Mathematics for Machine Learning

Vazgen Mikayelyan

July 18, 2020

FAST DISCOVERING THE FUTURE

Elementary transformations

Definition

Given a SLE

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases} .$$

$$A := \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$$ *is called the* **coefficient matrix** *and*

$\left[\ A \mid \mathbf{b}\ \right]$ *is called the* **augmented matrix**,

where $\mathbf{b} = \left[b_1, b_2, \ddots, b_m\right]^T$

Elementary transformations

Definition

A matrix is in **row echelon form (REF)** *if it satisfies the following properties:*
1. Any rows consisting entirely of zeros are at the bottom.
2. In each nonzero row, the first nonzero entry (called the **leading entry** *or* **pivot***) is in a column to the left of any leading entries below it.*

Elementary Row Operations

Example

The following matrices are in row echelon form:

$$\begin{bmatrix} 1 & 3 & 5 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & -2 & 7 \\ 0 & 4 & 1 \\ 0 & 0 & 3 \end{bmatrix} \quad \begin{bmatrix} 3 & 1 & -2 & 7 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

Assuming that each of these matrices is an augmented matrix, write out the corresponding SLE and solve them.

Elementary Row Operations

Example

The following matrices are in row echelon form:

$$\begin{bmatrix} 1 & 3 & 5 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & -2 & 7 \\ 0 & 4 & 1 \\ 0 & 0 & 3 \end{bmatrix} \quad \begin{bmatrix} 3 & 1 & -2 & 7 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

Assuming that each of these matrices is an augmented matrix, write out the corresponding SLE and solve them.

Definition

The following **elementary row operations** *can be performed on a matrix:*

1. *Interchange two rows.*
2. *Multiply a row by a nonzero constant.*
3. *Add a multiple of a row to another row.*

Elementary Row Operations

Shorthand notation

1. $R_i \leftrightarrow R_j$ means interchange rows i and j.
2. kR_i means multiply row i by k.
3. $R_i + kR_j$ means add k times row j to row i (and replace row i with the result).

Elementary Row Operations

Shorthand notation

1. $R_i \leftrightarrow R_j$ means interchange rows i and j.
2. kR_i means multiply row i by k.
3. $R_i + kR_j$ means add k times row j to row i (and replace row i with the result).

Example

Reduce the following matrix to echelon form:

$$\begin{bmatrix} 1 & -1 & 2 & 4 \\ 3 & -1 & 8 & 15 \\ -2 & 0 & -4 & -2 \end{bmatrix}.$$

Elementary Row Operations

Definition

The variables corresponding to the pivots in the row-echelon form are called **basic variables** *or* **leading variables**, *the other variables are* **free variables**.

Elementary Row Operations

Definition

The variables corresponding to the pivots in the row-echelon form are called **basic variables** *or* **leading variables**, *the other variables are* **free variables**.

Example

Let the REF of the augmented matrix be

$$\begin{bmatrix} 1 & -1 & 2 & 4 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The leading variables in row echelon form are x_1, x_3, *and the free variable is* x_2.

Gaussian Elimination

The following process is called **Gaussian elimination**.

Gaussian Elimination

The following process is called **Gaussian elimination**.

1. Write the augmented matrix of the SLE.

The following process is called **Gaussian elimination**.

1. Write the augmented matrix of the SLE.
2. Use elementary row operations to reduce the augmented matrix to row echelon form.

Gaussian Elimination

The following process is called **Gaussian elimination**.

1. Write the augmented matrix of the SLE.
2. Use elementary row operations to reduce the augmented matrix to row echelon form.
3. Using back substitution, solve the equivalent system that corresponds to the row-reduced matrix.

Gaussian Elimination

The following process is called **Gaussian elimination**.

1. Write the augmented matrix of the SLE.
2. Use elementary row operations to reduce the augmented matrix to row echelon form.
3. Using back substitution, solve the equivalent system that corresponds to the row-reduced matrix.

