Machine Learning

CSE 6363 (Fall 2019)

Lecture 2 Linear Algebra Review

Dajiang Zhu, Ph.D.
Department of Computer Science and Engineering

Most slides of this lecture courtesy: Dr. Heng Huang
Others courtesy: Dr. Kolter

Linear Algebra Review

- Why?
  - Solve linear equations?
  - Everything is about "Matrix"

- Example:

$$5x_1 + 5x_2 = 15$$
$$-x_1 + 7x_2 = 13$$

With matrix notation, we can write the above system as:

$$A\,x = B$$

$$A = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix}, B = \begin{bmatrix} 15 \\ 13 \end{bmatrix}$$

Linear Algebra Review

- Basic Notation
  - $A \in \mathbb{R}^{m \times n}$
  - $x \in \mathbb{R}^n$
  - x^T

- Matrix Multiplication

$$A \in \mathbb{R}^{m \times n} \text{ and } B \in \mathbb{R}^{n \times p}$$

$$C = AB \in \mathbb{R}^{m \times p}$$

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

Linear Algebra Review

- Vector-Vector Products
  - Inner product or dot product

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ x_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

  - Outer product

$$xy^T \in \mathbb{R}^{m \times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{bmatrix}$$

Linear Algebra Review

- How to understand "Matrix-Matrix products"

  ✓ "Rows-Columns"

$$C = AB = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_p \end{bmatrix}$$

  ✓ "Columns-Rows"

$$C = AB = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} — & b_1^T & — \\ — & b_2^T & — \\ & \vdots & \\ — & b_n^T & — \end{bmatrix} = \sum_{i=1}^{n} a_i b_i^T$$

Linear Algebra Review

- How to understand "Matrix-Matrix products"

  ✓ "Matrix-Columns"

$$C = AB = A\begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ Ab_1 & Ab_2 & \cdots & Ab_p \\ | & | & & | \end{bmatrix}$$

  ✓ "Rows-Matrix"

$$C = AB = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} B = \begin{bmatrix} — & a_1^T B & — \\ — & a_2^T B & — \\ & \vdots & \\ — & a_m^T B & — \end{bmatrix}$$

Linear Algebra Review

- Basic properties of matrix multiplication
  - Associative: $(AB)C = A(BC)$
  - Distributive: $A(B+C) = AB + AC$
  - NOT Commutative: $AB \neq BA$

Linear Algebra Review

- Identity matrix

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- Diagonal matrix

$$D_{ij} = \begin{cases} d_i & i = j \\ 0 & i \neq j \end{cases}$$

- Transpose

$$(A^T)^T = A$$

$$(AB)^T = B^T A^T$$

$$(A + B)^T = A^T + B^T$$

Linear Algebra Review

- Symmetric matrix $A = A^T$
- Trace $\mathrm{tr}A = \sum_{i=1}^{n} A_{ii}$

For $A \in \mathbb{R}^{n\times n}$, $\mathrm{tr}A = \mathrm{tr}A^T$.

For $A, B \in \mathbb{R}^{n\times n}$, $\mathrm{tr}(A+B) = \mathrm{tr}A + \mathrm{tr}B$.

For $A \in \mathbb{R}^{n\times n}$, $t \in \mathbb{R}$, $\mathrm{tr}(tA) = t\,\mathrm{tr}A$.

For A, B such that AB is square, $\mathrm{tr}AB = \mathrm{tr}BA$.

For A, B, C such that ABC is square, $\mathrm{tr}ABC = \mathrm{tr}BCA = \mathrm{tr}CAB$, and so on for the product of more matrices.

