ITCS 5356: Introduction to Machine Learning

Basic Linear Algebra

Razvan C. Bunescu

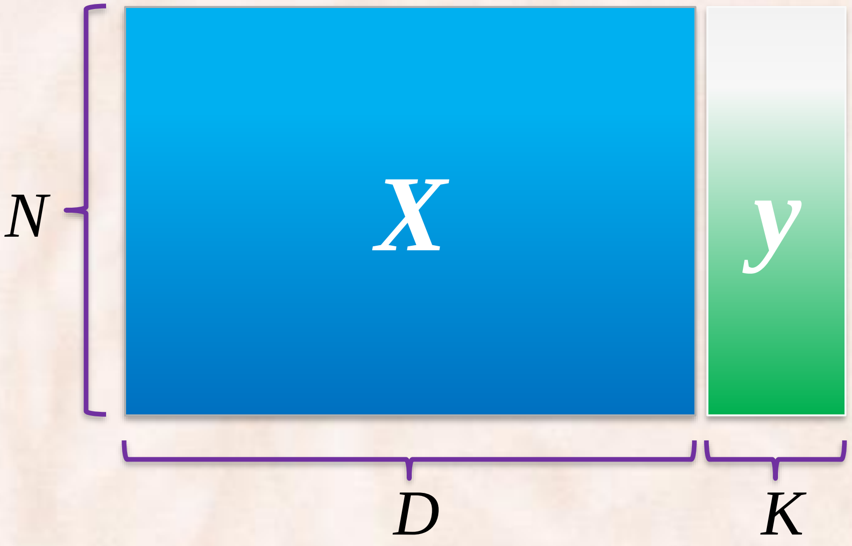
Department of Computer Science @ CCI

rbunescu@charlotte.edu

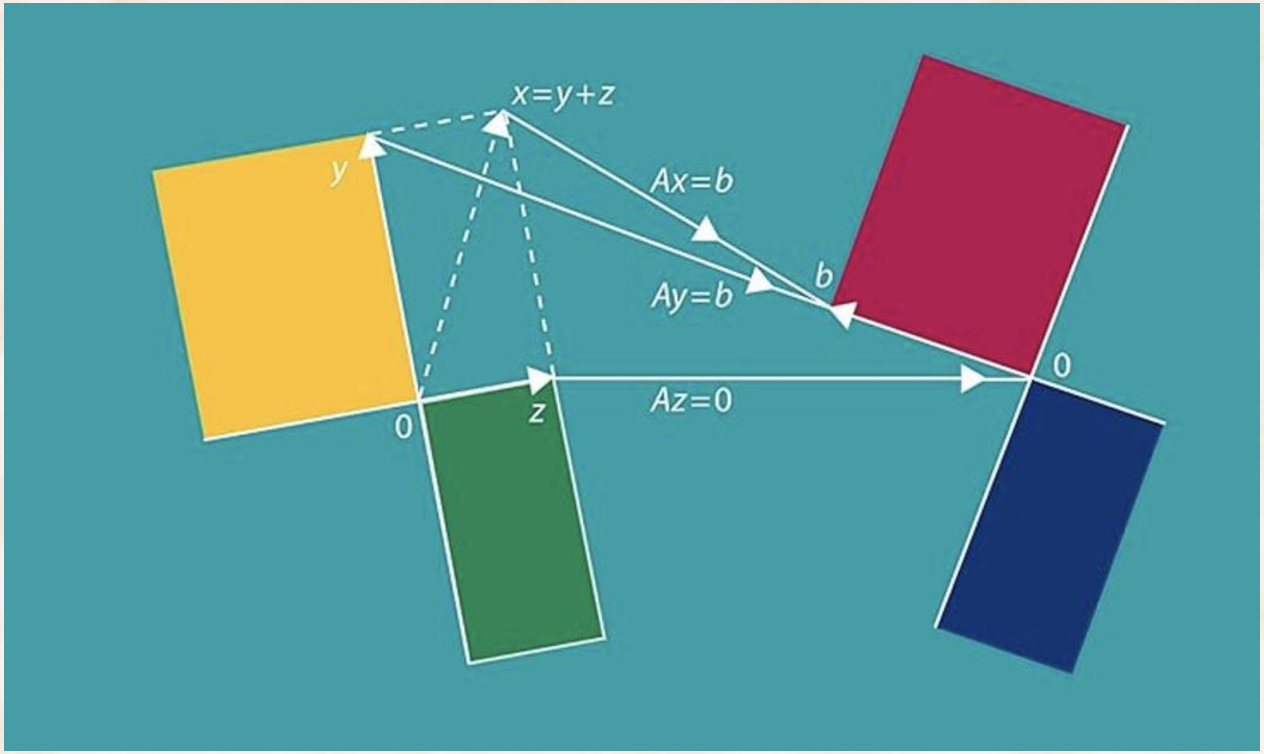
Slides adapted from 6.837 (Patrick Nicholas at MIT), MATH240 (Ryan Blair at UPenn), ITCS4156 (Minwoo Lee at UNCC), CS545 (Asa Ben-Hur at CSU), 10-601 (Matt Gormely at CMU)

