# 2 - LINEAR ALGEBRA

Slides covering content from the book “Deep Learning” by
Yoshua Bengio, Ian Goodfellow and Aaron Courville

# Topics

- 2.1 Scalars, Vectors, Matrices and Tensors
- 2.2 Multiplying Matrices and Vectors
- 2.3 Identity and Inverse Matrices
- 2.4 Linear Dependence, Span, and Rank
- 2.5 Norms
- 2.6 Special Kinds of Matrices and Vectors
- 2.7 Eigen-decomposition
- 2.8 Singular Value Decomposition
- 2.9 The Moore-Penrose Pseudoinverse
- 2.10 The Trace Operator
- 2.11 Determinant
- ~~2.12 Example: Principal Components Analysis~~

# Scalars

- A scalar is just a single number.
- Written in italics with a lower-case variable name.
- Eg. "Let $s \in \mathbb{R}$ be the slope of the line,"
- Here $\mathbb{R}$ represents the set of real numbers. Similarly…

# Vectors

- A vector is an array of numbers.
- Lower case names written in bold typeface.
- If each element is in R, and the vector has n elements,
- then the vector lies in the set formed by taking the Cartesian product of R n times, denoted as $R^n$.
- A cartesian product of 2 sets is a set where every element from the 1st set is paired with every element of the second set. (ordered pair).
- We can think of each n dimensional vector as an identifying point in n dimensional space, with each element giving the coordinate along a different axis.

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

# Matrices

- A matrix is a 2-D array of numbers
- Variable Name: Upper-case + Bold
- If a real-valued matrix A has a height of m and a width of n, then we say that A $\in$ $R^{m \times n}$
- Elements of a matrix: Upper-case + Italic (instead of bold)
- For example, $A_{1,1}$ is the upper left entry of **A** and $A_{m,n}$ is the bottom right entry.
- We can identify all of the numbers of a coordinate by writing a ":".
- A vector can also be thought of as a Matrix with a single column.

# Tensors

- Vector are 1-D arrays,
- Matrices are 2-D arrays
- In general,
  n-D arrays are called Tensors.

# Transpose of a Matrix

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{bmatrix} \Rightarrow \boldsymbol{A}^\top = \begin{bmatrix} a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,2} & a_{2,2} & a_{3,2} \end{bmatrix}$$

- Mirror Image about the main diagonal (top left to bottom right).
- Or simply every row becomes a column and every column becomes a row.
- To be precise, it is defined as:

$$(\boldsymbol{A}^\top)_{i,j} = A_{j,i}$$

# Operations on Matrices

- Matrix Addition – element by element
- Scalar Multiplication – every element multiplied by the scalar
- Matrix Multiplication:
- For $C = AB$:

$$c_{i,j} = \sum_k a_{i,k} b_{k,j}.$$

- Also we have a property,

$$(\boldsymbol{AB})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$$

# System of Linear Equations

$$\begin{aligned} a_{11}x_1 &+ a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 &+ a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots & \qquad\qquad\qquad\quad \vdots & \vdots \\ a_{m1}x_1 &+ a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

- $A\,x = b$

Image Source: Introduction to Octave by Dr. P.J.G.Lon
http://www-h.eng.cam.ac.uk/help/programs/octave/tutorial/

# Identity and Inverse

- Identity Matrix, elements of main diagonal 1, rest 0. Identity $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
- Any matrix $A$ multiplied to/by the Identity results into the same matrix $A$.

and

- The matrix inverse of $A$ is denoted as $A^{-1}$, and it is defined as the matrix such that

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

# Few related concepts

- Linear Dependence
- Span
- Rank

# Norms

A way to measure the “size” of a vector or matrix.

$$||\boldsymbol{x}||_p = \left( \sum_i |x_i|^p \right)^{\frac{1}{p}}$$

Most common:
p = 2 i.e. $L^2$ norm or Euclidean norm. Represents the distance from the origin.

- $f(\boldsymbol{x}) = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{0}$
- $f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$ (the *triangle inequality*)
- $\forall \alpha \in \mathbb{R}, f(\alpha \boldsymbol{x}) = |\alpha| f(\boldsymbol{x})$

# Special Norms

$$||\boldsymbol{x}||_1 = \sum_i |x_i|.$$

Also, $L^0$ norm (does not meet criterion 3)

$$||\boldsymbol{x}||_\infty = \max_i |x_i|.$$

$$\boldsymbol{x}^\top \boldsymbol{y} = ||\boldsymbol{x}||_2 ||\boldsymbol{y}||_2 \cos\theta$$

$$||A||_F = \sqrt{\sum_{i,j} a_{i,j}^2},$$

Frobenius norm

Matrix analog for $L^2$ norm.