Linear Algebra Review

- Norm
  - A realy value, e.g. $\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$

1. For all $x \in \mathbb{R}^n$, $f(x) \geq 0$ (non-negativity).
2. $f(x) = 0$ if and only if $x = 0$ (definiteness).
3. For all $x \in \mathbb{R}^n$, $t \in \mathbb{R}$, $f(tx) = |t| f(x)$ (homogeneity).
4. For all $x, y \in \mathbb{R}^n$, $f(x+y) \leq f(x) + f(y)$ (triangle inequality)

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \qquad \|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2} = \sqrt{\operatorname{tr}(A^T A)}$$

Linear Algebra Review

- L-0 Norm

$$\|x\|_0 = \sqrt[0]{\sum_i x_i^0} \Rightarrow \|x\|_0 = \#\, of\ (i \mid x_i \neq 0)$$

- L-1 Norm (Manhattan norm)

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

  – *Sum of Absolute Difference (SAD)*

$$SAD\,(x_1, x_2) = \|x_1 - x_2\|_1 = \sum |x_1 - x_2|$$

  – *Signal difference measurement - Mean-Absolute Error (MAE)*

Linear Algebra Review

- L-2 Norm (Euclidean norm)

$$\|x\|_2 = \sqrt[2]{\sum_i x_i^2}$$

  – *Sum of Squared Difference (SSD)*

$$SSD\ (x_1, x_2) = \|x_1 - x_2\|_2^2 = \sum_i (x_{1_i} - x_{2_i})^2$$

  – *Mean-Squared Error (MSE)*

- L-∞ Norm

$$\|x\|_\infty = \max_i |x_i|$$

Linear Algebra Review

- Linear dependent
  - If $x_n = \sum_{i=1}^{n-1} a_i x_i$, for some scalar values a_1, a_2, …, $a_{n-1} \in \mathbb{R}$, we say that x_1, …, x_n are linearly dependent; otherwise, those vectors are linearly independent.
- Rank
  - Size of the largest subset of rows/columns that constitute a linearly independent set is row/column rank.

Linear Algebra Review

❖ About 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 $A \in \mathbb{R}^{m \times n}$, $\operatorname{rank}(A) = \operatorname{rank}(A^T)$.
- 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 $A, B \in \mathbb{R}^{m \times n}$, $\operatorname{rank}(A + B) \leq \operatorname{rank}(A) + \operatorname{rank}(B)$.

Linear Algebra Review

- Inverse: invertible or non-singular

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

$$(A^{-1})^{-1} = A$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

$$(A^{-1})^T = (A^T)^{-1}.$$

Linear Algebra Review

- Orthogonal matrix

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

  - U $\in \mathbb{R}^{n\times n}$ is orthogonal if all its columns are orthogonal to each other and normalized.

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & \sqrt{2} & -\sqrt{2} \\ -1 & 1 & 1 \end{bmatrix}$$

  - $U^T U = I = U U^T$
  - $\|Ux\|_2 = \|x\|_2$

Linear Algebra Review

- Determinant

$$A = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \qquad |A| = ?$$

- **Properties of Determinants**
  - The determinant is a real number and can be negative
  - The determinant only exists for square
  - The inverse of a matrix will exist only if the determinant is not zero

- **The determinant of a matrix will be zero if**
  - An entire row is zero.
  - Two rows or columns are equal.
  - A row or column is a constant multiple of another row or column

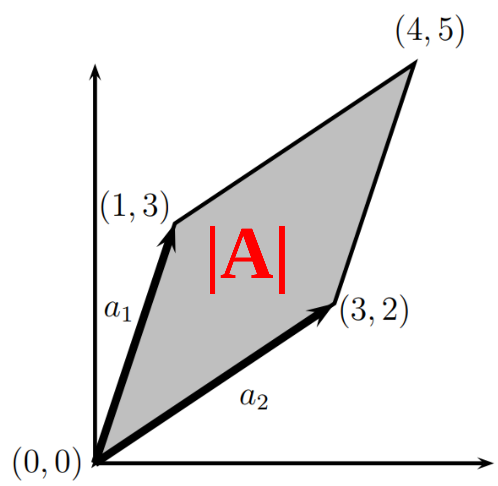
- **The determinant is the "size" of the output transformation**

Linear Algebra Review

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix} \begin{array}{l} \textbf{v1} \\ \textbf{v2} \end{array}$$

