# Lecture 1
# Basics of Linear Algebra

- Matrix Operations
- Matrix Derivative and Expectations
- Applications and Wrap-Up

# Contents

# Motivation

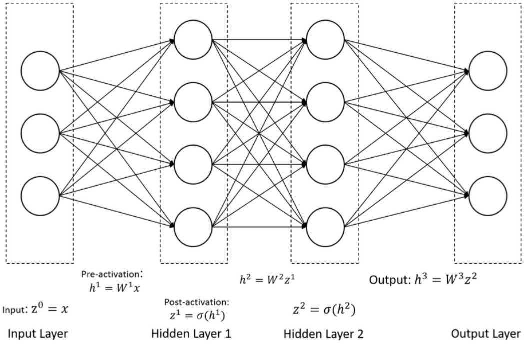


Figure: Example of a 3-layer fully-connected neural network. You should be able to understand its matrix representation.

# What is a Matrix?

Let $A = (a_{ij})$ be an $m \times n$ matrix.

- The $j$th column of $A$ is denoted by a column vector $\mathbf{a}_j$, i.e.,

$$\mathbf{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

- The $i$th row of $A$ is denoted by a row vector $\vec{\mathbf{a}}_i$, i.e.,

$$\vec{\mathbf{a}}_i = (a_{i1}, a_{i2}, \ldots, a_{in})$$

- Matrix $A$ can be represented in terms of either its columns and rows:

$$A = [\mathbf{a}_1, \cdots, \mathbf{a}_n] = \begin{bmatrix} \vec{\mathbf{a}}_1 \\ \vec{\mathbf{a}}_2 \\ \vdots \\ \vec{\mathbf{a}}_m \end{bmatrix}$$

# Matrix-Vector Multiplication

For an $m \times n$ matrix $A$ with the $i$th column $\mathbf{a}_i$, and a vector $\mathbf{u} = (u_1, u_2, \ldots, u_n)^\top$, the multiplication of $A$ and $\mathbf{u}$ is defined as

$$A\mathbf{u} = u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + \cdots + u_n\mathbf{a}_n$$

## Example

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 6 \\ -7 \\ 8 \\ -9 \end{bmatrix} = 6 \begin{bmatrix} 1 \\ 2 \end{bmatrix} - 7 \begin{bmatrix} 2 \\ 3 \end{bmatrix} + 8 \begin{bmatrix} 3 \\ 4 \end{bmatrix} - 9 \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

# Inner Product

- Given a vector $\mathbf{a} = (a_1, \ldots, a_n)^\top$ and a vector $\mathbf{b} = (b_1, \ldots, b_n)^\top$, following the rule of matrix-vector product, we have

$$\mathbf{a}^\top \mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots a_n b_n$$

- We call this special vector-vector multiplication the **inner product** (scalar product) of $\mathbf{a}$ and $\mathbf{b}$ (denoted by $\mathbf{a}^\top \mathbf{b}$ or $\langle \mathbf{a}, \mathbf{b} \rangle$)
- Properties: Commutative, bilinear
- Application: Cosine similarity, $\cos\theta = \frac{\mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}$

# Inner Product

- Given a vector $\mathbf{a} = (a_1, \ldots, a_n)^\top$ and a vector $\mathbf{b} = (b_1, \ldots, b_n)^\top$, following the rule of matrix-vector product, we have

$$\mathbf{a}^\top \mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots a_n b_n$$

- We call this special vector-vector multiplication the **inner product** (scalar product) of $\mathbf{a}$ and $\mathbf{b}$ (denoted by $\mathbf{a}^\top \mathbf{b}$ or $\langle \mathbf{a}, \mathbf{b} \rangle$)
- Properties: Commutative, bilinear
- Application: Cosine similarity, $\cos\theta = \frac{\mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}$

# Row Perspective of Multiplication

The matrix-vector multiplication $A\mathbf{u}$ has a row formula as

$$A\mathbf{u} = \begin{bmatrix} \vec{\mathbf{a}}_1\mathbf{u} \\ \vec{\mathbf{a}}_2\mathbf{u} \\ \vdots \\ \vec{\mathbf{a}}_m\mathbf{u} \end{bmatrix}$$

- Consider $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 6 & -7 & 8 & -9 \end{bmatrix}^\top$.
- We calculate

$$\vec{\mathbf{a}}_1\mathbf{u} = 6 \cdot 1 - 7 \cdot 2 + 8 \cdot 3 - 9 \cdot 4 = -20$$

$$\vec{\mathbf{a}}_2\mathbf{u} = 6 \cdot 2 - 7 \cdot 3 + 8 \cdot 4 - 9 \cdot 5 = -22$$

- We see that $A\mathbf{u} = \begin{bmatrix} -20 & -22 \end{bmatrix}^\top$

# Row Perspective of Multiplication

The matrix-vector multiplication $A\mathbf{u}$ has a row formula as

$$A\mathbf{u} = \begin{bmatrix} \vec{\mathbf{a}}_1 \mathbf{u} \\ \vec{\mathbf{a}}_2 \mathbf{u} \\ \vdots \\ \vec{\mathbf{a}}_m \mathbf{u} \end{bmatrix}$$