# Diagonal Matrix

- All elements except the main diagonal are 0.
- $A = \mathrm{diag}(v)$ denotes a matrix where:
  $a_{i,i} = v_i$ (and the remaining elements are all 0)
- Need not be square
- Computationally efficient to:
  - *Multiply by a vector/Matrix*
  - *Invert*

# More Matrices

- Symmetric Matrix:

  One where, $A = A^T$

- Often arises when generated by a function of 2 args that does not depend on the order of the args.

  Eg. Distance between two points

- Unit Vector:

  A vector with L2 norm being unity i.e. $||\boldsymbol{x}||_2 = 1$

  Same as the definition in Vector Math.

# Orthogonal Matrix

- Vectors $x$ and $y$ are orthogonal to each other if $x^T y = 0$
- If their norms are also unity, then we call the vectors Orthonormal.

- A square matrix whose rows are mutually orthonormal and who's columns are also mutually orthonormal is called an ORTHOGONAL matrix. Hence,

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}.$$

$$\boldsymbol{A}^{-1} = \boldsymbol{A}^\top,$$

# Eigen Decomposition

- Decomposition as a way of observation and understanding.
- An *eigenvector* of a square matrix A is a non-zero vector v such that multiplication by A alters only the scale of v:
- And the scaling factor or scalar lambda is known as the eigen value corresponding to this eigen vector.

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}.$$

- If the eigen vectors and values are known, we may use them to “recover” the original matrix

$$\boldsymbol{A} = \boldsymbol{V} \operatorname{diag}(\boldsymbol{\lambda}) \boldsymbol{V}^{-1}$$

# Eigen Decomposition

- Not all square matrices can be decomposed this way, and not all that can be decomposed will have real valued decompositions.

- A matrix whose eigenvalues are all positive is called positive definite. A matrix whose eigenvalues are all positive or zero-valued is called positive semidefinite.
- Positive semidefinite property:

$$\forall \boldsymbol{x},\ \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \geq 0$$

- Positive definite property:

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{0}.$$

# Singular Value Decomposition

- Eigen decomposition is limited to square matrices, so we have SCD for rectangular matrices.

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$$

- The matrices U and V are both defined to be orthogonal matrices. The matrix D is defined to be a diagonal matrix.
- The elements along the diagonal of D are known as the singular values of the matrix A. The columns of U are known as the left-singular vectors. The columns of V are known as as the right-singular vectors.

# SVD in terms of Eigen Decomposition

- We can actually interpret the singular value decomposition of A in terms of the eigendecomposition of functions of A.
- The left-singular vectors of A are the eigenvectors of $AA^T$.
- The right-singular vectors of A are the eigenvectors of $A^TA$.
- The non-zero singular values of A are the square roots of the eigenvalues of $A^TA$ or $AA^T$.

# Moore Penrose Pseudoinverse

- Inverse not defined for non-Square matrices, but suppose we want to find a matrix B corresponding to A such that,

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

$$\boldsymbol{x} = \boldsymbol{B}\boldsymbol{y}.$$

- Such a matrix may or may not exist, and if it exists, it may not necessarily be unique.
- Calculated using the formula: $\boldsymbol{A}^{+} = \boldsymbol{V}\boldsymbol{D}^{+}\boldsymbol{U}^{\top},$
- where U , D, and V are the singular value decomposition of A, and the pseudoinverse D+ of a diagonal matrix D is obtained by taking the reciprocal of its non-zero elements then taking the transpose of the resulting matrix

# Trace Operator

- Trace gives the sum of all diagonal entries of a matrix.
- It also has the following properties:

$$\mathrm{Tr}(\boldsymbol{A}) = \sum_i a_{i,i}.$$

$$||\boldsymbol{A}||_F = \sqrt{\mathrm{Tr}(\boldsymbol{A}^\top \boldsymbol{A})}.$$

$$\mathrm{Tr}(\boldsymbol{A}) = \mathrm{Tr}(\boldsymbol{A}^\top).$$

$$\mathrm{Tr}(\boldsymbol{ABC}) = \mathrm{Tr}(\boldsymbol{CAB}) = \mathrm{Tr}(\boldsymbol{BCA})$$

$$\mathrm{Tr}(\prod_{i=1}^{n} \boldsymbol{F}^{(i)}) = \mathrm{Tr}(\boldsymbol{F}^{(n)} \prod_{i=1}^{n-1} \boldsymbol{F}^{(i)}).$$

# Determinant

- It is a function mapping matrices to real scalars.
- It is equal to the product of the matrix's eigenvalues.
- The absolute value of the determinant can be thought of as a measure of how much multiplication by the matrix expands or contracts space. If the determinant is 0, then space is contracted completely along at least one dimension, causing it to lose all of its volume. If the determinant is 1, then the transformation is volume-preserving.