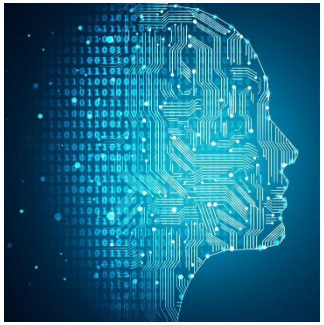
Principles of Machine Learning

Linear Algebra Review

Kevin Moon (kevin.moon@usu.edu)

STAT/CS 5810/6655



Motivation

- Linear algebra is essentially the study of matrices
- Most applied math, including machine learning, involves extensive use of linear algebra

Goals for Today

Not meant to be a comprehensive review. We'll focus on topics that are most important for getting started in machine learning

- Dot product and Euclidean norm
- Matrix-vector multiplication
- Linear combinations: span and linear independence
- Vector spaces
- Eigenvalue decomposition
- Positive (semi-)definite matrices

Dot Product

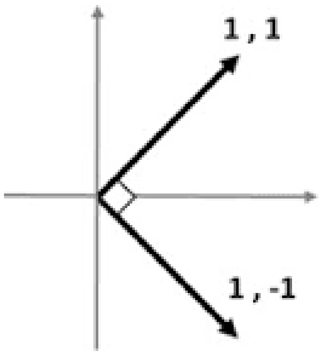
- Let

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_d \end{bmatrix}, \qquad \boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_d \end{bmatrix}$$

be two vectors. The *dot product* of $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$\begin{aligned} \langle \boldsymbol{u}, \boldsymbol{v} \rangle &:= u_1 v_1 + \cdots + u_d v_d \\ &= \sum_{i=1}^{d} u_i v_i \\ &= \boldsymbol{u}^T \boldsymbol{v} \end{aligned}$$

- We say $\boldsymbol{u}$ and $\boldsymbol{v}$ are *orthogonal* if $\boldsymbol{u} \neq \mathbf{0} \neq \boldsymbol{v}$, (where $\mathbf{0}$ denotes the zero vector), and $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$.



- There are generalizations of dot products in different (e.g. non-Euclidean) spaces known as inner products. This will come up later in the course.

Outer Product

- Let
$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_d \end{bmatrix}, \qquad \boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_d \end{bmatrix}$$
be two vectors. The *outer product* of $\boldsymbol{u}$ and $\boldsymbol{v}$ is
$$\boldsymbol{u}\boldsymbol{v}^T$$
which is a $d \times d$ matrix.

- Outer products arise when working with eigenvalue and singular value decompositions of a matrix

- Note that $\boldsymbol{u}\boldsymbol{v}^T \neq \boldsymbol{u}^T\boldsymbol{v}$

Euclidean Norm

- The *Euclidean norm* of a vector

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_d \end{bmatrix}$$

is

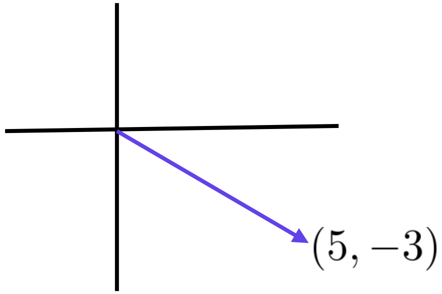
$$\begin{aligned} \|\boldsymbol{u}\| &:= \sqrt{u_1^2 + \cdots + u_d^2} \\ &= \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle}. \end{aligned}$$

- **Example:** If

$$\boldsymbol{u} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

then

$$\|\boldsymbol{u}\| = \sqrt{25 + 9} = \sqrt{34} \approx 5.8$$



- Similar to the dot product, the Euclidean norm is an example of a more general concept called a *norm*. In general, different norms are different ways of assessing the length of a vector.

Matrix-Vector Multiplication

- Let

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}, \qquad \boldsymbol{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n.$$

- Usual way to think about $A\boldsymbol{c}$:

$$A\boldsymbol{c} := \begin{bmatrix} \langle A_{1,:}, \boldsymbol{c} \rangle \\ \vdots \\ \langle A_{m,:}, \boldsymbol{c} \rangle \end{bmatrix} \in \mathbb{R}^m,$$

where $A_{i,:}$ denotes the i^{th} row of A.

- Alternate way to think about $A\boldsymbol{c}$:

$$A\boldsymbol{c} = c_1 A_{:,1} + \cdots + c_n A_{:,n}$$

where $A_{:,j}$ denotes the j^{th} column of A.