Example

Solve the system

$$\begin{cases} 2x_1 - x_2 + 5x_3 = -2 \\ x_1 - 2x_2 + 4x_3 = -7 \\ \quad\quad 3x_2 - 2x_3 = 9 \end{cases}.$$

Gauss-Jordan Elimination

Definition

A matrix is in **reduced row echelon form** *if it satisfies the following properties:*

1. *It is in row echelon form.*
2. *The leading entry in each non-zero row is a 1 (called a* **leading** *1).*
3. *Each column containing a leading 1 has zeros everywhere else.*

The entire process is called **Gauss-Jordan elimination**.

Gauss-Jordan Elimination

Example

The following matrix is in reduced row echelon form:

$$\begin{bmatrix} 0 & 1 & 3 & 0 & 0 & -1 & 4 & 0 & 6 \\ 0 & 0 & 0 & 1 & 0 & 3 & -2 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 2 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Gauss-Jordan Elimination

Example

The following matrix is in reduced row echelon form:

$$\begin{bmatrix} 0 & 1 & 3 & 0 & 0 & -1 & 4 & 0 & 6 \\ 0 & 0 & 0 & 1 & 0 & 3 & -2 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 2 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Example

Solve the system by Gauss-Jordan elimination.

$$\begin{cases} w - 2x + 3y + 2z = 1 \\ 2w - 4x + 7y + 2z = 4 \\ -3w + 6x - 7y - 10z = 1 \end{cases}.$$

The Gauss-Jordan Method for Computing the Inverse

The Gauss-Jordan Method for Computing the Inverse

Fundamental Theorem of Invertible Matrices

Let A be an $n \times n$ matrix. The following statements are equivalent:

1. A is invertible.
2. $A\mathbf{x} = \mathbf{b}$ has a unique solution for every vector $\mathbf{b} \in \mathbb{R}^n$.
3. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
4. The reduced row echelon form of A is I_n.
5. A is a product of elementary matrices.

The Gauss-Jordan Method for Computing the Inverse

Fundamental Theorem of Invertible Matrices

Let A be an $n \times n$ matrix. The following statements are equivalent:

1. A is invertible.
2. $A\mathbf{x} = \mathbf{b}$ has a unique solution for every vector $\mathbf{b} \in \mathbb{R}^n$.
3. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
4. The reduced row echelon form of A is I_n.
5. A is a product of elementary matrices.

Theorem

Let A be a square matrix. If a sequence of elementary row operations reduces A to I, then the same sequence of elementary row operations transforms I into A^{-1}.

The Gauss-Jordan Method for Computing the Inverse

Definition

Matrices A and B are called row equivalent if it is possible to obtain the matrix B by doing elementary row operations on A.

The Gauss-Jordan Method for Computing the Inverse

Definition

Matrices A and B are called row equivalent if it is possible to obtain the matrix B by doing elementary row operations on A.

If A is row equivalent to I, then elementary row operations will yield

$$[\ A \mid I\ ] \rightarrow [\ I \mid A^{-1}\ ].$$

The Gauss-Jordan Method for Computing the Inverse

Definition

Matrices A and B are called row equivalent if it is possible to obtain the matrix B by doing elementary row operations on A.

If A is row equivalent to I, then elementary row operations will yield

$$[\ A \mid I\ ] \to [\ I \mid A^{-1}\ ].$$

If A cannot be reduced to I, then the Fundamental Theorem guarantees us that A is not invertible.

Examples

Example

Find the inverse of

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

if it exists.

Examples

Example

Find the inverse of

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

if it exists.

Example

Find the inverse of

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

if it exists.

Determinant

Determinant

We will define determinant of a square matrix $A \in \mathbb{R}^{n\times n}$, $n \geq 2$ using mathematical induction.

Determinant

We will define determinant of a square matrix $A \in \mathbb{R}^{n\times n}$, $n \geq 2$ using mathematical induction. For $n = 2$, let

$$\det(A) = |A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} := a_{11}a_{22} - a_{12}a_{21}.$$

Determinant

We will define determinant of a square matrix $A \in \mathbb{R}^{n \times n}$, $n \geq 2$ using mathematical induction. For $n = 2$, let

$$\det(A) = |A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} := a_{11}a_{22} - a_{12}a_{21}.$$

Let $A_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the submatrix of A that we obtain when deleting row i and column j.