Notations

- Training data samples: $\boldsymbol{X} \in \mathbb{R}^{N \times D}$
- The number of data samples: N
- The number of attributes: D
- The target label: $\boldsymbol{y} \in \mathbb{R}^{N \times K}$
- The dimension of target label: K
- Data: $\boldsymbol{D} = [\boldsymbol{X}, \boldsymbol{y}] \in \mathbb{R}^{N \times (D+K)}$
- The i-th sample pair: $(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)})$
- Test sample pair: $(\boldsymbol{x}', \boldsymbol{y}')$
- Label predicted: $\hat{\boldsymbol{y}}$ i.e. predicted by function f: $\hat{\boldsymbol{y}} = f(x')$



Linear Algebra: Matrix Operations



Data to discover prediction function f

- $\hat{\boldsymbol{y}}^{(i)} = f(\mathbf{x}^{(i)})$

- Let $\mathbf{x}^{(i)} = \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ x_2^{(i)} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ and $\boldsymbol{y}^{(i)} = 6$

- What can be a good f?
  - Linear formulation $f(\mathbf{x}) = ax_0 + bx_1 + cx_2$
    - a, b, c are the parameters or weights.
      - $(a, b, c) = (1,2,1)$ or $(2,1,2)$ or $(1,1,3)$ or $(0,0,6)$ or …

Data to discover function f

- $\hat{\boldsymbol{y}} = f(\boldsymbol{X})$
- Let $\mathbf{X} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ and $\boldsymbol{y} = \begin{bmatrix} 6 \\ 4 \\ 5 \end{bmatrix}$
- What can be a good f?
  - If using linear formulation $f(x) = ax_0 + bx_1 + cx_2$
    - $(a, b, c) = (1,2,1)$

How can we model such solutions?

Linear Algebra

Matrix & Vector Operations

- Basic Operations
- Dot Product, Norm, Cosine
- Transpose
- Inverse
- Determinants
- Trace
- Rank & Linear Independence

Basic Operations

- **Scalar multiplication**:
  - Multiply a matrix with a number.
    - Distribute (broadcast) scalar multiplication to all entries.

$$3 \cdot \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \end{bmatrix} = \begin{bmatrix} 3 \cdot 1 & 3 \cdot 2 & 3 \cdot 3 \\ 3 \cdot 0 & 3 \cdot -1 & 3 \cdot -2 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 6 & 9 \\ 0 & -6 & -6 \end{bmatrix}$$

Basic Operations

- The **transpose** $A^{\top}$:
  - Rows of A become columns of $A^{\top}$ and columns of A become rows of $A^{\top}$.

$$A^{T}_{ij} = A_{ji}$$

- Definition:
  - A matrix is **symmetric** if $A^{\top} = A$
  - A matrix is **square** if it is of size $n \times n$
  - A matrix is **diagonal** if it is square and the only non-zero entries are a_{ii} for some i.

Definitions

- A matrix is **symmetric** if $A^\top = A$
- A matrix is **square** if it is of size $n \times n$
- A matrix is **diagonal** if it is square and the only non-zero entries are of the form a_{ii} for some i
- The **identity matrix** of dimension n, denoted I_n, is the $n \times n$ diagonal matrix where all the diagonal entries are 1.

Basic Operations

- **Addition** and **Subtraction**:
  - Both matrices must have same dimensions.
  - Add or subtract **element-wise**.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} - \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a-e & b-f \\ c-g & d-h \end{bmatrix}$$

Matrix Properties

Let A and B be two $m \times n$ matrices.

Let k and p be two scalars.

Let $\mathbf{0}$ be the $m \times n$ matrix of all 0's.

$$A + B = B + A$$

$$A + (B + C) = (A + B) + C$$

$$k(A + B) = kA + kB$$

$$(k + p)A = kA + pA$$

$$A + \mathbf{0} = A$$

$$A - A = \mathbf{0}$$

$$kA = \mathbf{0} \Rightarrow k = 0 \text{ or } A = \mathbf{0}$$

Basic Operations

- **Multiplication** $\mathrm{C} = A \cdot B$ or $\mathrm{C} = AB$:
  - Only possible if dimensions are $(m \times n) \cdot (n \times p)$.
    - Product dimensions for C are $(m \times p)$
  - Multiply each row in A by each column in B (row-column dot product).

