Machine Learning

CSE 6363 (Fall 2016)

Lecture 8 Linear Algebra Review

Heng Huang, Ph.D.
Department of Computer Science and Engineering

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

- 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 $$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

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)

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

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2} = \sqrt{\operatorname{tr}(A^T A)}$$

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

Linear Algebra Review

- Inverse: invertible or non-singular

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

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

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

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

- Orthogonal matrix

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

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

$$\|Ux\|_2 = \|x\|_2$$

Linear Algebra Review

- Quadratic form

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

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

i.e., only the symmetric part of A contributes to the quadratic form. For this reason, we often implicitly assume that the matrices appearing in a quadratic form are symmetric.

- Positive definite, positive semidefinite, negative definite

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 $D = \mathrm{diag}(d_1, \dots d_n)$ are just the diagonal entries $d_1, \dots d_n$.

Linear Algebra Review

- Derivatives of matrix

$$\frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{a}}{\partial \mathbf{X}} = \frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{a}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T$$

Derivatives of trace

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}\mathbf{A}) = \mathbf{A}^T$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}) = \mathbf{A}^T\mathbf{B}^T$$

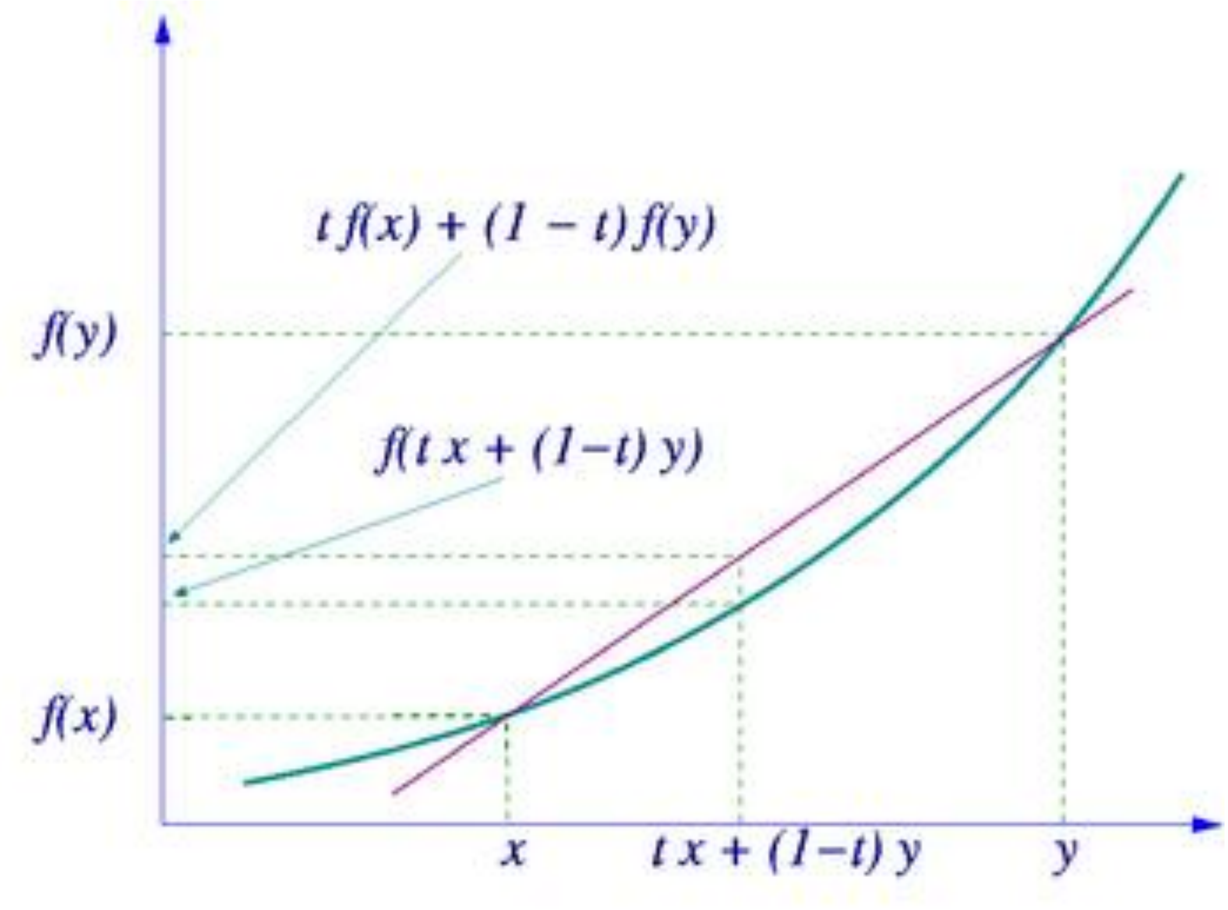
$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{A}\mathbf{X}^T\mathbf{B}) = \mathbf{B}\mathbf{A}$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}^T\mathbf{A}) = \mathbf{A}$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{A}\mathbf{X}^T) = \mathbf{A}$$

