Linear Algebra, Calculus, Optimization Review

Notation

$\mathbf{x}$ is an N-dimensional vector. We denote vectors with bold, lower-case letters.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_N \end{bmatrix}$$

$\mathbf{A}$ is an $N \times M$ matrix. We denote matrices with bold, upper-case letters.

$$\mathbf{A} = \begin{bmatrix} a_{11}, a_{12}, \dots, a_{1M} \\ a_{21}, a_{22}, \dots, a_{2M} \\ \dots \\ a_{N1}, a_{N2}, \dots, a_{NM} \end{bmatrix}$$

Let $\mathbf{A}[i,:]$ denote the i^{th} row and $\mathbf{A}[:,j]$ denote the j^{th} column of $\mathbf{A}$.

Linear Algebra

Matrix Multiplication

Matrix Multiplication is defined for two matrices, $\mathbf{A} \in \mathbb{R}^{N \times M}$ and $\mathbf{B} \in \mathbb{R}^{M \times D}$.

$$\mathbf{AB} = \mathbf{C}$$

where $\mathbf{C} \in \mathbb{R}^{N \times D}$. *Note that $\mathbf{BA}$ is **not** defined, as the inner dimensions of the matrices do not match.*

$$C_{ij} = \sum_{m=1}^{M} a_{im} * b_{mj}$$

This is equivalent to

$$C_{ij} = \mathbf{A}[i,:] \cdot \mathbf{B}[:,j]$$

Properties of Matrix Multiplication [1]

- Matrix Multiplication is associative: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
- Matrix Multiplication is distributive: $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$
- In general, Matrix Multiplication is **not** commutative. It is possible that $\mathbf{AB} \neq \mathbf{BA}$.

The Identity Matrix

The identity matrix is a square matrix, whose diagonal elements are 1, and whose off-diagonal elements are 0.

$$\mathbf{I} = \begin{bmatrix} 1, 0, \ldots, 0 \\ 0, 1, \ldots, 0 \\ \vdots, \ddots, \vdots \\ 0, 0, \ldots, 1 \end{bmatrix}$$

Properties

$$\mathbf{AI} = \mathbf{A} = \mathbf{IA}$$

The Transpose

The transpose of a matrix is obtained by switching the column and row indices of a matrix. ("Flipping the matrix over its diagonal.")[2]. For $\mathbf{A} \in \mathbb{R}^{N \times M}$, $\mathbf{A}^{\mathbf{T}} \in \mathbb{R}^{M \times N}$

$$A_{ij} = A_{ji}^{T}$$

Properties [1]

- $(\mathbf{A}^T)^T = \mathbf{A}$
- $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

- $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$

Vector Derivatives

Let $f(\mathbf{w})$ be a function of a vector. The first derivative, or gradient, of $f(\mathbf{w})$ is computed as:

$$\nabla_{\mathbf{w}} f(\mathbf{w}) = \begin{bmatrix} \frac{\partial f(\mathbf{w})}{\partial w_1} \\ \frac{\partial f(\mathbf{w})}{\partial w_2} \\ \vdots \\ \frac{\partial f(\mathbf{w})}{\partial w_N} \end{bmatrix}$$

Identities of Vector Derivatives [3]

1. $$\frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$
2. $$\frac{\partial \mathbf{x}^T \mathbf{B} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B} + \mathbf{B}^T)\mathbf{x}$$

Optimization

Find an extreme value (minimum or maximum) of a function.

Methods for optimizing a function:

1. Take the derivative. Solve for zero. (closed-form solution)
2. Iteratively update your solution, move along the gradient. (Gradient Descent)

Lagrange Multipliers: Constrained Optimization

A method for optimizing a function $f(\mathbf{w})$, subject to a constraint $g(\mathbf{w}) = k$.
If an optimal point in f is at $\mathbf{w_o}$ then the gradients of f and g at point $\mathbf{w_o}$ must be parallel:

$$\nabla f(\mathbf{w_o}) = \lambda \nabla g(\mathbf{w_o})$$

where λ a constant, called the *Lagrange multiplier*.

Practice Problems

1. For $\mathbf{x} \in \mathbb{R}^{N \times 1}$ and $\mathbf{a} \in \mathbb{R}^{N \times 1}$, show that

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

2. Find

$$\frac{\partial}{\partial \mathbf{c}} (\mathbf{a} + \mathbf{c}^T (\mathbf{B}\mathbf{D}^T \mathbf{E})^T + \mathbf{c}^T \mathbf{B}\mathbf{E}\mathbf{c})$$

3. Find local extrema of

$$\frac{d}{dx} |x^2 - 3|$$

4. Using a Lagrange multiplier: A rectangular box without a lid is to be made from 12 m^2 of cardboard. Find the maximum volume of such a box.

Resources

- Matrix Cookbook
- Stanford CS229: Linear Algebra Review

References

[1] Stanford CS229: Linear Algebra Review
[2] Wikipedia: Matrix Transpose
[3] Matrix Cookbook