Determinant

We will define determinant of a square matrix $A \in \mathbb{R}^{n \times n}$, $n \geq 2$ using mathematical induction. For $n = 2$, let

$$\det(A) = |A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} := a_{11}a_{22} - a_{12}a_{21}.$$

Let $A_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the submatrix of A that we obtain when deleting row i and column j. We will define $\det(A)$ in the following way:

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

Determinant

We will define determinant of a square matrix $A \in \mathbb{R}^{n\times n}$, $n \geq 2$ using mathematical induction. For $n = 2$, let

$$\det(A) = |A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} := a_{11}a_{22} - a_{12}a_{21}.$$

Let $A_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the submatrix of A that we obtain when deleting row i and column j. We will define $\det(A)$ in the following way:

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

Remark

It can be proved that $\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ji} \det(A_{ji}).$

Determinant

Example

Compute the following determinant

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Determinant

Example

Compute the following determinant

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Example

Compute determinant of an upper triangular matrix.

Invertible Matrix Theorem

Theorem

For any square matrix $A \in \mathbb{R}^{n \times n}$ *it holds that* A *is ivertible if and only if* $\det(A) \neq 0$.

Properties of Determinant

- $\det(AB) = \det(A)\det(B)$

Properties of Determinant

- $\det(AB) = \det(A)\det(B)$
- $\det(A) = \det\left(A^T\right)$

Properties of Determinant

- $\det(AB) = \det(A)\det(B)$
- $\det(A) = \det\left(A^T\right)$
- If A is invertible then $\det\left(A^{-1}\right) = \dfrac{1}{\det(A)}$.

Properties of Determinant

- $\det(AB) = \det(A)\det(B)$
- $\det(A) = \det\left(A^T\right)$
- If A is invertible then $\det\left(A^{-1}\right) = \dfrac{1}{\det(A)}$.
- Multiplication of a column/row with $\lambda \in \mathbb{R}$ scales $\det(A)$ by λ. In particular, $\det(\lambda A) = \lambda^n \det(A)$.

Properties of Determinant

- $\det(AB) = \det(A)\det(B)$
- $\det(A) = \det\left(A^T\right)$
- If A is invertible then $\det\left(A^{-1}\right) = \dfrac{1}{\det(A)}$.
- Multiplication of a column/row with $\lambda \in \mathbb{R}$ scales $\det(A)$ by λ. In particular, $\det(\lambda A) = \lambda^n \det(A)$.
- Swapping two rows/columns changes the sign of $\det(A)$.

Properties of Determinant

- $\det(AB) = \det(A)\det(B)$
- $\det(A) = \det\left(A^T\right)$
- If A is invertible then $\det\left(A^{-1}\right) = \dfrac{1}{\det(A)}$.
- Multiplication of a column/row with $\lambda \in \mathbb{R}$ scales $\det(A)$ by λ. In particular, $\det(\lambda A) = \lambda^n \det(A)$.
- Swapping two rows/columns changes the sign of $\det(A)$.
- Adding a multiple of a column/row to another one does not change $\det(A)$.

Properties of Determinant

- $\det(AB) = \det(A)\det(B)$
- $\det(A) = \det\left(A^T\right)$
- If A is invertible then $\det\left(A^{-1}\right) = \dfrac{1}{\det(A)}$.
- Multiplication of a column/row with $\lambda \in \mathbb{R}$ scales $\det(A)$ by λ. In particular, $\det(\lambda A) = \lambda^n \det(A)$.
- Swapping two rows/columns changes the sign of $\det(A)$.
- Adding a multiple of a column/row to another one does not change $\det(A)$.
- If $A = QBQ^{-1}$ then $\det(A) = \det(B)$.

