Linear Algebra Review - Part 2

STAT 415: Data Mining and Machine Learning

University of Michigan

Fall 2022

Outline

Linear Independence

- A set of vectors $\{x_1, x_2, \ldots x_n\} \subset \mathbb{R}^m$ is **linearly independent** if no vector can be represented as a linear combination of the remaining vectors.
- Conversely, if one vector belonging to the set can be represented as a linear combination of the remaining vectors, then the vectors are said to be **linearly dependent**. That is, if

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

for some scalar values $\alpha_1, \ldots, \alpha_{n-1} \in \mathbb{R}$, then we say that the vectors $x_1, \ldots, x_n$ are linearly dependent.

Linear Independence

Example:

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \quad x_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

Question: Are the vectors x_1, x_2, x_3 linearly dependent?

Hint: $x_3 = -2x_1 + x_2$

Rank

- **column rank**: the size of the largest subset of columns of matrix A that constitute a linearly independent set. (the number of linearly independent columns of matrix A.)
- **row rank**: the largest number of rows of A that constitute a linearly independent set.
- For any matrix $A \in \mathbb{R}^{m \times n}$, the column rank of A equals to the row rank of A, denoted as $\text{rank}(A)$.

Basic Properties of the Rank

- For $A \in \mathbb{R}^{m \times n}, \operatorname{rank}(A) \leq \min(m, n)$. If $\operatorname{rank}(A) = \min(m, n)$, then A is said to be full rank.

- For $A \in \mathbb{R}^{m \times n}, \operatorname{rank}(A) = \operatorname{rank}\left(A^T\right)$.

- For $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}, \operatorname{rank}(AB) \leq \min(\operatorname{rank}(A), \operatorname{rank}(B))$

- For $A, B \in \mathbb{R}^{m \times n}, \operatorname{rank}(A + B) \leq \operatorname{rank}(A) + \operatorname{rank}(B)$

Outline

Definition of the Inverse

- The **inverse** of a square matrix $A \in \mathbb{R}^{n \times n}$ is denoted A^{-1}, and is the unique matrix such that

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

- Note: Not all matrices have inverses.
  - Non-square matrices do not have inverses by definition.
  - For some square matrices A, A^{-1} may not exist. In this case,we say that A is **non-invertible** or **singular**.
- In order for a square matrix A to have an inverse A^{-1}, then A must be full rank.

Properties of the inverse

Assume that $A, B \in \mathbb{R}^{n\times n}$ are non-singular.

- $\left(A^{-1}\right)^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $\left(A^{-1}\right)^{T} = \left(A^{T}\right)^{-1}$.
  For this reason this matrix is often denoted A^{-T}.

Example: Consider the linear system of equations, $Ax = b$ where $A \in \mathbb{R}^{n\times n}$, and $x, b \in \mathbb{R}^{n}$. If A is nonsingular (i.e., invertible), then $x = A^{-1}b$.

Outline

Orthogonal Matrices

- **Orthogonal of vectors**:
  Two **vectors** $x, y \in \mathbb{R}^n$ are orthogonal if $x^T y = 0$.

- **Normalized**: A vector $x \in \mathbb{R}^n$ is normalized if $\|x\|_2 = 1$.
  (the Euclidean or ℓ_2 norm: $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{x^T x}$)

- **Orthogonal of matrices**:
  A **square matrix** $U \in \mathbb{R}^{n \times n}$ is orthogonal if all its columns are orthogonal to each other and are normalized. (the columns are then referred to as being **orthonormal**).

Orthogonal Matrices

- From the definition of orthogonality and normality, we have

$$U^T U = I = U U^T$$

  In other words, the inverse of an orthogonal matrix is its transpose.

- Note that if U is not square - i.e., $U \in \mathbb{R}^{m \times n}, \quad n < m$ - but its columns are still orthonormal, then $U^T U = I$, but $U U^T \neq I$. We generally only use the term orthogonal to describe the case where U is square.
- $\|Ux\|_2 = \|x\|_2$ for any $x \in \mathbb{R}^n, U \in \mathbb{R}^{n \times n}$ orthogonal.

Outline

Span and Projection

- The **span** of a set of vectors $\{x_1, x_2, \dots x_n\}$ is the set of all vectors that can be expressed as a linear combination of $\{x_1, \dots, x_n\}$. That is,

$$\text{span}\left(\{x_1, \dots x_n\}\right) = \left\{ v : v = \sum_{i=1}^{n} \alpha_i x_i, \quad \alpha_i \in \mathbb{R} \right\}$$

- The **projection** of a vector $y \in \mathbb{R}^m$ onto $\text{span}\left(\{x_1, \dots x_n\}\right)$ (each $x_i \in \mathbb{R}^m$) is the vector $v \in \text{span}\left(\{x_1, \dots x_n\}\right)$, such that v is as close as possible to y, as measured by the Euclidean norm $\|v - y\|_2$. We define it formally as,

$$\text{Proj}\left(y; \{x_1, \dots x_n\}\right) = \text{argmin}_{v \in \text{span}(\{x_1, \dots, x_n\})} \|y - v\|_2$$