- Consider $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 6 & -7 & 8 & -9 \end{bmatrix}^\top$.
- We calculate

$$\vec{\mathbf{a}}_1 \mathbf{u} = 6 \cdot 1 - 7 \cdot 2 + 8 \cdot 3 - 9 \cdot 4 = -20$$

$$\vec{\mathbf{a}}_2 \mathbf{u} = 6 \cdot 2 - 7 \cdot 3 + 8 \cdot 4 - 9 \cdot 5 = -22$$

- We see that $A\mathbf{u} = \begin{bmatrix} -20 & -22 \end{bmatrix}^\top$

# Row Perspective of Multiplication

The matrix-vector multiplication $A\mathbf{u}$ has a row formula as

$$A\mathbf{u} = \begin{bmatrix} \vec{\mathbf{a}}_1\mathbf{u} \\ \vec{\mathbf{a}}_2\mathbf{u} \\ \vdots \\ \vec{\mathbf{a}}_m\mathbf{u} \end{bmatrix}$$

- Consider $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 6 & -7 & 8 & -9 \end{bmatrix}^\top$.
- We calculate

$$\vec{\mathbf{a}}_1\mathbf{u} = 6 \cdot 1 - 7 \cdot 2 + 8 \cdot 3 - 9 \cdot 4 = -20$$

$$\vec{\mathbf{a}}_2\mathbf{u} = 6 \cdot 2 - 7 \cdot 3 + 8 \cdot 4 - 9 \cdot 5 = -22$$

- We see that $A\mathbf{u} = \begin{bmatrix} -20 & -22 \end{bmatrix}^\top$

# Linear Systems as Matrix Equations

Write the following linear systems into compact matrix form:

$$\begin{cases} 2x_1 + x_2 + x_3 = 5 \\ 4x_1 - 6x_2 = -2 \\ -2x_1 + 7x_2 + 2x_3 = 9 \end{cases} \qquad \Rightarrow A\mathbf{x} = \mathbf{b}$$

where

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}$$

# Rank of a Matrix

- The rank of a matrix $A$ is the number of linearly independent columns
- Equivalently, it is the number of linearly independent rows
- Example: $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ has rank 1
- Full rank: rank$(A) = \min(m, n)$ for $A \in \mathbb{R}^{m \times n}$
- Application: Determines solvability of linear systems $A\mathbf{x} = \mathbf{b}$

# Rank of a Matrix

- The rank of a matrix $A$ is the number of linearly independent columns
- Equivalently, it is the number of linearly independent rows
- Example: $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ has rank 1
- Full rank: $\text{rank}(A) = \min(m, n)$ for $A \in \mathbb{R}^{m \times n}$
- Application: Determines solvability of linear systems $A\mathbf{x} = \mathbf{b}$

# Identity Matrix

- The identity matrix of order $k$, denoted by $I$ or $I_k$, is a $k \times k$ square matrix whose diagonal elements are 1's and whose nondiagonal elements are 0's

$$I = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

- Properties: $AI = A$ for any compatible matrix $A$.

# Identity Matrix

- The identity matrix of order $k$, denoted by $I$ or $I_k$, is a $k \times k$ square matrix whose diagonal elements are 1's and whose nondiagonal elements are 0's

$$I = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

- Properties: $AI = A$ for any compatible matrix $A$.

# Inverse of a Matrix

- Let $A$ be a $k \times k$ matrix. The inverse of $A$, denoted by $A^{-1}$, is another $k \times k$ matrix such that

$$AA^{-1} = A^{-1}A = I$$

- If the inverse exists, it is unique
- Existence: $A^{-1}$ exists if and only if $\det(A) \neq 0$ (or equivalently, rank$(A) = k$)
- For $2 \times 2$ matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

# Inverse of a Matrix

- Let $A$ be a $k \times k$ matrix. The inverse of $A$, denoted by $A^{-1}$, is another $k \times k$ matrix such that

$$AA^{-1} = A^{-1}A = I$$

- If the inverse exists, it is unique
- Existence: $A^{-1}$ exists if and only if $\det(A) \neq 0$ (or equivalently, rank$(A) = k$)
- For $2 \times 2$ matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

# Inverse of a Matrix

- Let $A$ be a $k \times k$ matrix. The inverse of $A$, denoted by $A^{-1}$, is another $k \times k$ matrix such that

$$AA^{-1} = A^{-1}A = I$$

- If the inverse exists, it is unique
- Existence: $A^{-1}$ exists if and only if $\det(A) \neq 0$ (or equivalently, rank$(A) = k$)
- For $2 \times 2$ matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

# Inverse of a Matrix

- Let $A$ be a $k \times k$ matrix. The inverse of $A$, denoted by $A^{-1}$, is another $k \times k$ matrix such that

$$AA^{-1} = A^{-1}A = I$$

- If the inverse exists, it is unique
- Existence: $A^{-1}$ exists if and only if $\det(A) \neq 0$ (or equivalently, rank$(A) = k$)
- For $2 \times 2$ matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

# Transpose of a Matrix

- Let $A$ be an $n \times k$ matrix. The transpose of $A$, denoted by $A^\top$, is a $k \times n$ matrix whose columns are the rows of $A$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{bmatrix} \Rightarrow A^\top = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{nk} \end{bmatrix}$$