Vector Spaces and Subspaces

Definition

Let V be a set on which addition and scalar multiplication have been defined. If the following axioms hold for all $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ in V and for all scalars c and $d \in \mathbb{R}$ then V is called a **vector space** *and its elements are called* **vectors.**

1. $\mathbf{u}+\mathbf{v}$ *is in* V.	*Closure under addition*
2. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$	*Commutativity*
3. $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$	*Associativity*
4. $\exists \mathbf{0} \in V$, (*called a* **zero vector**),	$s.t. \mathbf{u}+\mathbf{0}=\mathbf{u}$
5. $\forall \mathbf{u} \in V, \exists -\mathbf{u} \in V$	$s.t.\ \mathbf{u}+(-\mathbf{u})=\mathbf{0}$
6. $c\mathbf{u} \in V$	*Closure under scalar mult.*
7. $c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$	*Distributivity*
8. $(c+d)\mathbf{u}=c\mathbf{u}+d\mathbf{u}$	*Distributivity*
9. $c(d\mathbf{u})=(cd)\mathbf{u}$	
10. $1\mathbf{u}=\mathbf{u}$	

Exercise

Prove that if V *is a vector space then* $0\mathbf{u} = \mathbf{0}$.

Exercise

Prove that if V *is a vector space then* $0\mathbf{u} = \mathbf{0}$.

Example

For any $n \geq 1$, $\mathbb{R}^n$ *is a vector space with the usual operations of addition and scalar multiplication.*

Vector Spaces and Subspaces

Exercise

Prove that if V *is a vector space then* $0\mathbf{u} = \mathbf{0}$.

Example

For any $n \geq 1$, $\mathbb{R}^n$ *is a vector space with the usual operations of addition and scalar multiplication.*

Example

For any natural m *and* n, *the set of all* $m \times n$ *matrices* $R^{m \times n}$ *forms a vector space with the usual operations of matrix addition and matrix scalar multiplication. Here the "vectors" are actually matrices.*

Vector Spaces and Subspaces

Exercise

Prove that if V is a vector space then $0\mathbf{u} = \mathbf{0}$.

Example

For any $n \geq 1$, $\mathbb{R}^n$ is a vector space with the usual operations of addition and scalar multiplication.

Example

For any natural m and n, the set of all $m \times n$ matrices $R^{m \times n}$ forms a vector space with the usual operations of matrix addition and matrix scalar multiplication. Here the "vectors" are actually matrices.

Example

The set $\mathbb{Z}$ of integers with the usual operations is **not** *a vector space.*

Example

Let $\mathcal{P}_3$ denote the set of all polynomials of degree 3 or less with real coefficients.

Example

Let $\mathcal{P}_3$ denote the set of all polynomials of degree 3 or less with real coefficients. Define addition and scalar multiplication in the usual way.

Example

Let $\mathcal{P}_3$ denote the set of all polynomials of degree 3 or less with real coefficients. Define addition and scalar multiplication in the usual way. If

$$p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3,\ q(x) = b_0 + b_1 x + b_2 x^2 + b_3 x^3$$

are in $\mathcal{P}_3$, then

$$p(x) + q(x) =$$

$$= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + (a_3 + b_3)x^3$$

is also in $\mathcal{P}_3$.

Example

Let $\mathcal{P}_3$ denote the set of all polynomials of degree 3 or less with real coefficients. Define addition and scalar multiplication in the usual way. If

$$p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3, \; q(x) = b_0 + b_1 x + b_2 x^2 + b_3 x^3$$

are in $\mathcal{P}_3$, then

$$p(x) + q(x) =$$

$$= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + (a_3 + b_3)x^3$$

is also in $\mathcal{P}_3$. If c is a scalar, then

$$cp(x) = ca_0 + ca_1 x + ca_2 x^2 + ca_3 x^3.$$

Example

Let $\mathcal{P}_3$ denote the set of all polynomials of degree 3 or less with real coefficients. Define addition and scalar multiplication in the usual way. If

$$p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3, \; q(x) = b_0 + b_1 x + b_2 x^2 + b_3 x^3$$

are in $\mathcal{P}_3$, then

$$p(x) + q(x) =$$

$$= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + (a_3 + b_3)x^3$$

is also in $\mathcal{P}_3$. If c is a scalar, then

$$cp(x) = ca_0 + ca_1 x + ca_2 x^2 + ca_3 x^3.$$

In general, for any fixed $n \geq 0$, the set $\mathcal{P}_n$ of all polynomials of degree less than or equal to n is a vector space, as is the set $\mathcal{P}$ of all polynomials.

