Math Review

Peifan Wu

September 10, 2020

1 Basic Linear Algebra

1.1 Definitions of Vectors and Matrices

Vector $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$ is a column vector.

The transpose of x: $x^T = x' = [x_1, x_2, \dots, x_n]$ is a row vector.

Matrix $A = \begin{bmatrix} A_{11} & \dots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nm} \end{bmatrix}$ is a $n \times m$ matrix. Each element of the matrix is represented by $A_{ij}, i = 1, \dots, n, j = 1, \dots, m$.

The transpose of matrix A is denoted as A^T, A' where $A_{ij}^T = A_{ji}$. Notice that $\left(A^T\right)^T = A$.

1.2 Matrix Operations

Addition/Subtraction Addition/subtraction of matrices are applied element-wise. If A, B are both $n \times m$ matrices then

$$A + B = \begin{bmatrix} A_{11} + B_{11} & \dots & A_{1m} + B_{1m} \\ \vdots & \ddots & \vdots \\ A_{n1} + B_{n1} & \cdots & A_{nm} + B_{nm} \end{bmatrix}$$
$$A - B = \begin{bmatrix} A_{11} - B_{11} & \dots & A_{1m} - B_{1m} \\ \vdots & \ddots & \vdots \\ A_{n1} - B_{n1} & \cdots & A_{nm} - B_{nm} \end{bmatrix}$$

Properties:

- Commutativity: $A + B = B + A$
- Associativity: $(A + B) + C = A + (B + C)$

Multiplication For matrices A with size $n \times m$ and B with size $m \times p$, we have $C = A \times B$ where C is a $n \times p$ matrix,

$$C_{ik} = \sum_{j=1}^{n} A_{ij} B_{jk}$$

Matrix multiplication is not commutative!
Properties:

- $(AB)^T = B^T A^T$
- Associativity: $(AB)\,C = A\,(BC)$
- Distributivity: $(A+B)\,C = AC + BC$, $C\,(A+B) = CA + CB$
- Communtativity only holds for scalar multiplication, but not for matrix multiplication:

$$\alpha A = A\alpha$$
$$AB \neq BA$$

- The product of a matrix and a vector can be dot product.

Inverse We first define Identity matrix:

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

- $n \times n$ square matrix with "1"s on the diagonal, and "0"s everywhere
- $\forall A$ that is a $n \times n$ matrix, $A \times I = I \times A = A$

Then we define the inverse of a $n \times n$ square matrix A.

- A $n \times n$ matrix, denoted by A^{-1}, is called the inverse of A if $A^{-1}A = I$
- Easy to prove $AA^{-1} = I$
- Such A^{-1} doesn't always exist. When it exists we call A invertible or non-singular. Otherwise, A is singular

1.3 Matrix and Linear Equations

A linear equation system can be expressed as $A \cdot x = b$ where A is a $m \times n$ matrix, x is a column vector of n elements, and b is a column vector of m elements.

- If $n = m$ (then A is a square matrix) and A is invertible (the rank of A equals n), then the solution to the linear equation system is $x = A^{-1}b$ by left multiplying A^{-1} to both sides of the equation

- If $n > m$, the number of unknowns x is more than the number of equations m. If the rank of A is greater than m as well, then x has multiple solutions

- If $n < m$, the number of unknowns x is less than the number of equations m. In general there's no solution. However, we can choose x subject to a loss function,

$$\min_{x} ||Ax - b||$$

  gives $x = \left(A^T A\right)^{-1} A^T b$. A way to interpret the OLS regression.

2 Basic Optimization

2.1 Unconstrained Optimization

Assume x is a vector with n elements, and f is a function $\mathbb{R}^n \to \mathbb{R}$. We want x such that

$$\min_{x} f(x)$$

If f is differentiable then the first-order necessary condition is

$$\partial_i f(x) = f_{x_i}(x) = 0$$

i.e., the partial derivative of f with respect to any element in x is 0. This is only a necessary condition: you need to verify the point is local minimum instead of local maximum.

2.2 Constrained Optimization

Assume x is a vector with n elements. Almost all the constrained problems can be reduced to the same form,

$$\begin{aligned} &\min_{x} f(x) \\ s.t.\, &g(x) \leqslant 0 \\ &h(x) = 0 \end{aligned}$$

- $f : \mathbb{R}^n \to \mathbb{R}$, the objective function

- $g : \mathbb{R}^n \to \mathbb{R}^m$, constraints may or may not bind

- $h : \mathbb{R}^n \to \mathbb{R}^p$, constraints that are always binding

We use the Lagrangian to solve the optimization problem.

First, the definition of Lagragian,

$$\mathcal{L} = f(x) + \mu g(x) + \lambda h(x)$$

where μ and λ are called Lagrange multipliers. These are also the so called "shadow prices" under some economic setting. The values of μ and λ shows how tight the constraints are.

Then we write the first-order necessary condition, $\partial\mathcal{L}(x) = 0$. i.e.,

$$\begin{aligned} &\partial f(x) + \mu \partial g(x) = 0 \\ &\qquad s.t. \begin{cases} \mu > 0, g(x) = 0 & g \text{ constraints binding} \\ \mu = 0, g(x) < 0 & g \text{ constraints not binding} \end{cases} \end{aligned}$$

Check Kuhn-Tucker Theorem for more details.