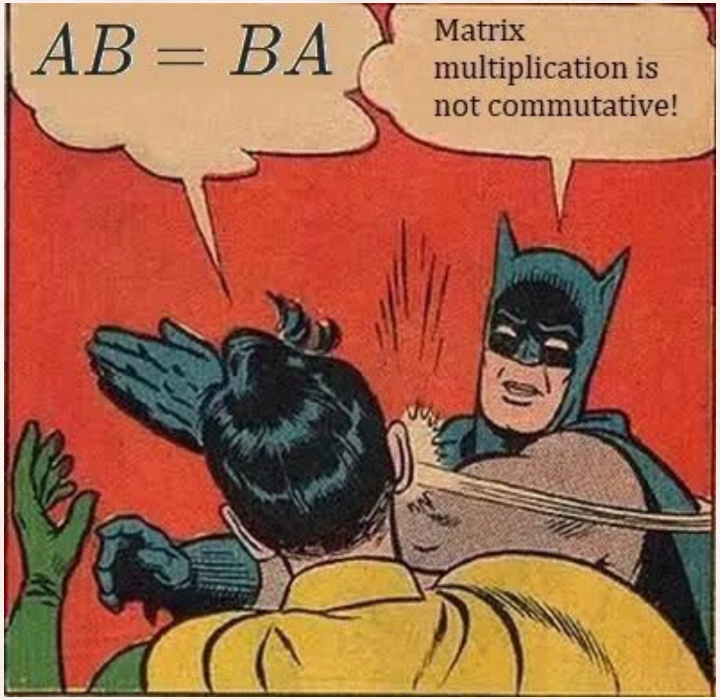
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$$

Basic Operations

- Is $AB = BA$?

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & \dots \\ \dots & \dots \end{bmatrix}$$

$$\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ea+fc & \dots \\ \dots & \dots \end{bmatrix}$$



Matrix Properties

Let A and B be two $m \times n$ matrices.

Let k and p be two scalars.

Let $\mathbf{0}$ be the $m \times n$ matrix of all 0's.

$A(BC) = (AB)C$

$A(B + C) = AB + AC$

$(A + B)C = AC + BC$

$k(AB) = (kA)B = A(kB)$

$I_m \cdot A = A$

$A \cdot I_n = A$

$kA = \mathbf{0} \ \Rightarrow k = 0 \text{ or } A = \mathbf{0}$

$(A^\top)^\top = A$

$(kA)^\top = kA^\top$

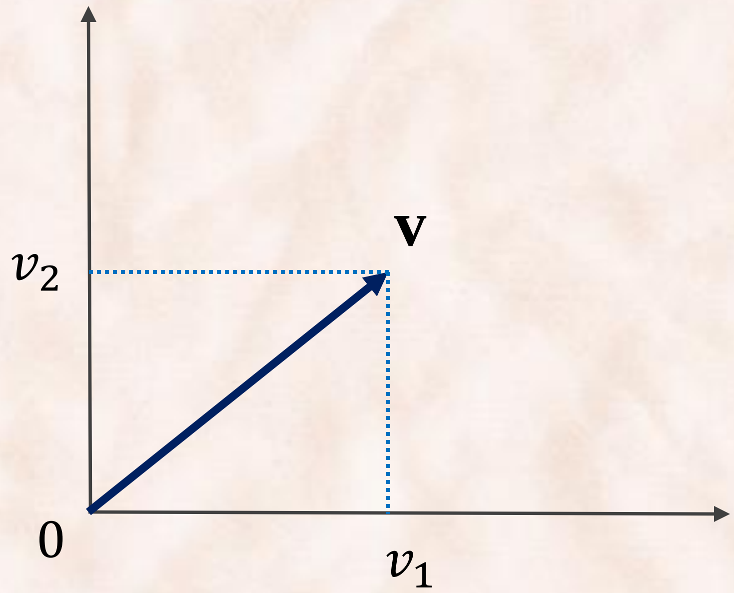
$(A + B)^\top = A^\top + B^\top$

$(AB)^\top = B^\top A^\top$

Vector Operations

- A **vector** is a $d \times 1$ matrix:
  - A point in a *d* dimensional space.
    - A directed segment connected the origin with this point

$$\mathbf{v} = \vec{\boldsymbol{v}} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix}$$



Vector Operations

- **Dot product**:

$$a \cdot b = a^\top b = [a\ b\ c]\begin{bmatrix} d \\ e \\ f \end{bmatrix} = ad + be + cf$$

the dot product is a matrix multiplication

- **Magnitude** or Euclidean / L2 **norm**:

$$\|a\| = \sqrt{a^\top a} = \sqrt{a^2 + b^2 + c^2}$$

the magnitude is the square root of the dot product of a vector with itself.

- **Angle**:

$$\cos(\theta) = \frac{a^\top b}{\|a\|\|b\|}$$

the cosine of the angle between two vectors is the dot product between their normalized versions.

