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.
  - $\boldsymbol{x} = [x_1, x_2, \dots, x_d]^T, \boldsymbol{y} = [y_1, y_2, \dots, y_d]^T, \boldsymbol{z} = [z_1, z_2, \dots, z_d]^T$
- **Matrices**: Matrices are 2-D array of numbers. Typically denoted as boldface uppercase variable names.
  - $\boldsymbol{A} = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} B_{1,1} & \cdots & B_{1,n} \\ \vdots & \ddots & \vdots \\ B_{m,1} & \cdots & B_{m,n} \end{bmatrix}$
- **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 $\boldsymbol{A}$ is denoted as $\boldsymbol{A}^T$, where $(\boldsymbol{A}^T)_{i,j} = \boldsymbol{A}_{j,i}$
- **Addition**: Matrices can be added as long as they have the same shape, by adding their corresponding elements.
  - $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$, where $\boldsymbol{C}_{i,j} = \boldsymbol{A}_{i,j} + \boldsymbol{B}_{i,j}$
- **Multiplication**: In order for the product of the two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ to be defined, $\boldsymbol{A}$ must have the same number of columns as that of the rows of $\boldsymbol{B}$. If $\boldsymbol{A}$ is of shape $m \times n$ and $\boldsymbol{B}$ is of shape $n \times p$ then $\boldsymbol{C}$ is of shape $m \times p$, the product operation $\boldsymbol{C} = \boldsymbol{AB}$ is defined by,

$$\boldsymbol{C}_{i,j} = \sum_k \boldsymbol{A}_{i,k} \boldsymbol{B}_{k,j}$$

Matrix Operations

- **Elementwise or Hadamard Product**: It's a matrix containing the product of the individual elements. It is denoted as $\boldsymbol{A} \odot \boldsymbol{B}$
- **Dot Product**: The dot product between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of the same dimensionality is the matrix product $\boldsymbol{x}^T \boldsymbol{y}$
- Matrix product is not commutative ($\boldsymbol{AB} = \boldsymbol{BA}$ does not always hold**).** However, the dot product between two vectors is commutative, i.e., $\boldsymbol{x}^T \boldsymbol{y} = \boldsymbol{y}^T \boldsymbol{x}$
- Let us consider a system of linear equations as follows,

$$\begin{gathered}\boldsymbol{A}_{1,1}x_1 + \boldsymbol{A}_{1,2}x_2 + \cdots + \boldsymbol{A}_{1,n}x_n = b_1 \\ \boldsymbol{A}_{2,1}x_1 + \boldsymbol{A}_{2,2}x_2 + \cdots + \boldsymbol{A}_{2,n}x_n = b_2 \\ \cdots \\ \boldsymbol{A}_{m,1}x_1 + \boldsymbol{A}_{m,2}x_2 + \cdots + \boldsymbol{A}_{m,n}x_n = b_m\end{gathered}$$

System of Linear Equations

- $$\boldsymbol{A}_{1,1}x_1 + \boldsymbol{A}_{1,2}x_2 + \cdots + \boldsymbol{A}_{1,n}x_n = b_1$$
$$\boldsymbol{A}_{2,1}x_1 + \boldsymbol{A}_{2,2}x_2 + \cdots + \boldsymbol{A}_{2,n}x_n = b_2$$
$$\cdots$$
$$\boldsymbol{A}_{m,1}x_1 + \boldsymbol{A}_{m,2}x_2 + \cdots + \boldsymbol{A}_{m,n}x_n = b_m$$
- We can write these as,
- $$\begin{bmatrix} \boldsymbol{A}_{1,1} & \boldsymbol{A}_{1,2} & \cdots & \boldsymbol{A}_{1,n} \\ \boldsymbol{A}_{2,1} & \boldsymbol{A}_{2,2} & \cdots & \boldsymbol{A}_{2,n} \\ . & . & . & . \\ \boldsymbol{A}_{m,1} & \boldsymbol{A}_{m,2} & \cdots & \boldsymbol{A}_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ . \\ b_m \end{bmatrix}$$
- $$\boldsymbol{Ax} = \boldsymbol{b}$$

System of Linear Equations

$$\boldsymbol{Ax} = \boldsymbol{b}$$

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

Eigenvalues and Eigenvectors

- Suppose $\boldsymbol{A}$ is a matrix. The question is – does there exist any vector $\boldsymbol{x}$ for $\boldsymbol{A}$ so that the operation $\boldsymbol{Ax}$ gives a vector which is nothing but a stretched (and not rotated) version of the vector $\boldsymbol{x}$? i.e., $\boldsymbol{Ax} = \lambda\boldsymbol{x}$, or $(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$.
- For non-trivial solution $\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = |\lambda\boldsymbol{I} - \boldsymbol{A}| = 0$
- If $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, then $|\lambda\boldsymbol{I} - \boldsymbol{A}| = 0$ 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 $\boldsymbol{x}$'s are the eigenvectors.
- Remember that eigenvectors are not unique. This is because if $\boldsymbol{x}$ is an eigenvector, then a$\boldsymbol{x}$ is also the same eigenvector (as it satisfies $\boldsymbol{A}a\boldsymbol{x} = \lambda a\boldsymbol{x}$). So, we are satisfied with the direction of the eigenvectors only.