$|A| = ?$

$$S = \alpha\, v1 + \beta\, v2,\ \alpha \in [0,1],\ \beta \in [0,1]$$



Linear Algebra Review

- Quadratic form: A$\in \mathbb{R}^{n\times n}, x \in \mathbb{R}^n$

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

- Positive definite: A$\in\mathbb{S}^n$, for all non-zero vectors $x \in \mathbb{R}^n, x^T A x > 0.$
- Positive semidefinite: A$\in\mathbb{S}^n$, for all non-zero vectors $x \in \mathbb{R}^n, x^T A x \geq 0.$
- Negative definite: A$\in\mathbb{S}^n$, for all non-zero vectors $x \in \mathbb{R}^n, x^T A x < 0.$
- Negative semidefinite: A$\in\mathbb{S}^n$, for all non-zero vectors $x \in \mathbb{R}^n, x^T A x \leq 0.$

❖ One important property of positive definite and negative definite matrices is that they are always full rank, and hence, invertible.

Linear Algebra Review

- Gram matrix
  - Positive semidefinite $G = A^T A$
- Eigenvectors and eigenvalues $Ax = \lambda x, \quad x \neq 0$
- The trace of a A is equal to the sum of its eigenvalues,

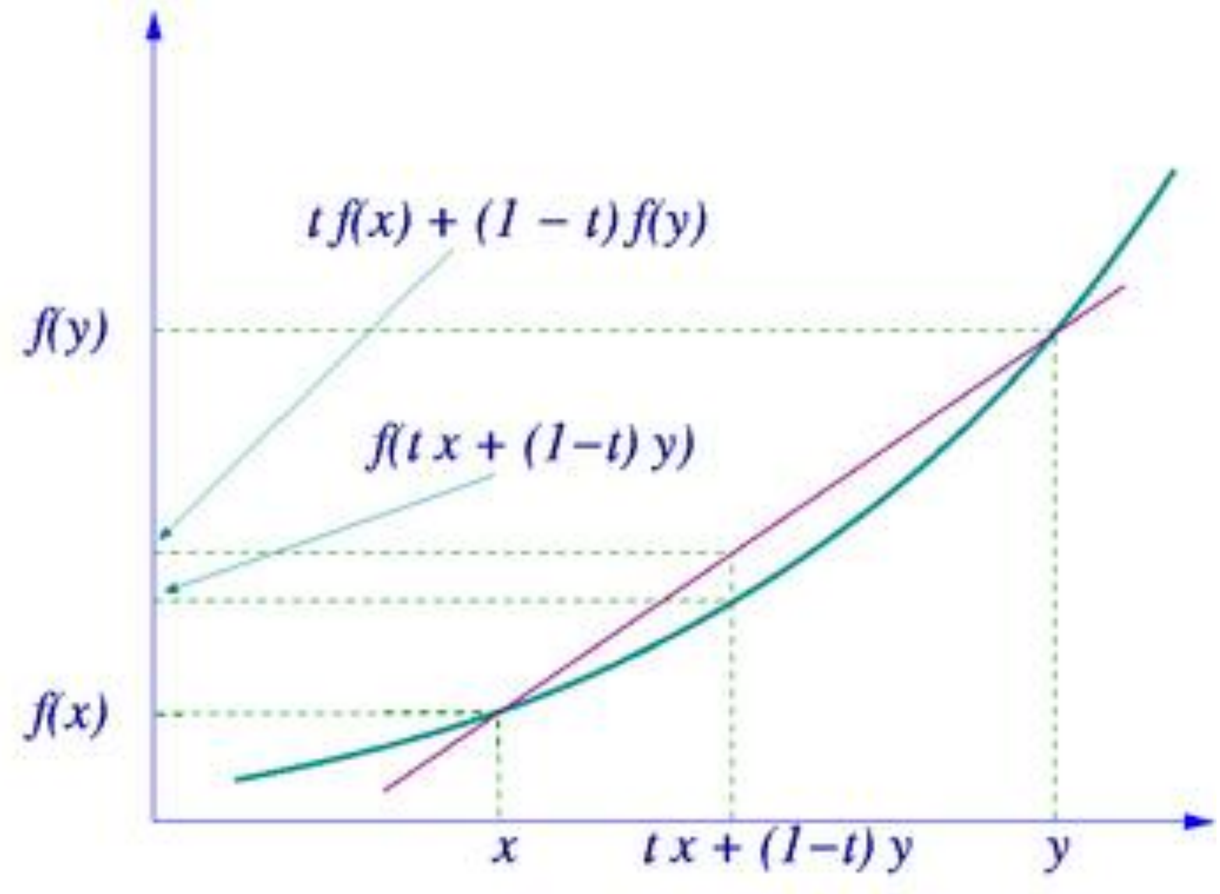
$$\mathrm{tr}A = \sum_{i=1}^{n} \lambda_i.$$

