Linear algebra and gradients

Overview

- Notation: vector, matrix
- Definition: vector and matrix operations
- Gradients and chain rule

What is a vector?

- An array of numbers
- Notation: $\mathbf{v}$
  - bold lower case
- Size: $\text{size}(\mathbf{v}) = n$
- Order: $\dim(\mathbf{v}) = 1$
- Indexing: $\mathbf{v}_i$

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 4.3 \\ 1.2 \\ 9.9 \\ 2.3 \end{bmatrix}$$

$$\begin{bmatrix} 1.2 \\ 4.4 \end{bmatrix}$$

What is a matrix?

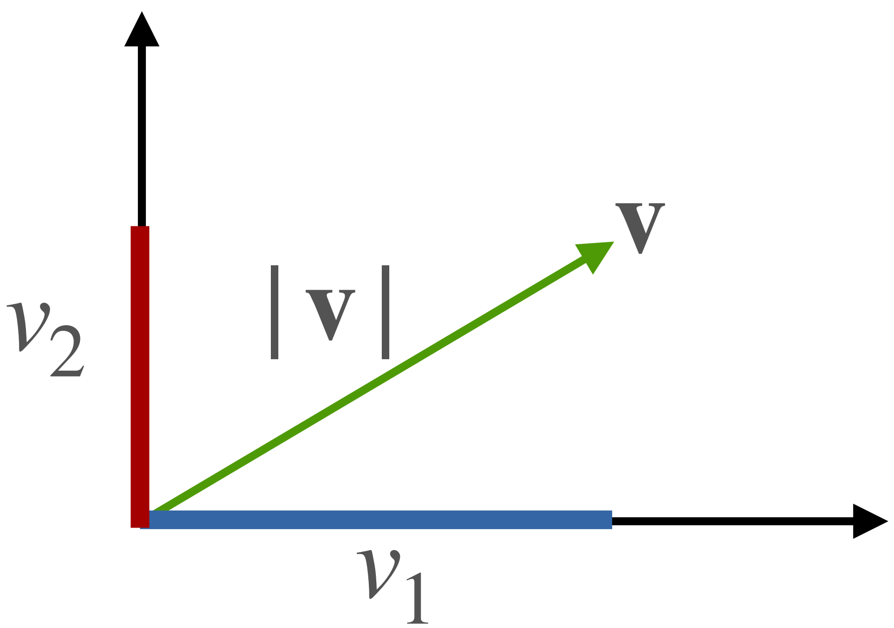
- An 2D array of numbers
- Notation: $\mathbf{M}$
  - bold upper case
- Size: $\text{size}(\mathbf{M}) = n \times m$
- Order: $\dim(\mathbf{M}) = 2$
- Indexing: $\mathbf{M}_{ij}$

$$\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

$$\begin{bmatrix} 1.2 & 1.1 \\ 4.4 & 3.2 \end{bmatrix}$$

Vector norm

$$|\mathbf{v}| = \sqrt{\sum_{i}^{n} v_i^2}$$



Element wise operations

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} v_0 + w_0 \\ v_1 + w_1 \\ \cdots \\ v_n + w_n \end{bmatrix} \quad \mathbf{v} * \mathbf{w} = \begin{bmatrix} v_0 * w_0 \\ v_1 * w_1 \\ \cdots \\ v_n * w_n \end{bmatrix} \quad \cdots$$

$$\mathbf{v} - \mathbf{w} = \begin{bmatrix} v_0 - w_0 \\ v_1 - w_1 \\ \cdots \\ v_n - w_n \end{bmatrix} \quad \mathbf{v}/\mathbf{w} = \begin{bmatrix} v_0/w_0 \\ v_1/w_1 \\ \cdots \\ v_n/w_n \end{bmatrix} \quad \cdots$$

Inner product

$$\mathbf{v} \cdot \mathbf{w} = v_0 \cdot w_0 + v_1 \cdot w_1 + \ldots + v_n \cdot w_n$$