Standard Results on Eigenvalues and Eigenvectors

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

Standard Results on Eigenvalues and Eigenvectors

- A matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is positive definite if $\forall \boldsymbol{x} \neq \boldsymbol{0} \in \mathbb{R}^n, \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > \boldsymbol{0}$. It is positive semi-definite if $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{0}$
- A matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is negative definite if $\forall \boldsymbol{x} \neq \boldsymbol{0} \in \mathbb{R}^n, \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} < \boldsymbol{0}$. It is negative semi-definite if $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0}$
- If $\boldsymbol{A}$ is positive definite, $\lambda_i > 0\ \forall i$
- If $\boldsymbol{A}$ is positive semi-definite, $\lambda_i \geq 0\ \forall i$
- If $\boldsymbol{A}$ is negative definite, $\lambda_i < 0\ \forall i$
- If $\boldsymbol{A}$ 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.
  - $||\boldsymbol{x}|| > 0$ and $||\boldsymbol{x}|| = 0$ only if $\boldsymbol{x} = 0$
  - $||\alpha\boldsymbol{x}|| = |\alpha|||\boldsymbol{x}||$
  - $||\boldsymbol{x}+\boldsymbol{y}|| \leq ||\boldsymbol{x}|| + ||\boldsymbol{y}||$ → Triangle inequality
- L_p norm: $||\boldsymbol{x}||_p = \left(\sum_{i=1}^{n} |\boldsymbol{x}_i|^p\right)^{\frac{1}{p}}$
- L_0 norm: $||\boldsymbol{x}||_0$ = Number of non-zero elements in $\boldsymbol{x}$
- L_∞ norm: $||\boldsymbol{x}||_\infty = \max_i |\boldsymbol{x}_i|$

Matrix/Induced Norms

- $||\boldsymbol{A}|| = \max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{||\boldsymbol{A}\boldsymbol{x}||}{||\boldsymbol{x}||}$

- You can think of matrix norm as the multiplying capacity of the matrix.

- $||\boldsymbol{A}|| > 0$ and $||\boldsymbol{A}|| = 0$ only if $\mathbf{A} = \mathbf{0}$
- $\alpha||\boldsymbol{A}|| = |\alpha|||\boldsymbol{A}||$
- $||\boldsymbol{A} + \boldsymbol{B}|| \leq ||\boldsymbol{A}|| + ||\boldsymbol{B}||$ → Triangle inequality
- $||\boldsymbol{A}\boldsymbol{B}|| \leq ||\boldsymbol{A}||||\boldsymbol{B}||$ → Additional

Orthogonality

- **Vectors**: Two vectors $\boldsymbol{u}$ and $\boldsymbol{v} \in \mathbb{R}^n$ are **orthogonal** if $\boldsymbol{u}^T\boldsymbol{v} = 0$. We write $\boldsymbol{u} \perp \boldsymbol{v}$.

- In general, we have $\left|\left|\boldsymbol{u} + \boldsymbol{v}\right|\right| \leq ||\boldsymbol{u}|| + ||\boldsymbol{v}||$

- However, $\boldsymbol{u} \perp \boldsymbol{v} \Rightarrow \left|\left|\boldsymbol{u} + \boldsymbol{v}\right|\right|^{\boldsymbol{2}} = \left|\left|\boldsymbol{u}\right|\right|^2 + \left|\left|\boldsymbol{v}\right|\right|^2$

- This is the **Pythagorean theorem**.

Orthogonality

- **Matrix**: A matrix $\boldsymbol{A}$ is **orthogonal** when it has orthogonal columns

$$A = \begin{pmatrix} | & | & & | \\ \overrightarrow{a_1} & \overrightarrow{a_2} & \ldots & \overrightarrow{a_n} \\ | & | & & | \end{pmatrix}$$

$$\overrightarrow{a_i^T}.\overrightarrow{a_i} = 1$$

$$\overrightarrow{a_i^T}.\overrightarrow{a_j} = 0, i \neq j$$

- Properties:
  - $\boldsymbol{AA^T = A^TA = I}$
  - $\boldsymbol{A^{-1} = A^T}$
  - $\left|\left|\boldsymbol{Au}\right|\right|_2^2 = \boldsymbol{u^TA^TAu = u^Tu} = \left|\left|\mathbf{u}\right|\right|_2^2$ → Length preserving.

System of Linear Equations

$$\boldsymbol{Ax} = \boldsymbol{b}$$

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

Thank You!!