- Properties: $(A^\top)^\top = A$, $(AB)^\top = B^\top A^\top$

# Transpose of a Matrix

- Let $A$ be an $n \times k$ matrix. The transpose of $A$, denoted by $A^\top$, is a $k \times n$ matrix whose columns are the rows of $A$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{bmatrix} \Rightarrow A^\top = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{nk} \end{bmatrix}$$

- Properties: $(A^\top)^\top = A$, $(AB)^\top = B^\top A^\top$

# Symmetric Matrices

- Let $A$ be a $k \times k$ matrix. $A$ is said to be symmetric if

$$A = A^\top$$

- Examples: Covariance matrices, Hessian matrices
- Properties: Real eigenvalues, orthogonal eigenvectors
- Spectral theorem: $A = Q\Lambda Q^\top$ where $Q$ is orthogonal and $\Lambda$ is diagonal

# Symmetric Matrices

- Let $A$ be a $k \times k$ matrix. $A$ is said to be symmetric if

$$A = A^\top$$

- Examples: Covariance matrices, Hessian matrices
- Properties: Real eigenvalues, orthogonal eigenvectors
- Spectral theorem: $A = Q\Lambda Q^\top$ where $Q$ is orthogonal and $\Lambda$ is diagonal

# Symmetric Matrices

- Let $A$ be a $k \times k$ matrix. $A$ is said to be symmetric if

$$A = A^\top$$

- Examples: Covariance matrices, Hessian matrices
- Properties: Real eigenvalues, orthogonal eigenvectors
- Spectral theorem: $A = Q\Lambda Q^\top$ where $Q$ is orthogonal and $\Lambda$ is diagonal

# Symmetric Matrices

- Let $A$ be a $k \times k$ matrix. $A$ is said to be symmetric if

$$A = A^\top$$

- Examples: Covariance matrices, Hessian matrices
- Properties: Real eigenvalues, orthogonal eigenvectors
- Spectral theorem: $A = Q \Lambda Q^\top$ where $Q$ is orthogonal and $\Lambda$ is diagonal

# Idempotent Matrices

- Let $A$ be a $k \times k$ matrix. $A$ is called idempotent if

$$A = AA$$

- If $A$ is also symmetric, then $A$ is called symmetric idempotent
- If $A$ is symmetric idempotent, then $I - A$ is also symmetric idempotent
- Example: Projection matrices $P = X(X^\top X)^{-1}X^\top$

# Idempotent Matrices

- Let $A$ be a $k \times k$ matrix. $A$ is called idempotent if

$$A = AA$$

- If $A$ is also symmetric, then $A$ is called symmetric idempotent
- If $A$ is symmetric idempotent, then $I - A$ is also symmetric idempotent
- Example: Projection matrices $P = X(X^\top X)^{-1}X^\top$

# Idempotent Matrices

- Let $A$ be a $k \times k$ matrix. $A$ is called idempotent if

$$A = AA$$

- If $A$ is also symmetric, then $A$ is called symmetric idempotent
- If $A$ is symmetric idempotent, then $I - A$ is also symmetric idempotent
- Example: Projection matrices $P = X(X^\top X)^{-1} X^\top$

# Idempotent Matrices

- Let $A$ be a $k \times k$ matrix. $A$ is called idempotent if

$$A = AA$$

- If $A$ is also symmetric, then $A$ is called symmetric idempotent
- If $A$ is symmetric idempotent, then $I - A$ is also symmetric idempotent
- Example: Projection matrices $P = X(X^\top X)^{-1}X^\top$

# Orthonormal Matrices

- Let $A$ be a $k \times k$ matrix. If $A$ is an orthonormal matrix, then

$$A^\top A = I$$

- As a consequence, if $A$ is an orthonormal matrix, then

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

- Properties: Preserves norms and angles ($\|A\mathbf{x}\| = \|\mathbf{x}\|$)
- Examples: Rotation matrices, permutation matrices

# Orthonormal Matrices

- Let $A$ be a $k \times k$ matrix. If $A$ is an orthonormal matrix, then

$$A^\top A = I$$

- As a consequence, if $A$ is an orthonormal matrix, then

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

- Properties: Preserves norms and angles ($\|A\mathbf{x}\| = \|\mathbf{x}\|$)
- Examples: Rotation matrices, permutation matrices

# Orthonormal Matrices

- Let $A$ be a $k \times k$ matrix. If $A$ is an orthonormal matrix, then

$$A^\top A = I$$

- As a consequence, if $A$ is an orthonormal matrix, then

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

- Properties: Preserves norms and angles ($\|A\mathbf{x}\| = \|\mathbf{x}\|$)
- Examples: Rotation matrices, permutation matrices

# Orthonormal Matrices

- Let $A$ be a $k \times k$ matrix. If $A$ is an orthonormal matrix, then

$$A^\top A = I$$

- As a consequence, if $A$ is an orthonormal matrix, then

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

- Properties: Preserves norms and angles ($\|A\mathbf{x}\| = \|\mathbf{x}\|$)
- Examples: Rotation matrices, permutation matrices