Matrix-Vector Multiplication

- Example

$$A = \begin{bmatrix} 3 & 4 & 1 \\ -2 & 5 & -4 \\ 1 & -2 & 3 \end{bmatrix} \qquad \boldsymbol{c} = \begin{bmatrix} 10 \\ 15 \\ -12 \end{bmatrix}$$

- Definition of matrix multiplication

$$A\boldsymbol{c} = \begin{bmatrix} 3 \cdot 10 + 4 \cdot 15 + 1 \cdot (-12) \\ -2 \cdot 10 + 5 \cdot 15 + (-4) \cdot (-12) \\ 1 \cdot 10 + (-2) \cdot 15 + 3 \cdot (-12) \end{bmatrix}$$

- Alternate way

$$A\boldsymbol{c} = 10 \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix} + 15 \begin{bmatrix} 4 \\ 5 \\ -2 \end{bmatrix} + (-12) \begin{bmatrix} 1 \\ -4 \\ 3 \end{bmatrix}$$

Span

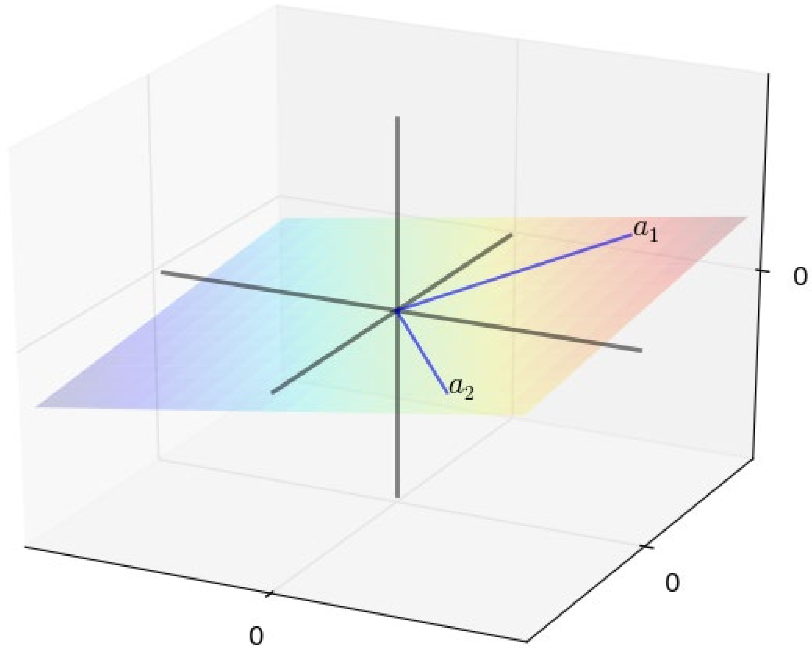
- Consider an arbitrary collection of vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbb{R}^m$.
- A *linear combination* of these vectors is any vector of the form

$$\sum_{j=1}^{n} c_j \boldsymbol{a}_j,$$

where $c_1, \ldots, c_n \in \mathbb{R}$.

- In this context, c_j is called a *scalar*
- The *span* of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ is the set of *all* linear combinations of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, and is denoted

$$\text{span}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$$



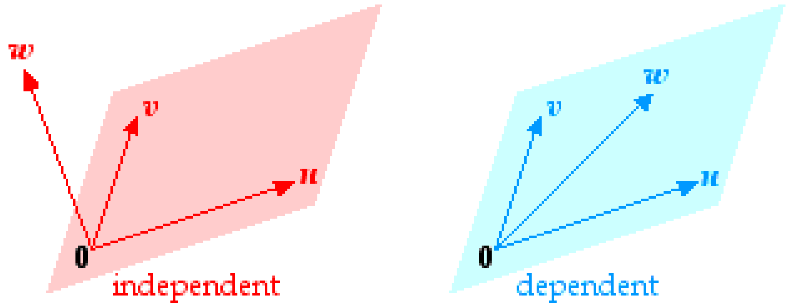
$m = 3, n = 2$

Linear Independence

- Vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbb{R}^m$ are *linearly independent* iff the following implication holds:

$$\sum_{j=1}^{n} c_j \boldsymbol{a}_j = 0 \implies c_j = 0 \; \forall j$$

- If $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are not linearly independent, they are said to be *linearly dependent*.



