CS-563 Deep Learning

Matrix and Vector Calculus

Nazar Khan
Department of Computer Science
University of the Punjab

Matrix Calculus

- Specialised notation for multivariate calculus.
- Simplifies operations such as finding the minimum of a multivariate function.
- Different conventions exist. You may choose any as long as you remain consistent.
- Purpose of these slides is to set the convention for the rest of the course.

Notation

- Scalars are denoted by lower-case letters like s, a, b.
- Vectors are denoted by lower-case bold letters like $\mathbf{x}, \mathbf{y}, \mathbf{v}$.
- Matrices are denoted by upper-case bold letters like $\mathbf{M}, \mathbf{D}, \mathbf{A}$.
- Any vector $\mathbf{x} \in \mathbb{R}^d$ is by default a column vector.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

- The corresponding row vector is obtained as $\mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & \dots & x_d \end{bmatrix}$.

Vectors

For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and $\mathbf{z} \in \mathbb{R}^k$

- *Inner product* $\mathbf{x}^T\mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_d y_d$ is a scalar value. Also called *dot product* or *scalar product*.
- Other representations: $\mathbf{x} \cdot \mathbf{y}$, $(\mathbf{x}, \mathbf{y})$ and $< \mathbf{x}, \mathbf{y} >$.
- Represents similarity of vectors.
  - If $\mathbf{x}^T\mathbf{y} = 0$, then $\mathbf{x}$ and $\mathbf{y}$ are orthogonal vectors (in 2D, this means they are perpendicular).
- *Euclidean norm* of vector

  $$\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{x_1 x_1 + x_2 x_2 + \cdots + x_d x_d}$$

  represents the magnitude of the vector.
- *Unit vector* has norm 1. Also called *normalised vector*.
- If $\|\mathbf{x}\| = 1$ and $\|\mathbf{y}\| = 1$, and $\mathbf{x}^T\mathbf{y} = 0$, then $\mathbf{x}$ and $\mathbf{y}$ are *orthonormal vectors*.
- *Outer-product* $\mathbf{x}\mathbf{z}^T$ is a $d \times k$ matrix.

Matrix and Vector Calculus

For scalars $x, y \in \mathbb{R}$, vectors $\mathbf{x} \in \mathbb{R}^d, \mathbf{y} \in \mathbb{R}^k$ and matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m \times n}$, we will use the following conventions for writing matrix and vector derivatives.

$$\text{Scalar w.r.t vector: } \nabla_{\mathbf{x}} y = \frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_d} \end{bmatrix}$$

$$\text{Vector w.r.t scalar: } \nabla_{x} \mathbf{y} = \frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_k}{\partial x} \end{bmatrix}$$

$$\text{Vector w.r.t vector: } \nabla_{\mathbf{x}} \mathbf{y} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \nabla_{\mathbf{x}} y_1 \\ \nabla_{\mathbf{x}} y_2 \\ \vdots \\ \nabla_{\mathbf{x}} y_k \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_d} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_k}{\partial x_1} & \frac{\partial y_k}{\partial x_2} & \cdots & \frac{\partial y_k}{\partial x_d} \end{bmatrix}}_{k \times d}$$

Matrix and Vector Calculus

Scalar w.r.t matrix: $\nabla_{\mathbf{X}} y = \frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{m1}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{m2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1n}} & \frac{\partial y}{\partial x_{2n}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{bmatrix}$

Matrix w.r.t scalar: $\nabla_{x} \mathbf{Y} = \frac{\partial \mathbf{Y}}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{bmatrix}$

Matrix and Vector Calculus

For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and matrices $\mathbf{M} \in \mathbb{R}^{k \times d}$ and $\mathbf{A} \in \mathbb{R}^{d \times d}$

- $\nabla_{\mathbf{x}}(\mathbf{y}^T\mathbf{x}) = \nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{y}) = \mathbf{y}^T$
- $\nabla_{\mathbf{x}}(\mathbf{M}\mathbf{x}) = \mathbf{M}$
- $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = \mathbf{x}^T(\mathbf{A}^T + \mathbf{A})$
- For symmetric $\mathbf{A}$, $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = 2(\mathbf{A}\mathbf{x})^T$

Prove all of the derivatives given above.