# Quadratic Forms

- Let $\mathbf{y}$ be a $k \times 1$ vector, and let $A$ be a $k \times k$ matrix. The function

$$\mathbf{y}^\top A\mathbf{y} = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij} y_i y_j$$

  is called a quadratic form

- Geometric interpretation: Ellipsoids in $k$-dimensional space

- Example: Energy in physical systems, Mahalanobis distance

# Quadratic Forms

- Let $\mathbf{y}$ be a $k \times 1$ vector, and let $A$ be a $k \times k$ matrix. The function

$$\mathbf{y}^\top A\mathbf{y} = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij} y_i y_j$$

  is called a quadratic form

- Geometric interpretation: Ellipsoids in $k$-dimensional space

- Example: Energy in physical systems, Mahalanobis distance

# Positive Definite and Positive Semidefinite Matrices

Let $A$ be a $k \times k$ matrix.

- $A$ is said to be *positive definite* if

  (a) $A = A^\top$ (A is symmetric)

  (b) $\mathbf{y}^\top A\mathbf{y} > 0 \quad \forall \mathbf{y} \in \mathbb{R}^k, \mathbf{y} \neq 0$

- $A$ is said to be *positive semidefinite* if:

  (a) $A = A^\top$ (A is symmetric)

  (c) $\mathbf{y}^\top A\mathbf{y} \geq 0 \quad \forall \mathbf{y} \in \mathbb{R}^k$

- Tests: Eigenvalues $> 0$ (positive definite), eigenvalues $\geq 0$ (positive semidefinite)

- Application: Convex optimization, kernel methods

# Positive Definite and Positive Semidefinite Matrices

Let $A$ be a $k \times k$ matrix.

- $A$ is said to be *positive definite* if

(a) $A = A^\top$ (A is symmetric)

(b) $\mathbf{y}^\top A\mathbf{y} > 0 \quad \forall \mathbf{y} \in \mathbb{R}^k, \mathbf{y} \neq 0$

- $A$ is said to be *positive semidefinite* if:

(a) $A = A^\top$ (A is symmetric)

(c) $\mathbf{y}^\top A\mathbf{y} \geq 0 \quad \forall \mathbf{y} \in \mathbb{R}^k$

- Tests: Eigenvalues $> 0$ (positive definite), eigenvalues $\geq 0$ (positive semidefinite)
- Application: Convex optimization, kernel methods

# Positive Definite and Positive Semidefinite Matrices

Let $A$ be a $k \times k$ matrix.

- $A$ is said to be *positive definite* if

  (a) $A = A^\top$ (A is symmetric)

  (b) $\mathbf{y}^\top A\mathbf{y} > 0 \quad \forall \mathbf{y} \in \mathbb{R}^k, \mathbf{y} \neq 0$

- $A$ is said to be *positive semidefinite* if:

  (a) $A = A^\top$ (A is symmetric)

  (c) $\mathbf{y}^\top A\mathbf{y} \geq 0 \quad \forall \mathbf{y} \in \mathbb{R}^k$

- Tests: Eigenvalues $> 0$ (positive definite), eigenvalues $\geq 0$ (positive semidefinite)

- Application: Convex optimization, kernel methods

# Positive Definite and Positive Semidefinite Matrices

Let $A$ be a $k \times k$ matrix.

- $A$ is said to be *positive definite* if

(a) $A = A^\top$ (A is symmetric)

(b) $\mathbf{y}^\top A\mathbf{y} > 0 \quad \forall \mathbf{y} \in \mathbb{R}^k, \mathbf{y} \neq 0$

- $A$ is said to be *positive semidefinite* if:

(a) $A = A^\top$ (A is symmetric)

(c) $\mathbf{y}^\top A\mathbf{y} \geq 0 \quad \forall \mathbf{y} \in \mathbb{R}^k$

- Tests: Eigenvalues $> 0$ (positive definite), eigenvalues $\geq 0$ (positive semidefinite)

- Application: Convex optimization, kernel methods

# Trace of a Matrix

Let $A$ be a $k \times k$ matrix. The *trace* of $A$, denoted by trace$(A)$ or tr$(A)$, is the sum of the diagonal elements of $A$; thus,

$$\text{trace}(A) = \sum_{i=1}^{k} a_{ii}$$

**Properties:**

1. If $A$ is an $m \times n$ matrix and $B$ is an $n \times m$ matrix, then

$$\text{trace}(AB) = \text{trace}(BA)$$

2. If the matrices are appropriately conformable, then

$$\text{trace}(ABC) = \text{trace}(CAB)$$

3. If $A$ and $B$ are $k \times k$ matrices and $a$ and $b$ are scalars, then

$$\text{trace}(aA + bB) = a\text{trace}(A) + b\text{trace}(B)$$

# Trace of a Matrix

Let $A$ be a $k \times k$ matrix. The *trace* of $A$, denoted by trace$(A)$ or tr$(A)$, is the sum of the diagonal elements of $A$; thus,