Inverse of a Matrix

- **Identity matrix** (I)
  - $A \cdot I = I \cdot A = A$

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Some matrices have an **inverse**:
  - $A \cdot A^{-1} = A^{-1} \cdot A = I$
  - Both A and A^{-1} must be square.

- Properties of inversion operator:
  - $(A^{-1})^{-1} = A$
  - $(AB)^{-1} = B^{-1} A^{-1}$

Determinant of a Matrix

- A real number associated with a square matrix:
  - If det(A) = 0, then A has no inverse.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \det(A) = ad - bc \qquad A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$\det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = a \cdot \det\begin{bmatrix} e & f \\ h & i \end{bmatrix} - b \cdot \det\begin{bmatrix} d & f \\ g & i \end{bmatrix} + c \cdot \det\begin{bmatrix} d & e \\ g & h \end{bmatrix}$$

  - See Laplace expansion for recursive computation of determinants, or Leibniz formula based on permutations.

Orthonormal Basis

- A set of vectors B in a vector space V is called a **basis** iff every vector in V can be written as a linear combination of basis vectors from B.
  - **Orthonormal** basis: *Orthogonal + Normal*
    - ***Orthogonal*** basis: if all pairs of basis vectors are orthogonal, i.e. their dot-product is zero.
    - ***Normal*** basis: is all basis vectors have magnitude of one.

***Orthonormal** basis in 3D vector space*

$$x = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$$
$$y = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$$
$$z = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$$

***Orthogonal** basis vectors*

$$x \cdot y = 0$$
$$x \cdot z = 0$$
$$y \cdot z = 0$$

***Normal** basis vectors*

$$\|x\| = 1$$
$$\|y\| = 1$$
$$\|z\| = 1$$

Trace of a Matrix

- The **trace** of a square matrix A is the sum of the diagonal entries of A.

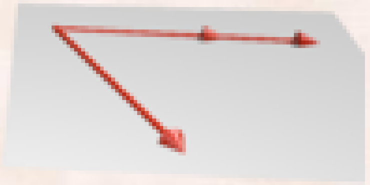
$$A = \begin{bmatrix} 1 & 4 & -1 \\ -2 & -3 & 2 \\ 3 & -4 & 7 \end{bmatrix}$$

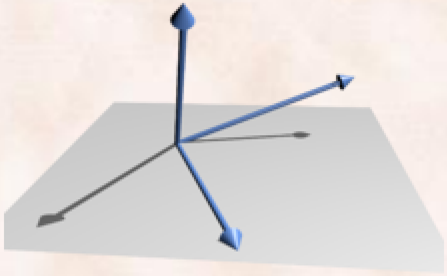
$$\mathrm{tr}(A) = 1 + (-3) + 7 = 5$$

Linearly Independence

- The vectors $\mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v}_m$ are **linearly dependent** if there exist scalars $a_1, a_2, \ldots, a_m$, not all zero, such that $a_1\mathrm{v}_1 + a_2\mathrm{v}_2 + \cdots + a_m\mathrm{v}_m = \mathbf{0}$.

- The vectors $\mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v}_m$ are **linearly independent** if $a_1\mathrm{v}_1 + a_2\mathrm{v}_2 + \cdots + a_m\mathrm{v}_m = \mathbf{0}$ implies $a_1 = a_2 = \cdots = a_m = 0$.





Rank of a MAtrix

- The **rank** of an $m \times n$ matrix A is the maximum number of linearly independent rows.
  - It is also the maximum number of linearly independent columns.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad \text{rank}(A) = 2$$

$$-\,row_1 + 2\,row_2 - row_3 = 0$$

Matrix Invertibility

- A square $n \times n$ matrix A is invertible *iff* $\text{rank}(A) = N$.
- A square $n \times n$ matrix A is invertible *iff* $\det(A) \neq 0$.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

- Columns are linearly independent => $\text{rank}(A) = 2$ => A is invertible.
- $\det(A) = 1 \cdot 4 - 3 \cdot 2 = -2 \neq 0$ => A is invertible

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \end{bmatrix}$$

- The 3 columns are NOT linearly independent ($\text{rank}(A) = 2$) => A is not invertible.
- $\det(A) = 1 \cdot 1 \cdot 4 - 1 \cdot 1 \cdot 4 - 1 \cdot 2 \cdot 4 + 1 \cdot 3 \cdot 3 + 2 \cdot 2 \cdot 1 - 2 \cdot 3 \cdot 1 = 0$ => A is not invertible.

Euclidean Distance

- The **Euclidean distance** between two vectors **x** and **y** is:

$$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2 = \sqrt{(\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y})}$$

- Compute the distance between two vectors A and B below:

$$A = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad d(A,B) = ?$$

Supplemental Readings

- Chapter 2 in DL textbook on Linear Algebra.