Outer product

$$\mathbf{v} \otimes \mathbf{w} = \begin{bmatrix} v_0 w_0 & v_0 w_1 & \cdots & v_0 w_n \\ v_1 w_0 & v_1 w_1 & \cdots & v_1 w_n \\ & & \cdots & \\ v_n w_0 & v_n w_1 & \cdots & v_n w_n \end{bmatrix}$$

Frobenius norm

$$|\mathbf{M}| = \sqrt{\sum_{i}^{n} \sum_{j}^{m} M_{ij}^2}$$

Transpose

$$\mathbf{M} = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix}$$

$$\mathbf{M}^\top = \begin{bmatrix} a_{00} & a_{10} \\ a_{01} & a_{11} \\ a_{02} & a_{12} \end{bmatrix}$$

Matrix multiplication

$$\begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \end{bmatrix} \begin{bmatrix} b_{00} & b_{01} & b_{02} \\ b_{10} & b_{11} & b_{12} \\ b_{20} & b_{21} & b_{22} \\ b_{30} & b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} c_{00} & c_{01} & c_{02} \\ c_{10} & c_{11} & c_{12} \end{bmatrix}$$

2x4 4x3 2x3

Matrix-vector multiplication

$$\begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \end{bmatrix} \begin{bmatrix} b_{00} \\ b_{10} \\ b_{20} \\ b_{30} \end{bmatrix} = \begin{bmatrix} c_{00} \\ c_{10} \end{bmatrix}$$

2x4 4 2

Vector operations

- Length / norm: $|\mathbf{v}|$
- Element-wise operations: $\mathbf{v} + \mathbf{w}, \mathbf{v} - \mathbf{w}, \ldots$
- Inner (dot) product: $\mathbf{v}^\top \mathbf{w} = \mathbf{v} \cdot \mathbf{w}$
- Outer product: $\mathbf{v}\mathbf{w}^\top = \mathbf{v} \otimes \mathbf{w}$

Matrix Operations

- Frobenius norm $|\mathbf{M}|$
- Transpose $\mathbf{M}^\top$
- Matrix multiplication $\mathbf{AB}$ $\mathbf{Av}$

Functions of vectors

- Definition: $f : \mathbf{x} \rightarrow \mathbf{y}$

Derivatives of vector valued functions

- Gradient of scaler function

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{x}_0} & \frac{\partial f}{\partial \mathbf{x}_1} & \cdots & \frac{\partial f}{\partial \mathbf{x}_n} \end{bmatrix}$$

scaler

- **Jacobian**: Derivative of vector $\mathbf{g}(\mathbf{x})$ by vector $\mathbf{x}$:

$$\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{g}}{\partial \mathbf{x}_0} & \frac{\partial \mathbf{g}}{\partial \mathbf{x}_1} & \cdots & \frac{\partial \mathbf{g}}{\partial \mathbf{x}_n} \end{bmatrix}$$

vector

Chain rule for scalar functions

- Nested functions: $f(g(x))$
  - where $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$
  - and $y = g(x)$
- Derivative: $$\frac{\partial f(g(x))}{\partial x} = \frac{\partial f(y)}{\partial y} \frac{\partial g(x)}{\partial x}$$

Chain rule for vector-valued functions

- Nested functions: $f(g(\mathbf{x}))$
  - where $f : \mathbb{R}^m \to \mathbb{R}^n$ and $g : \mathbb{R}^p \to \mathbb{R}^m$
  - and $\mathbf{y} = g(\mathbf{x})$
- Derivative: $$\frac{\partial f(g(\mathbf{x}))}{\partial \mathbf{x}} = \frac{\partial f(\mathbf{y})}{\partial \mathbf{y}} \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}$$

n x p   n x m   m x p

Summary

- Vector **v**
- Matrix **M**
- Gradients $\frac{\partial f(x)}{\partial x}$
- Chain rule $\frac{\partial f(g(x))}{\partial x} = \frac{\partial f(y)}{\partial y} \frac{\partial g(x)}{\partial x}$
- More reading
  [The Matrix Cookbook, Petersen and Pedersen 2012]