$$\text{trace}(A) = \sum_{i=1}^{k} a_{ii}$$

**Properties:**

1. If $A$ is an $m \times n$ matrix and $B$ is an $n \times m$ matrix, then

$$\text{trace}(AB) = \text{trace}(BA)$$

2. If the matrices are appropriately conformable, then

$$\text{trace}(ABC) = \text{trace}(CAB)$$

3. If $A$ and $B$ are $k \times k$ matrices and $a$ and $b$ are scalars, then

$$\text{trace}(aA + bB) = a\text{trace}(A) + b\text{trace}(B)$$

# Trace of a Matrix

Let $A$ be a $k \times k$ matrix. The *trace* of $A$, denoted by trace$(A)$ or tr$(A)$, is the sum of the diagonal elements of $A$; thus,

$$\text{trace}(A) = \sum_{i=1}^{k} a_{ii}$$

**Properties:**

1. If $A$ is an $m \times n$ matrix and $B$ is an $n \times m$ matrix, then

$$\text{trace}(AB) = \text{trace}(BA)$$

2. If the matrices are appropriately conformable, then

$$\text{trace}(ABC) = \text{trace}(CAB)$$

3. If $A$ and $B$ are $k \times k$ matrices and $a$ and $b$ are scalars, then

$$\text{trace}(aA + bB) = a\text{trace}(A) + b\text{trace}(B)$$

# Rank of an Idempotent Matrix

- Let $A$ be an idempotent matrix. The rank of $A$ is equal to its trace

$$\text{rank}(A) = \text{trace}(A)$$

- Proof sketch: Use the fact that idempotent matrices are diagonalizable with eigenvalues 0 or 1
- Application: In regression, $\text{rank}(X) = \text{trace}(H)$ where $H = X(X^\top X)^{-1}X^\top$ is the hat matrix

# Rank of an Idempotent Matrix

- Let $A$ be an idempotent matrix. The rank of $A$ is equal to its trace

$$\text{rank}(A) = \text{trace}(A)$$

- Proof sketch: Use the fact that idempotent matrices are diagonalizable with eigenvalues 0 or 1
- Application: In regression, $\text{rank}(X) = \text{trace}(H)$ where $H = X(X^\top X)^{-1}X^\top$ is the hat matrix

# An Important Identity for a Partitioned Matrix

Let $\mathbf{X}$ be an $n \times p$ matrix partitioned such that

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$$

We note that

$$\begin{aligned} \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X} &= \mathbf{X} \\ \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top[\mathbf{X}_1 \ \mathbf{X}_2] &= \mathbf{X} \\ \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top[\mathbf{X}_1 \ \mathbf{X}_2] &= [\mathbf{X}_1 \ \mathbf{X}_2] \end{aligned}$$

Consequently,

$$\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}_1 = \mathbf{X}_1 \quad \text{and} \quad \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}_2 = \mathbf{X}_2$$

Similarly,

$$\mathbf{X}_1^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top = \mathbf{X}_1^\top \quad \text{and} \quad \mathbf{X}_2^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top = \mathbf{X}_2^\top$$

# Inverse of a Partitioned Matrix

Consider a matrix of the form

$$\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} \mathbf{X}_1^\top \mathbf{X}_1 & \mathbf{X}_1^\top \mathbf{X}_2 \\ \mathbf{X}_2^\top \mathbf{X}_1 & \mathbf{X}_2^\top \mathbf{X}_2 \end{bmatrix}$$

It can be shown that the inverse of this matrix is $(\mathbf{X}^\top X)^{-1}$ that equals

$$\begin{bmatrix} (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} + (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{X}_2 G \mathbf{X}_2^\top \mathbf{X}_1 (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} & -(\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{X}_2 G \\ -(\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{X}_2 G & G \end{bmatrix}$$

where

$$\mathbf{H}_1 = \mathbf{X}_1 (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \quad \text{and} \quad G = \left[\mathbf{X}_2^\top (\mathbf{I} - \mathbf{H}_1) \mathbf{X}_2\right]^{-1}$$

Application: Regression analysis with multiple groups of predictors

# Determinant

- The determinant of a square matrix $A$, denoted $\det(A)$ or $|A|$, is a scalar value
- Geometric interpretation: Scaling factor of the linear transformation
- For $2 \times 2$ matrix: $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$
- Properties:
  - $\det(AB) = \det(A)\det(B)$
  - $\det(A^{-1}) = 1/\det(A)$
  - $\det(A^\top) = \det(A)$
- Application: Testing invertibility, change of variables in integration

# Determinant

- The determinant of a square matrix $A$, denoted $\det(A)$ or $|A|$, is a scalar value
- Geometric interpretation: Scaling factor of the linear transformation
- For $2 \times 2$ matrix: $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$
- Properties:
  - $\det(AB) = \det(A)\det(B)$
  - $\det(A^{-1}) = 1/\det(A)$
  - $\det(A^\top) = \det(A)$
- Application: Testing invertibility, change of variables in integration

# Determinant

- The determinant of a square matrix $A$, denoted $\det(A)$ or $|A|$, is a scalar value
- Geometric interpretation: Scaling factor of the linear transformation
- For $2 \times 2$ matrix: $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$
- Properties:
  - $\det(AB) = \det(A)\det(B)$
  - $\det(A^{-1}) = 1/\det(A)$
  - $\det(A^\top) = \det(A)$
- Application: Testing invertibility, change of variables in integration

# Determinant

- The determinant of a square matrix $A$, denoted $\det(A)$ or $|A|$, is a scalar value
- Geometric interpretation: Scaling factor of the linear transformation
- For $2 \times 2$ matrix: $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$
- Properties:
  - $\det(AB) = \det(A)\det(B)$
  - $\det(A^{-1}) = 1/\det(A)$
  - $\det(A^\top) = \det(A)$
- Application: Testing invertibility, change of variables in integration

# Contents

# Matrix Derivatives

Let $\boldsymbol{A}$ be a $k \times k$ matrix of constants, $\boldsymbol{a}$ be a $k \times 1$ vector of constants, and $\boldsymbol{y}$ be a $k \times 1$ vector of variables.