Vector spaces

A vector space is a set V and a field F with two operations:

1. Addition $+$: Adding two vectors gives you another vector.
   - Mathematically, $\forall \boldsymbol{v}, \boldsymbol{w} \in V$, $\boldsymbol{v} + \boldsymbol{w} \in V$
2. Scalar multiplication: multiplying a vector with a scalar gives another vector
   - Mathematically, $\forall \boldsymbol{v} \in V$ and $a \in F$, $a\boldsymbol{v} \in V$

Vector Spaces

- The two operations need to satisfy certain axioms
- **Example**: Need a zero vector
  - $\exists \mathbf{0} \in V$ s.t. $\boldsymbol{v} + \mathbf{0} = \boldsymbol{v}$ for all $\boldsymbol{v} \in V$
- **Example**: Additive inverse
  - $\forall \boldsymbol{v} \in V$, there exists an element $-\boldsymbol{v} \in V$ s.t. $\boldsymbol{v} + (-\boldsymbol{v}) = \mathbf{0}$
- See Wikipedia for the other axioms

Vector Spaces

Example: d-dimensional Euclidean space is a vector space

- Vector addition is defined as component-wise addition
- Scalar multiplication multiplies each entry in the vector by the scalar
- The zero vector has zero in each entry
- The additive inverse of a vector is obtained by multiplying the vector by -1

Subspaces

- A subspace of a vector space is a subset of that vector space that is itself a vector space
- **Example**: a 2-dimensional plane that intersects the origin is a subspace of $\mathbb{R}^3$
- **Example**: a 1-dimensional line that intersects the origin is a subspace of $\mathbb{R}^2$
- Is a 1-dimensional line that <u>does not</u> intersect the origin a subspace of $\mathbb{R}^2$?

- Every matrix has two fundamental associated subspaces: the **image** and the **nullspace**

Image of a Matrix

- Let A be an $m \times n$ matrix. The **image** of A is the set

$$\text{image}(A) := \{A\boldsymbol{c} | \boldsymbol{c} \in \mathbb{R}^n\}.$$

- This is a subspace of $\mathbb{R}^m$
  - Proof on the board
- Let the columns of A be denoted $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. Recall

$$A\boldsymbol{c} = \sum_{j=1}^{n} c_1 \boldsymbol{a}_j \,.$$

- Thus the image of A is the span of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$
- For this reason, we refer to the image as the **column span** of A

Nullspace of a Matrix

- Let A be an $m \times n$ matrix. The **nullspace** of A is the set

$$N(A) := \{\boldsymbol{c} \in \mathbb{R}^n | A\boldsymbol{c} = \boldsymbol{0}\}.$$

- This is a subspace of $\mathbb{R}^n$

Suppose $Au_1 = 0$ and $Au_2 = 0$.

Then $A(u_1 + u_2) = Au_1 + Au_2 = 0 + 0 + 0$

$\implies u_1 + u_2 \in N(A)$

- $N(A) = \{\boldsymbol{0}\}$ iff (if and only if) the columns of A are linearly independent

$N(A) = \{0\} \iff (Ac = 0 \implies c = 0)$

$\iff$ Colums of A are LI.

Examples

- $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$
  - $\text{image}(A) = \mathbb{R}^2$
  - $N(A) = \text{span}\left\{\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}\right\}$
- $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}$
  - $\text{image}(A) = x\text{-}y$ plane
  - $N(A) = \{0\}$

Group Exercise

Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

1. Determine the image
2. Determine the nullspace
3. Find a vector $\boldsymbol{c}$ such that $A\boldsymbol{c} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$
4. Determine the set of all $\boldsymbol{c}$ such that $A\boldsymbol{c} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

Dimension

- A *basis* for a subspace $S \subset \mathbb{R}^m$ is a set of linearly independent vectors