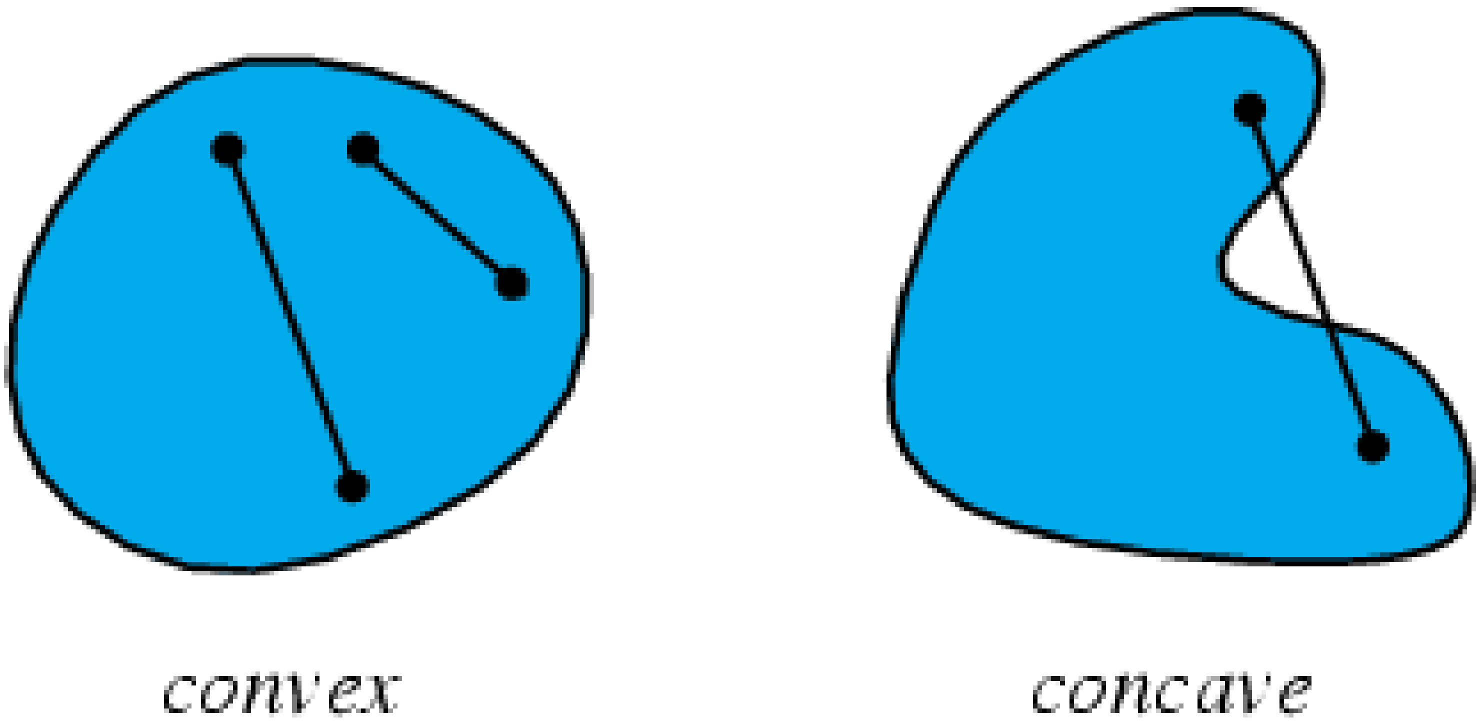
- Exercise

Convex Function



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

Convex Set



Region above a convex function is a convex set.

Programming

- Objective function to be minimized/maximized.
- Constraints to be satisfied.

Example

Objective function

$$
\begin{aligned}
\text{maximize} \quad & z = x_1 + x_2 \\
\text{subject to} \quad & 4x_1 - x_2 \leq 8 \\
& 2x_1 + x_2 \leq 10 \\
& 5x_1 - 2x_2 \geq -2 \\
& x_1, x_2 \geq 0
\end{aligned}
$$

Constraints

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.

Quadratic Programming

- QP Problem:

$$Ax \le b$$
$$Gx = h$$
$$J = \frac{1}{2} x^T H x + f^T x \to \min$$

- Matlab Optimization Toolbox: **`QUADPROG`**
- Same feasibility issues as for LP