1. If $z = \boldsymbol{a}^\top \boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial(\boldsymbol{a}^\top \boldsymbol{y})}{\partial \boldsymbol{y}} = \boldsymbol{a}$$

2. If $z = \boldsymbol{y}^\top \boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial(\boldsymbol{y}^\top \boldsymbol{y})}{\partial \boldsymbol{y}} = 2\boldsymbol{y}$$

3. If $z = \boldsymbol{a}^\top \boldsymbol{A} \boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial(\boldsymbol{a}^\top \boldsymbol{A} \boldsymbol{y})}{\partial \boldsymbol{y}} = \boldsymbol{A}^\top \boldsymbol{a}$$

4. If $z = \boldsymbol{y}^\top \boldsymbol{A} \boldsymbol{y}$ and $\boldsymbol{A}$ is symmetric, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial(\boldsymbol{y}^\top \boldsymbol{A} \boldsymbol{y})}{\partial \boldsymbol{y}} = 2\boldsymbol{A}\boldsymbol{y}$$

# Matrix Derivatives

Let $\boldsymbol{A}$ be a $k \times k$ matrix of constants, $\boldsymbol{a}$ be a $k \times 1$ vector of constants, and $\boldsymbol{y}$ be a $k \times 1$ vector of variables.

1. If $z = \boldsymbol{a}^\top \boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial(\boldsymbol{a}^\top \boldsymbol{y})}{\partial \boldsymbol{y}} = \boldsymbol{a}$$
2. If $z = \boldsymbol{y}^\top \boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial(\boldsymbol{y}^\top \boldsymbol{y})}{\partial \boldsymbol{y}} = 2\boldsymbol{y}$$
3. If $z = \boldsymbol{a}^\top \boldsymbol{A}\boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial(\boldsymbol{a}^\top \boldsymbol{A}\boldsymbol{y})}{\partial \boldsymbol{y}} = \boldsymbol{A}^\top \boldsymbol{a}$$
4. If $z = \boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}$ and $\boldsymbol{A}$ is symmetric, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial(\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y})}{\partial \boldsymbol{y}} = 2\boldsymbol{A}\boldsymbol{y}$$

# Matrix Derivatives

Let $\boldsymbol{A}$ be a $k \times k$ matrix of constants, $\boldsymbol{a}$ be a $k \times 1$ vector of constants, and $\boldsymbol{y}$ be a $k \times 1$ vector of variables.

1. If $z = \boldsymbol{a}^\top \boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial(\boldsymbol{a}^\top \boldsymbol{y})}{\partial \boldsymbol{y}} = \boldsymbol{a}$$
2. If $z = \boldsymbol{y}^\top \boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial(\boldsymbol{y}^\top \boldsymbol{y})}{\partial \boldsymbol{y}} = 2\boldsymbol{y}$$
3. If $z = \boldsymbol{a}^\top \boldsymbol{A}\boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial(\boldsymbol{a}^\top \boldsymbol{A}\boldsymbol{y})}{\partial \boldsymbol{y}} = \boldsymbol{A}^\top \boldsymbol{a}$$
4. If $z = \boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}$ and $\boldsymbol{A}$ is symmetric, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial(\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y})}{\partial \boldsymbol{y}} = 2\boldsymbol{A}\boldsymbol{y}$$

# Matrix Derivatives

Let $\boldsymbol{A}$ be a $k \times k$ matrix of constants, $\boldsymbol{a}$ be a $k \times 1$ vector of constants, and $\boldsymbol{y}$ be a $k \times 1$ vector of variables.

1. If $z = \boldsymbol{a}^\top \boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial (\boldsymbol{a}^\top \boldsymbol{y})}{\partial \boldsymbol{y}} = \boldsymbol{a}$$
2. If $z = \boldsymbol{y}^\top \boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial (\boldsymbol{y}^\top \boldsymbol{y})}{\partial \boldsymbol{y}} = 2\boldsymbol{y}$$
3. If $z = \boldsymbol{a}^\top \boldsymbol{A} \boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial (\boldsymbol{a}^\top \boldsymbol{A} \boldsymbol{y})}{\partial \boldsymbol{y}} = \boldsymbol{A}^\top \boldsymbol{a}$$
4. If $z = \boldsymbol{y}^\top \boldsymbol{A} \boldsymbol{y}$ and $\boldsymbol{A}$ is symmetric, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial (\boldsymbol{y}^\top \boldsymbol{A} \boldsymbol{y})}{\partial \boldsymbol{y}} = 2\boldsymbol{A}\boldsymbol{y}$$