$$\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$$

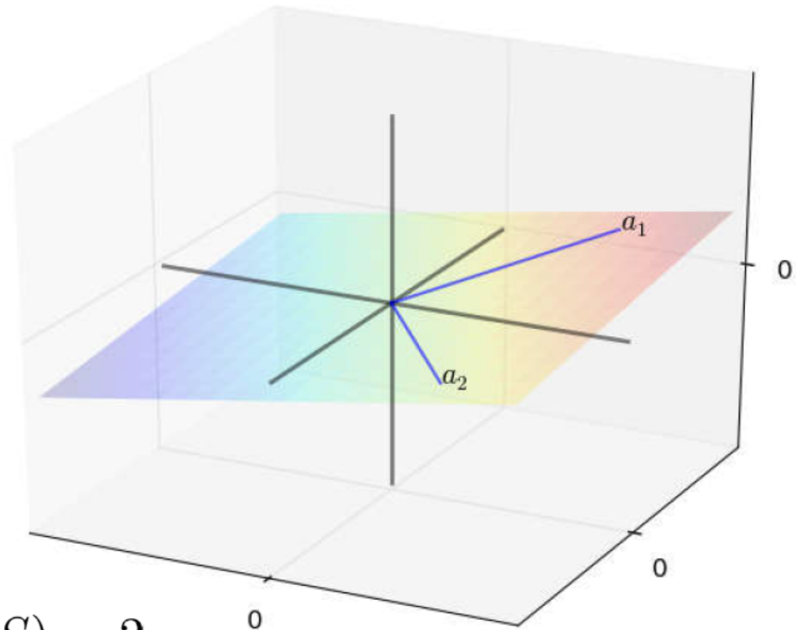
that span S.

- Fun fact: Any two bases for a subspace have the same number of elements. That number is called the *dimension* of the subspace, and is denoted $\dim(S)$.

- **Example:** $\dim(\mathbb{R}^n) = n$, and the *standard basis*

$$\boldsymbol{e}_i = [0, \ldots, 0, \underbrace{1}_{i^{th} \text{ position}}, 0, \ldots, 0]^\mathsf{T},$$

$i = 1, \ldots, n$, is one possible basis.



$$m = 3, \dim(S) = 2$$

Rank and Nullity

- For any matrix A, define

$$\text{rank}(A) := \dim(\text{colspan}(A))$$

  and

$$\text{nullity}(A) := \dim(N(A)).$$

- **Rank Plus Nullity Theorem:** For any matrix $A \in \mathbb{R}^{m \times n}$,

$$\text{rank}(A) + \text{nullity}(A) = n.$$

Example

- Consider

$$A = \begin{bmatrix} 1 & 0 & 0 & 4 & 1 & 2 & 3 \\ 0 & 1 & 0 & 5 & 4 & -7 & -4 \\ 0 & 0 & 1 & -3 & 2 & 9 & -5 \end{bmatrix}$$

- What is the dimension of the nullspace?

- $n = 7$
- $\text{rank}(A) = 3$
- $\text{nullity}(A) = 7 - 3 = 4$

Orthogonal Complements

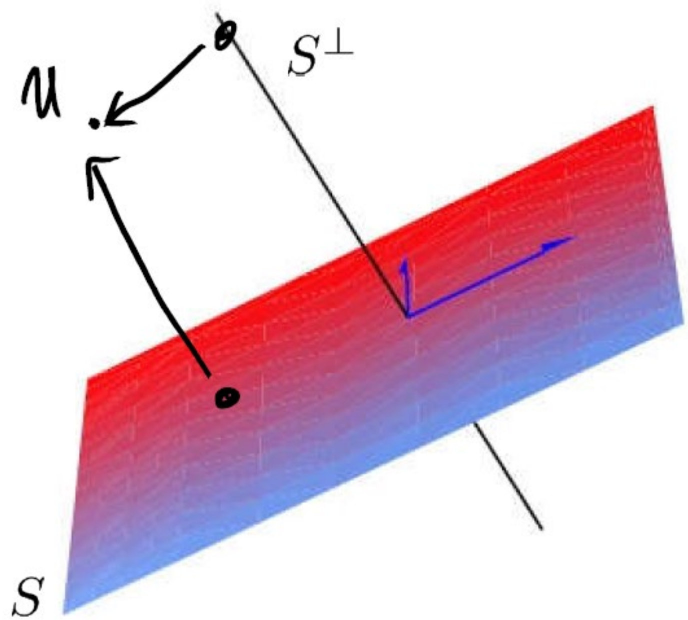
- If $S \subseteq \mathbb{R}^n$ is a subspace, the *orthogonal complement* of S is

$$S^\perp = \{\boldsymbol{w} \in \mathbb{R}^d : \langle \boldsymbol{w}, \boldsymbol{v} \rangle = 0 \ \forall \boldsymbol{v} \in S\}$$

- **Projection Theorem:** $\mathbb{R}^n = \mathbf{S} \bigoplus \mathbf{S}^\perp$, which means that every $\boldsymbol{u} \in \mathbb{R}^n$ can be written

$$\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{w}$$

for *unique* $\boldsymbol{v} \in S$ and $\boldsymbol{w} \in S^\perp$.



