CS 342: Neural Networks

Logistics

Logistics – Online classes

- Lectures and office hours on Zoom (for now)
- Zoom recordings
- If possible, we will have classes in JGB 2.216

Logistics – Tools

- Website: most course materials
  - gavlegoat.github.io/cs342
- Canvas: any non-public course material
- Piazza: discussion
- Zoom: lectures and office hours (for now)

Logistics – TAs

- Office hours
  - Wednesday 3:30-4:30
  - Friday 2:00-3:00
  - My office hours TBD – Please fill out the poll on Canvas
- Piazza

Logistics – Mechanics

- Lecture for about half of each class
- Coding exercises for the remaining half
  - Whole-class or small groups
- Quizzes in class on Mondays
- Homework
- Final Project

Background

Vectors

Vectors: Arrays of numbers

$$v = \begin{pmatrix} 2.3 \\ 9.1 \\ 5.2 \\ 6.1 \end{pmatrix}$$

Basic vector operations:

Size: $\mathrm{size}(v) = 4$

Dimension: $\dim(v) = 1$

Indexing: $v_2 = 9.1$

Euclidean spaces: $v \in \mathbb{R}^4 = \mathbb{R}^{\mathrm{size}(v)}$

Matrices

Matrices: 2D arrays of numbers

$$M = \begin{pmatrix} -0.1 & 1.2 & -3.1 & 0.5 \\ 1.5 & -5.2 & -4.1 & 2.1 \\ -0.1 & 2.4 & 7.1 & -4.2 \end{pmatrix}$$

Basic matrix operations:

Size: $\mathrm{size}(M) = 3\times 4$

Dimension: $\dim(M) = 2$

Indexing: $M_{3,2} = 2.4$ $\quad M_{ij}$

Matrix spaces: $M \in \mathbb{R}^{3\times 4} = \mathbb{R}^{\mathrm{size}(M)}$

Norms

p-norm: $|\mathrm{v}|_p = \left(\sum_{i=1}^{\mathrm{size}(\mathrm{v})} |\mathrm{v}_i|^p \right)^{1/p}$

Special case – 2-norm: $|\mathrm{v}| = |\mathrm{v}|_2 = \sqrt{\sum_{i=0}^{\mathrm{size}(\mathrm{v})} \mathrm{v}_i^2}$

Special case – ∞-norm: $|\mathrm{v}|_\infty = \max_{1 \le i \le \mathrm{size}(\mathrm{v})} |\mathrm{v}_\mathrm{i}|$

For matrices –
Frobenius Norm: $|\mathrm{M}| = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} \mathrm{M}_{ij}^2}$ where $\mathrm{size}(\mathrm{M}) = n \times m$

Vector Operations

Element-wise operations:

$$v + w = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix} \qquad v - w = \begin{pmatrix} v_1 - w_1 \\ v_2 - w_2 \\ \vdots \\ v_n - w_n \end{pmatrix}$$

Inner (dot) product:

$$v \cdot w = \sum_{i=1}^{\mathrm{size}(v)} v_i w_i$$

Outer product:

$$v \otimes w = \begin{pmatrix} v_1 w_1 & v_1 w_2 & \cdots & v_1 w_m \\ v_2 w_1 & v_2 w_2 & \cdots & v_2 w_m \\ \vdots & \vdots & \ddots & \vdots \\ v_n w_1 & v_n w_2 & \cdots & v_n w_m \end{pmatrix}$$

Matrix Operations

Transpose: $\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}^T = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix}$ $\quad \mathrm{M}_{ij}^T = \mathrm{M}_{ji}$

Matrix multiplication: $(\mathrm{AB})_{ij} = \sum_{k=1}^{n} \mathrm{A}_{ik} \mathrm{B}_{kj}$ $\quad \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$

$$(2 \times 3)(3 \times 2) \rightarrow (2 \times 2)$$

Matrix-vector multiplication:

$$(\mathrm{Av})_i = \sum_{k=1}^{n} \mathrm{A}_{ik} \mathrm{v}_k \qquad \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

Inner and outer product are
Also matrix multiplication:

$$\mathrm{v} \cdot \mathrm{w} = \mathrm{v}^T \mathrm{w} \qquad \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

$$\mathrm{v} \otimes \mathrm{w} = \mathrm{v}\,\mathrm{w}^{\mathrm{T}}$$

Vector Functions

We will usually be working with functions that take vectors as input and return vectors as output

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$f\begin{pmatrix} 1.4 \\ 0.2 \\ -2.3 \end{pmatrix} = \begin{pmatrix} 2.1 \\ -0.4 \end{pmatrix}$$

Gradients

Scalar function (gradient):

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \frac{\partial f(\mathrm{v})}{\partial \mathrm{v}} = \nabla f(\mathrm{v}) = \begin{pmatrix} \frac{\partial f}{\partial \mathrm{v}_1} & \frac{\partial f}{\partial \mathrm{v}_2} & \cdots & \frac{\partial f}{\partial \mathrm{v}_n} \end{pmatrix}$$

Vector function (Jacobian):

$$g : \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad g(\mathrm{v}) = \begin{pmatrix} g_1(\mathrm{v}) \\ \vdots \\ g_m(\mathrm{v}) \end{pmatrix}, \quad \frac{\partial g(\mathrm{v})}{\partial \mathrm{v}} = \mathbf{J}(g) = \begin{pmatrix} \frac{\partial g_1}{\partial \mathrm{v}_1} & \cdots & \frac{\partial g_1}{\partial \mathrm{v}_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial \mathrm{v}_1} & \cdots & \frac{\partial g_m}{\partial \mathrm{v}_n} \end{pmatrix}$$

Gradients – Chain Rule (scalar)

For computing derivatives of the composed function $f(g(x))$

$$y = g(x)$$

$$\frac{\mathrm{d}f(g(x))}{\mathrm{d}x} = \frac{\mathrm{d}f(y)}{\mathrm{d}x} = \frac{\mathrm{d}f(y)}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}f(y)}{\mathrm{d}y}\frac{\mathrm{d}g(x)}{\mathrm{d}x}$$

Gradients – Chain Rule (vector)

For composition of vector valued functions $f(g(x))$

where $f: \mathbb{R}^m \to \mathbb{R}^n, g: \mathbb{R}^p \to \mathbb{R}^m$

$$y = g(x)$$

$$\underbrace{\frac{\partial f(g(x))}{\partial x}}_{n \times p} = \underbrace{\frac{\partial f(y)}{\partial y}}_{n \times m} \underbrace{\frac{\partial g(x)}{\partial x}}_{m \times p}$$

Special Topics

- Your interests
- Special topics