# More Derivative Rules

- Application: Gradient descent optimization

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \nabla f(\mathbf{w}_t)$$

  where $\nabla f(\mathbf{w})$ is the gradient of the objective function

- Example: For linear regression with loss $L(\mathbf{w}) = \|\mathbf{y} - X\mathbf{w}\|^2$, the gradient is

$$\nabla L(\mathbf{w}) = -2X^\top(\mathbf{y} - X\mathbf{w})$$

- Chain rule for matrix derivatives: If $z = f(\mathbf{y})$ and $\mathbf{y} = g(\mathbf{x})$, then

$$\frac{\partial z}{\partial \mathbf{x}} = \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right)^\top \frac{\partial z}{\partial \mathbf{y}}$$

# More Derivative Rules

- Application: Gradient descent optimization

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \nabla f(\mathbf{w}_t)$$

  where $\nabla f(\mathbf{w})$ is the gradient of the objective function

- Example: For linear regression with loss $L(\mathbf{w}) = \|\mathbf{y} - X\mathbf{w}\|^2$, the gradient is

$$\nabla L(\mathbf{w}) = -2X^\top(\mathbf{y} - X\mathbf{w})$$

- Chain rule for matrix derivatives: If $z = f(\mathbf{y})$ and $\mathbf{y} = g(\mathbf{x})$, then

$$\frac{\partial z}{\partial \mathbf{x}} = \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right)^\top \frac{\partial z}{\partial \mathbf{y}}$$

# Expectations of Random Vectors

Let $\boldsymbol{A}$ be a $k \times k$ matrix of constants, $\boldsymbol{a}$ be a $k \times 1$ vector of constants, and $\boldsymbol{y}$ be a $k \times 1$ random vector with mean $\boldsymbol{\mu}$ and nonsingular variance–covariance matrix $V$.

1. $\mathbb{E}(\boldsymbol{a}^\top \boldsymbol{y}) = \boldsymbol{a}^\top \boldsymbol{\mu}$
2. $\mathbb{E}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{\mu}$
3. $\mathrm{Var}(\boldsymbol{a}^\top \boldsymbol{y}) = \boldsymbol{a}^\top V \boldsymbol{a}$
4. $\mathrm{Var}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A} V \boldsymbol{A}^\top$

   *Note:* If $V = \sigma^2 I$, then $\mathrm{Var}(\boldsymbol{A}\boldsymbol{y}) = \sigma^2 \boldsymbol{A}\boldsymbol{A}^\top$
5. $\mathbb{E}(\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}) = \mathrm{trace}(AV) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}$

   *Note:* If $V = \sigma^2 I$, then $\mathbb{E}(\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}) = \sigma^2 \mathrm{trace}(A) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}$

# Expectations of Random Vectors

Let $\boldsymbol{A}$ be a $k \times k$ matrix of constants, $\boldsymbol{a}$ be a $k \times 1$ vector of constants, and $\boldsymbol{y}$ be a $k \times 1$ random vector with mean $\boldsymbol{\mu}$ and nonsingular variance–covariance matrix $V$.

1. $\mathbb{E}(\boldsymbol{a}^\top \boldsymbol{y}) = \boldsymbol{a}^\top \boldsymbol{\mu}$
2. $\mathbb{E}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{\mu}$
3. $\text{Var}(\boldsymbol{a}^\top \boldsymbol{y}) = \boldsymbol{a}^\top V \boldsymbol{a}$
4. $\text{Var}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A} V \boldsymbol{A}^\top$

   *Note:* If $V = \sigma^2 I$, then $\text{Var}(\boldsymbol{A}\boldsymbol{y}) = \sigma^2 \boldsymbol{A}\boldsymbol{A}^\top$
5. $\mathbb{E}(\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}) = \text{trace}(AV) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}$

   *Note:* If $V = \sigma^2 I$, then $\mathbb{E}(\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}) = \sigma^2 \text{trace}(A) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}$

# Expectations of Random Vectors

Let $\boldsymbol{A}$ be a $k \times k$ matrix of constants, $\boldsymbol{a}$ be a $k \times 1$ vector of constants, and $\boldsymbol{y}$ be a $k \times 1$ random vector with mean $\boldsymbol{\mu}$ and nonsingular variance–covariance matrix $V$.

1. $\mathbb{E}(\boldsymbol{a}^\top \boldsymbol{y}) = \boldsymbol{a}^\top \boldsymbol{\mu}$
2. $\mathbb{E}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{\mu}$
3. $\text{Var}(\boldsymbol{a}^\top \boldsymbol{y}) = \boldsymbol{a}^\top V \boldsymbol{a}$
4. $\text{Var}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A} V \boldsymbol{A}^\top$

   *Note:* If $V = \sigma^2 I$, then $\text{Var}(\boldsymbol{A}\boldsymbol{y}) = \sigma^2 \boldsymbol{A}\boldsymbol{A}^\top$
5. $\mathbb{E}(\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}) = \text{trace}(AV) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}$

   *Note:* If $V = \sigma^2 I$, then $\mathbb{E}(\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}) = \sigma^2 \text{trace}(A) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}$

# Expectations of Random Vectors

Let $\boldsymbol{A}$ be a $k \times k$ matrix of constants, $\boldsymbol{a}$ be a $k \times 1$ vector of constants, and $\boldsymbol{y}$ be a $k \times 1$ random vector with mean $\boldsymbol{\mu}$ and nonsingular variance–covariance matrix $V$.

