
- If A is positive semi-definite, $\lambda_i \geq 0 \forall i$
- If A is negative definite, $\lambda_i < 0 \forall i$
- If A is negative semi-definite, $\lambda_i \leq 0 \forall i$

Vector Norms

- Vector norm is a real valued function (i.e., its output is always a real number) with the following properties.
 - $||\mathbf{x}|| > 0$ and $||\mathbf{x}|| = 0$ only if $\mathbf{x} = 0$
 - $||\alpha\mathbf{x}|| = |\alpha| ||\mathbf{x}||$
 - $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}|| \rightarrow$ Triangle inequality
- L_p norm: $||\mathbf{x}||_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$
- L_0 norm: $||\mathbf{x}||_0 =$ Number of non-zero elements in \mathbf{x}
- L_∞ norm: $||\mathbf{x}||_\infty = \max_i |x_i|$



Matrix/Induced Norms

- $||A|| = \max_{x \neq 0} \frac{||Ax||}{||x||}$
- You can think of matrix norm as the multiplying capacity of the matrix.
- $||A|| > 0$ and $||A|| = 0$ only if $A = 0$
- $\alpha ||A|| = |\alpha| ||A||$
- $||A + B|| \leq ||A|| + ||B|| \rightarrow$ Triangle inequality
- $||AB|| \leq ||A|| ||B|| \rightarrow$ Additional



Orthogonality

- **Vectors:** Two vectors \mathbf{u} and $\mathbf{v} \in \mathbb{R}^n$ are **orthogonal** if $\mathbf{u}^T \mathbf{v} = 0$. We write $\mathbf{u} \perp \mathbf{v}$.
- In general, we have $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$
- However, $\mathbf{u} \perp \mathbf{v} \Rightarrow ||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$
- This is the **Pythagorean theorem**.

Orthogonality

- **Matrix:** A matrix A is **orthogonal** when it has orthogonal columns

$$A = \begin{pmatrix} | & | & \dots & | \\ \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n \\ | & | & \dots & | \end{pmatrix}$$

$$\vec{a}_i^T \cdot \vec{a}_i = 1$$

$$\vec{a}_i^T \cdot \vec{a}_j = 0, i \neq j$$

- Properties:

- $AA^T = A^T A = I$

- $A^{-1} = A^T$

- $||Au||_2^2 = \mathbf{u}^T A^T A \mathbf{u} = \mathbf{u}^T \mathbf{u} = ||\mathbf{u}||_2^2 \rightarrow \text{Length preserving.}$



System of Linear Equations

$$Ax = b$$

- $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$
- There can be 3 possibilities
 - $m = n$ and $\det(A) \neq 0$, the solution is unique, $x = A^{-1}b$. When is $\det(A) = 0$?
 - $m < n$ - underdetermined problem (No. of equations < No. of variables). Infinitely many solutions. What can be a meaningful solution?
 - minimize $J = \|x\|_2$ subject to $Ax = b \rightarrow x = A^T(AA^T)^{-1}b$
 - $m > n$ - overdetermined problem (No. of equations > No. of variables). No solution. What can be a meaningful solution?
 - Minimize $J = \|Ax - b\|_2 \rightarrow x = (A^T A)^{-1}A^T b$



Thank You!!