Exercise

Consider

$$A = \begin{bmatrix} 1 & 0 & 0 & 4 & 1 & 2 & 3 \\ 0 & 1 & 0 & 5 & 4 & -7 & -4 \\ 0 & 0 & 1 & -3 & 2 & 9 & -5 \end{bmatrix}$$

Express the nullspace of A as the orthogonal complement of another subspace.

- $N(A) = \text{rowspan}(A)^{\perp}$

Orthonormal Bases

- A set of vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ is an *orthonormal basis* for a subspace $S \subset \mathbb{R}^n$ if
  - $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ is a basis for S
  - For all i and j

$$\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Eigenvalues and Eigenvectors

- Let $A \in \mathbb{R}^{d \times d}$. If
$$A\boldsymbol{u} = \lambda \boldsymbol{u}$$
for some $\lambda \in \mathbb{R}$ and $\boldsymbol{u} \in \mathbb{R}^d$, we say λ is an *eigenvalue* of A and $\boldsymbol{u}$ is a corresponding *eigenvector*.

- If A is *symmetric*, i.e. $A = A^T$, we can characterize A in terms of its eigenvalues and eigenvectors.

Spectral Theorem

- A matrix $U \in \mathbb{R}^{d\times d}$ is said to be an *orthogonal matrix* if $U^T U = UU^T = I$, i.e., the transpose of U is its inverse.
- **Spectral Theorem:** If $A \in \mathbb{R}^{d\times d}$ is symmetric, then

$$A = U\Lambda U^T$$

  where U is an orthogonal matrix and Λ is a diagonal matrix.
- The expression $U\Lambda U^T$ is called the *spectral decomposition* or *eigenvalue decomposition* of A.
- Connection to eigenvalues/eigenvectors?

Spectral Theorem

- Multiplying $A = U\Lambda U^T$ on the right by U, we have

$$AU = U\Lambda.$$

Let

$$U = \begin{bmatrix} | & & | \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_d \\ | & & | \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_d \end{bmatrix}$$

If we look at the matrix equation $AU = U\Lambda$ one column at a time, we have

$$A\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i, i = 1, \ldots, d$$

Thus, λ_i are eigenvalues of A, and $\boldsymbol{u}_i$ are corresponding eigenvectors. Since U is an orthogonal matrix, $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d$ are an orthonormal basis of $\mathbb{R}^d$. Thus, the spectral theorem implies the existence of an ONB consisting of eigenvectors of A.

Group Exercise

An important identity is

$$U\Lambda U^T = \sum_{i=1}^{d} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T.$$

This can be verified by working out expressions for the entries on each side, and observing that they are equal.

1. For any fixed i, what is the rank of $\boldsymbol{u}_i \boldsymbol{u}_i^T$?
2. Define

   $$r = \#\{i : \lambda_i \neq 0\}.$$

   Argue that $\text{rank}(A) = r$.
3. (Optional) Show that if two square matrices produce the same results when applied to a basis, the two matrices are equal.
4. (Optional) Use the previous result to give an alternate proof of the identity in the first line above.

A bit more on rank

- The *rank* of a matrix A is equal to the number of columns in A that are linearly independent

- A matrix is said to be *full rank* if its rank is equal to the total number of columns

- If A is symmetric, then the rank of A is equal to the number of nonzero eigenvalues of A (proved in the previous exercises)

Positive (Semi-)Definite Matrices

- Let A be a $d \times d$ matrix. We say that A is *positive definite* if $\boldsymbol{x}^T A \boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$. We say that A is *positive semi-definite* if $\boldsymbol{x}^T A \boldsymbol{x} \geq 0$ for all $\boldsymbol{x}$.

- PD and PSD matrices arise frequently in ML, for example
  - Gram matrices
  - Kernel matrices
  - Covariance matrices
  - Hessian matrices

$$\begin{bmatrix} \langle \boldsymbol{x}_1, \boldsymbol{x}_1 \rangle & \cdots & \langle \boldsymbol{x}_1, \boldsymbol{x}_d \rangle \\ \vdots & \ddots & \vdots \\ \langle \boldsymbol{x}_d, \boldsymbol{x}_1 \rangle & \cdots & \langle \boldsymbol{x}_d, \boldsymbol{x}_d \rangle \end{bmatrix}$$

- PD/PSD matrices are not necessarily symmetric, e.g.,

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

However, in this course we will only consider PD/PSD matrices that are also symmetric.

Further Reading

- Linear Algebra review (available on Canvas)