1. $\mathbb{E}(\boldsymbol{a}^\top \boldsymbol{y}) = \boldsymbol{a}^\top \boldsymbol{\mu}$
2. $\mathbb{E}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{\mu}$
3. $\text{Var}(\boldsymbol{a}^\top \boldsymbol{y}) = \boldsymbol{a}^\top V \boldsymbol{a}$
4. $\text{Var}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A} V \boldsymbol{A}^\top$

   *Note:* If $V = \sigma^2 I$, then $\text{Var}(\boldsymbol{A}\boldsymbol{y}) = \sigma^2 \boldsymbol{A}\boldsymbol{A}^\top$
5. $\mathbb{E}(\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}) = \text{trace}(AV) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}$

   *Note:* If $V = \sigma^2 I$, then $\mathbb{E}(\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}) = \sigma^2 \text{trace}(A) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}$

# Expectations of Random Vectors

Let $\boldsymbol{A}$ be a $k \times k$ matrix of constants, $\boldsymbol{a}$ be a $k \times 1$ vector of constants, and $\boldsymbol{y}$ be a $k \times 1$ random vector with mean $\boldsymbol{\mu}$ and nonsingular variance–covariance matrix $V$.

1. $\mathbb{E}(\boldsymbol{a}^\top \boldsymbol{y}) = \boldsymbol{a}^\top \boldsymbol{\mu}$
2. $\mathbb{E}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{\mu}$
3. $\mathrm{Var}(\boldsymbol{a}^\top \boldsymbol{y}) = \boldsymbol{a}^\top V \boldsymbol{a}$
4. $\mathrm{Var}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A} V \boldsymbol{A}^\top$

   *Note:* If $V = \sigma^2 I$, then $\mathrm{Var}(\boldsymbol{A}\boldsymbol{y}) = \sigma^2 \boldsymbol{A}\boldsymbol{A}^\top$
5. $\mathbb{E}(\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}) = \mathrm{trace}(AV) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}$

   *Note:* If $V = \sigma^2 I$, then $\mathbb{E}(\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}) = \sigma^2 \mathrm{trace}(A) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}$

# Applications of Matrix Expectations

- Portfolio variance: For portfolio returns $\mathbf{r}$ with weights $\mathbf{w}$,

$$\mathrm{Var}(\mathbf{w}^\top \mathbf{r}) = \mathbf{w}^\top \Sigma \mathbf{w}$$

  where $\Sigma$ is the covariance matrix of returns
- Risk estimation: For quadratic loss functions
- Signal processing: For estimating power in transformed signals
- Econometrics: In GMM and other estimation methods

# Applications of Matrix Expectations

- Portfolio variance: For portfolio returns $\mathbf{r}$ with weights $\mathbf{w}$,

$$\text{Var}(\mathbf{w}^\top \mathbf{r}) = \mathbf{w}^\top \Sigma \mathbf{w}$$

  where $\Sigma$ is the covariance matrix of returns
- Risk estimation: For quadratic loss functions
- Signal processing: For estimating power in transformed signals
- Econometrics: In GMM and other estimation methods

# Applications of Matrix Expectations

- Portfolio variance: For portfolio returns $\mathbf{r}$ with weights $\mathbf{w}$,

$$\text{Var}(\mathbf{w}^\top \mathbf{r}) = \mathbf{w}^\top \Sigma \mathbf{w}$$

  where $\Sigma$ is the covariance matrix of returns

- Risk estimation: For quadratic loss functions
- Signal processing: For estimating power in transformed signals
- Econometrics: In GMM and other estimation methods

# Applications of Matrix Expectations

- Portfolio variance: For portfolio returns $\mathbf{r}$ with weights $\mathbf{w}$,

$$\text{Var}(\mathbf{w}^\top \mathbf{r}) = \mathbf{w}^\top \Sigma \mathbf{w}$$

  where $\Sigma$ is the covariance matrix of returns
- Risk estimation: For quadratic loss functions
- Signal processing: For estimating power in transformed signals
- Econometrics: In GMM and other estimation methods

# Contents

# Applications in AI

- Neural networks: Weight matrices and activation functions

$$\mathbf{h}^{(l)} = f(W^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)})$$

- Principal Component Analysis (PCA): Eigendecomposition of covariance matrix

$$\Sigma = Q\Lambda Q^\top$$

- Linear regression: Least squares solution

$$\hat{\beta} = (X^\top X)^{-1}X^\top y$$

- Support Vector Machines: Quadratic optimization with linear constraints

# Further Reading

- Strang, G. (2016). *Introduction to Linear Algebra*
- Boyd, S. & Vandenberghe, L. (2018). *Introduction to Applied Linear Algebra*
- MIT OpenCourseWare: Linear Algebra

Next lecture: Derivative of Neural Network Functions