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

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

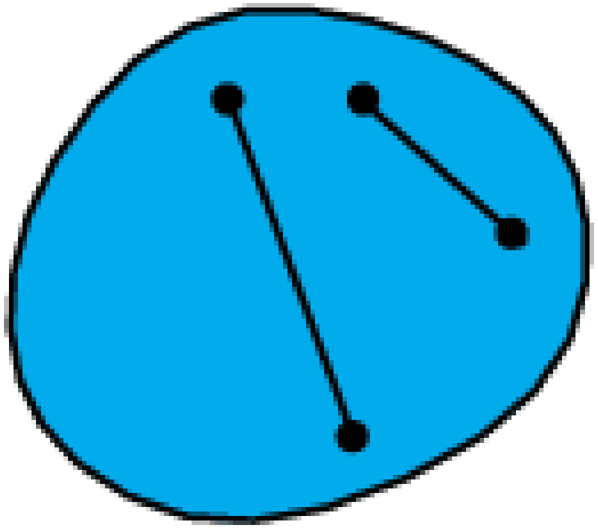
- The rank of A is equal to the number of non-zero eigenvalues of A.
- If A is non-singular then $1/\lambda_i$ is an eigenvalue of A^{-1} with associated eigenvector x_i, i.e., $A^{-1}x_i = (1/\lambda_i)x_i$.
- The eigenvalues of a diagonal matrix $D = \mathrm{diag}(d_1, \ldots d_n)$ are just the diagonal entries $d_1, \ldots d_n$.

Convex Function

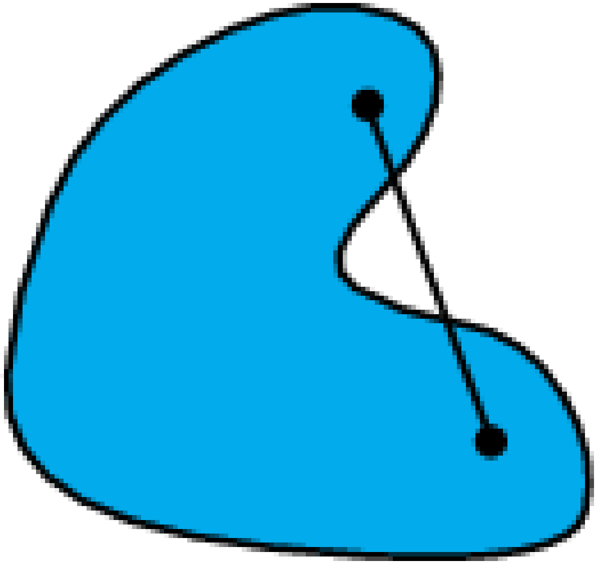


$$f(t\,x + (1\text{-}t)\,y) <= t\,f(x) + (1\text{-}t)\,f(y)$$

Convex Set



convex



concave

Region above a convex function is a convex set.

Convex Programming

- Convex optimization function
- Convex feasible region
- Why is it so important ???
    - Global optimum can be found in polynomial time.
  - Many practical problems are convex
    - Non-convex problems can be relaxed to convex ones.