Range

- The **range** (sometimes also called the **column space**) of a matrix $A \in \mathbb{R}^{m \times n}$, denoted $\mathcal{R}(A)$, is the the span of the columns of A:

$$\mathcal{R}(A) = \{v \in \mathbb{R}^m : v = Ax, x \in \mathbb{R}^n\}$$

- Assume A is full rank and $n < m$, the projection of a vector $y \in \mathbb{R}^m$ onto the range of A is given by,

$$\mathrm{Proj}(y; A) = \mathrm{argmin}_{v \in \mathcal{R}(A)} \|v - y\|_2 = A\left(A^T A\right)^{-1} A^T y$$

- When A contains only a single column, $a \in \mathbb{R}^m$, this gives the special case for a projection of a vector on to a line:

$$\mathrm{Proj}(y; a) = \frac{aa^T}{a^T a} y$$

Nullspace

- The **nullspace** of a matrix $A \in \mathbb{R}^{m \times n}$, denoted $\mathcal{N}(A)$ is the set of all vectors that equal 0 when multiplied by A, i.e.,

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$$

- Vectors in $\mathcal{R}(A)$ are of size m, while vectors in the $\mathcal{N}(A)$ are of size n, so vectors in $\mathcal{R}\left(A^T\right)$ and $\mathcal{N}(A)$ are both in $\mathbb{R}^n$.

$$\left\{w : w = u + v, u \in \mathcal{R}\left(A^T\right), v \in \mathcal{N}(A)\right\} = \mathbb{R}^n$$

$$\mathcal{R}\left(A^T\right) \cap \mathcal{N}(A) = \{\mathbf{0}\}.$$

Nullspace

- $\mathcal{R}\left(A^T\right)$ and $\mathcal{N}(A)$ are disjoint subsets that together span the entire space of $\mathbb{R}^n$.

- Sets of this type are called **orthogonal complements**, and we denote this $\mathcal{R}\left(A^T\right) = \mathcal{N}(A)^\perp$

Outline

Several Properties of the Determinant

- For $A \in \mathbb{R}^{n \times n}, |A| = \left|A^T\right|$.
- For $A, B \in \mathbb{R}^{n \times n}, |AB| = |A||B|$.
- For $A \in \mathbb{R}^{n \times n}, |A| = 0$ if and only if A is singular (i.e., non-invertible).
- For $A \in \mathbb{R}^{n \times n}$ and A is non-singular, $\left|A^{-1}\right| = 1/|A|$.

Outline

Quadratic Forms

Given a square matrix $A \in \mathbb{R}^{n \times n}$ and a vector $x \in \mathbb{R}^n$, the scalar value $x^T A x$ is called a **quadratic form**.

$$x^T A x = \sum_{i=1}^{n} x_i (Ax)_i = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} A_{ij} x_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j .$$

$$x^T A x = \left(x^T A x \right)^T = x^T A^T x = x^T \left(\frac{1}{2} A + \frac{1}{2} A^T \right) x$$

(Hint for 1st equality: the transpose of a scalar is equal to itself.)

Only the symmetric part of A contributes to the quadratic form and we often assume that the matrices in a quadratic form are symmetric.

Positive Semidefinite Matrices

- A symmetric matrix $A \in \mathbb{S}^n$ is **positive definite** (PD) if for all non-zero vectors $x \in \mathbb{R}^n, x^T Ax > 0$.

  This is denoted $A \succ 0$ (or just $A > 0$). The set of all positive definite matrices is denoted $\mathbb{S}^n_{++}$.

- A symmetric matrix $A \in \mathbb{S}^n$ is **positive semidefinite** (PSD) if for all vectors $x^T Ax \geq$ 0 .

  This is written $A \succeq 0$ (or just $A \geq 0$), and the set of all positive semidefinite matrices is often denoted $\mathbb{S}^n_+$.

Negative Semidefinite Matrices

- A symmetric matrix $A \in \mathbb{S}^n$ is **negative definite** (ND), denoted $A \prec 0$ (or just $A < 0$) if for all non-zero $x \in \mathbb{R}^n, x^T A x < 0$.
- A symmetric matrix $A \in \mathbb{S}^n$ is **negative semidefinite** (NSD), denoted $A \preceq 0$ (or just $A \leq 0$) if for all $x \in \mathbb{R}^n, x^T A x \leq 0$.
- Finally, a symmetric matrix $A \in \mathbb{S}^n$ is **indefinite**, if it is neither positive semidefinite nor negative semidefinite - i.e., if there exists $x_1, x_2 \in \mathbb{R}^n$ such that $x_1^T A x_1 > 0$ and $x_2^T A x_2 < 0$

Properties of Positive Definite Matrices

- Positive definite and negative definite matrices are always full rank, and hence, invertible.

- Given any matrix $A \in \mathbb{R}^{m \times n}$ (not necessarily symmetric or even square), the Gram matrix $G = A^T A$ is always positive semidefinite.

- Further, if $m \geq n$ (and we assume for convenience that A is full rank), then $G = A^T A$ is positive definite.