Subspaces

Definition

A subset W of a vector space V is called a subspace of V if W is itself a vector space with the same scalars, addition, and scalar multiplication as V.

Subspaces

Definition

A subset W of a vector space V is called a subspace of V if W is itself a vector space with the same scalars, addition, and scalar multiplication as V.

Example

A line through the origin is a subspace of $\mathbb{R}^2$.

Subspaces

Definition

A subset W of a vector space V is called a subspace of V if W is itself a vector space with the same scalars, addition, and scalar multiplication as V.

Example

A line through the origin is a subspace of $\mathbb{R}^2$.

Theorem

Let V be a vector space and let W be a nonempty subset of V. Then W is a subspace of V if and only if the following conditions hold:
a. If $\mathbf{u}$ and $\mathbf{v}$ are in W, then $\mathbf{u} + \mathbf{v}$ is in W.
b. If $\mathbf{u}$ is in W and c is a scalar, then $c\mathbf{u}$ is in W.

Subspaces

Let us have a look at some subspaces.

Subspaces

Let us have a look at some subspaces.

- For every vector space V the trivial subspaces are V itself and $\{0\}$.

Subspaces

Let us have a look at some subspaces.

- For every vector space V the trivial subspaces are V itself and $\{0\}$.
- The solution set of a homogeneous SLE $A\mathbf{x} = \mathbf{0}$ with n unknowns $\mathbf{x} = [x_1, \ldots, x_n]$ is a subspace of $\mathbb{R}^n$.

Subspaces

Let us have a look at some subspaces.

- For every vector space V the trivial subspaces are V itself and $\{0\}$.
- The solution set of a homogeneous SLE $A\mathbf{x} = \mathbf{0}$ with n unknowns $\mathbf{x} = [x_1, \ldots, x_n]$ is a subspace of $\mathbb{R}^n$.
- The solution of an inhomogeneous SLE $A\mathbf{x} = \mathbf{b}; \mathbf{b} \neq \mathbf{0}$ is not a subspace of $\mathbb{R}^n$.

Subspaces

Let us have a look at some subspaces.

- For every vector space V the trivial subspaces are V itself and $\{0\}$.
- The solution set of a homogeneous SLE $A\mathbf{x} = \mathbf{0}$ with n unknowns $\mathbf{x} = [x_1, \ldots, x_n]$ is a subspace of $\mathbb{R}^n$.
- The solution of an inhomogeneous SLE $A\mathbf{x} = \mathbf{b}; \mathbf{b} \neq \mathbf{0}$ is not a subspace of $\mathbb{R}^n$.
- The intersection of arbitrarily many subspaces is a subspace itself.

Subspaces

Let us have a look at some subspaces.

- For every vector space V the trivial subspaces are V itself and $\{0\}$.
- The solution set of a homogeneous SLE $A\mathbf{x} = \mathbf{0}$ with n unknowns $\mathbf{x} = [x_1, \ldots, x_n]$ is a subspace of $\mathbb{R}^n$.
- The solution of an inhomogeneous SLE $A\mathbf{x} = \mathbf{b}; \mathbf{b} \neq \mathbf{0}$ is not a subspace of $\mathbb{R}^n$.
- The intersection of arbitrarily many subspaces is a subspace itself.

Example

Which of the following is a subspace of R^2?

1. *a line passing through the origin*
2. *two distinct lines passing through the origin*
3. *a unit circle*
4. *a line not passing through the origin*