Outline

Eigenvalues and Eigenvectors

Given a square matrix $A \in \mathbb{R}^{n \times n}$, we say that $\lambda \in \mathbb{C}$ is an **eigenvalue** of A and $x \in \mathbb{C}^n$ is the corresponding **eigenvector** if

$$Ax = \lambda x, \quad x \neq 0$$

- Let $\lambda_1, \ldots, \lambda_n$ are all the eigenvalues of the matrix A, they may not be distinct.
- To find the eigenvector corresponding to the eigenvalue λ_i, we solve the linear equation $(\lambda_i I - A)\, x = 0$, which is guaranteed to have a non-zero solution because $\lambda_i I - A$ is singular.
- We usually assume that the eigenvector is normalized to have length 1.

Properties of Eigenvalues and Eigenvectors

Assume $A \in \mathbb{R}^{n \times n}$ has eigenvalues

- The trace of a A is equal to the sum of its eigenvalues,

$$\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i$$

- The determinant of A is equal to the product of its eigenvalues,

$$|A| = \prod_{i=1}^{n} \lambda_i$$

- The rank of A is equal to the number of non-zero eigenvalues of A.

Properties of Eigenvalues and Eigenvectors

Assume $A \in \mathbb{R}^{n \times n}$ has eigenvalues

- Suppose A is non-singular with eigenvalue λ and an associated eigenvector x.
  Then $1/\lambda$ is an eigenvalue of A^{-1} with an associated eigenvector x, i.e., $A^{-1}x = (1/\lambda)x$.

- The eigenvalues of a diagonal matrix $D = \operatorname{diag}(d_1, \ldots d_n)$ are just the diagonal entries $d_1, \ldots d_n$.

Outline

Properties

Assume A is a symmetric real matrix.

- All eigenvalues of A are real numbers. Denote them by $\lambda_1, \ldots, \lambda_n$.
- There exists a set of eigenvectors $u_1, \ldots, u_n$ such that

  a) for all i, u_i is an eigenvector with eigenvalue λ_i

  b) $u_1, \ldots, u_n$ are unit vectors and orthogonal to each other.

Diagonalization of the matrix

Let U be the orthonormal matrix that contains u_i's as columns:

$$U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}$$

Let $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ be the diagonal matrix and we can verify

$$AU = \begin{bmatrix} | & | & & | \\ Au_1 & Au_2 & \cdots & Au_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1 u_1 & \lambda_2 u_2 & \cdots & \lambda_n u_n \\ | & | & & | \end{bmatrix}$$

$$= U \operatorname{diag}(\lambda_1, \ldots, \lambda_n) = U\Lambda$$

Diagonalization of the matrix

Recalling that orthonormal matrix U satisfies that $UU^T = I$, we have

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

- This new presentation of A as $U\Lambda U^T$ is often called the **diagonalization** of the matrix A.
- We can often effectively treat a symmetric matrix A as a diagonal matrix w.r.t the basis defined by the eigenvectors U.

Background: representing vector w.r.t. another basis

Any orthonormal matrix $U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}$ defines a new basis (coordinate system) of $\mathbb{R}^n$ in the following sense:

For any vector $x \in \mathbb{R}^n$ can be represented as a linear combination of $u_1, \ldots, u_n$ with coefficient $\hat{x}_1, \ldots, \hat{x}_n$:

$$x = \hat{x}_1 u_1 + \cdots + \hat{x}_n u_n = U\hat{x}$$

Background: representing vector w.r.t. another basis

Indeed, such $\hat{x}$ uniquely exists

$$x = U\hat{x} \Leftrightarrow U^T x = \hat{x}$$

In other words, the vector $\hat{x} = U^T x$ can serve as another representation of the vector x w.r.t the basis defined by U.

"Diagonalizing" quadratic form

The quadratic form $x^T A x$ can also be simplified under the new basis

$$x^T A x = x^T U \Lambda U^T x = \hat{x}^T \Lambda \hat{x} = \sum_{i=1}^{n} \lambda_i \hat{x}_i^2$$

- With the old representation, $x^T A x = \sum_{i=1,j=1}^{n} x_i x_j A_{ij}$ involves a sum of n^2 terms instead of n terms in the equation above.

- With this viewpoint, we can also show that the definiteness of the matrix A depends entirely on the sign of its eigenvalues.

"Diagonalizing" quadratic form

- If all $\lambda_i > 0$, then the matrix A is positive definite because $x^T A x = \sum_{i=1}^{n} \lambda_i \hat{x}_i^2 > 0$ for any $\hat{x} \neq 0$.

- If all $\lambda_i \geq 0$, it is positive semidefinite because $x^T A x = \sum_{i=1}^{n} \lambda_i \hat{x}_i^2 \geq 0$ for all $\hat{x}$.

- Likewise, if all $\lambda_i < 0$ or $\lambda_i \leq 0$, then A is negative definite or negative semidefinite respectively.

- Finally, if A has both positive and negative eigenvalues, say $\lambda_i > 0$ and $\lambda_j < 0$, then it is indefinite.