CIS3333: Mathematics of Machine Learning Fall 2025

Homework 7

Due: November 12, 2025, 11:59 PM ET

Submission Instructions: Submit a single PDF to Gradescope. Show key steps and justify your answers conceptually.

Collaboration & AI Policy: You may discuss approaches with classmates, but write up your own solutions and list collaborators. If you use computational tools (including LLMs) for checking, cite them and ensure the reasoning is your own.

Problem 1: Differentiability vs. Partial Derivatives (7 points)

Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

1. (2 points) Calculate the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ at the origin $(0, 0)$ using the limit definition.
2. (5 points) Show that f is not differentiable at $(0, 0)$.

   Hint: Recall that a function $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at $(0, 0)$ if

   $$f(h) = f(0, 0) + \nabla f(0, 0)^\top h + o(\|h\|) \quad \textit{for } \nabla f(0, 0) = \begin{bmatrix} \frac{\partial f}{\partial x}(0, 0) \\ \frac{\partial f}{\partial y}(0, 0) \end{bmatrix}$$

   and this is equivalent to saying that the directional derivative $D_h f(0, 0) = \nabla f(0, 0)^\top h$ for all $h \in \mathbb{R}^2$.

Problem 2: Taylor Approximations and Optimization (10 points)

Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x) = e^{x_1+x_2}$.

1. (2 points) Find the gradient $\nabla f(x)$ and evaluate it at the point $x_0 = (0, 0)$. Then write down the first-order (linear) Taylor approximation $f_1(x)$ at this point:

   $$f_1(x) = f(x_0) + \nabla f(x_0)^\top (x - x_0)$$

2. (3 points) Compute the Hessian matrix $\nabla^2 f(x)$ and evaluate it at $x_0 = (0, 0)$. Then write down the second-order (quadratic) Taylor approximation $f_2(x)$ at x_0:

   $$f_2(x) = f(x_0) + \nabla f(x_0)^\top (x - x_0) + \frac{1}{2}(x - x_0)^\top \nabla^2 f(x_0)(x - x_0)$$

3. (5 points) Suppose we want to minimize f starting from $(0, 0)$ using gradient descent with step size η. The update would be:

$$x_1 = x_0 - \eta \nabla f(x_0).$$

Recall that from a linear approximation, we do not know how far to step. We can use the quadratic approximation $f_2(x)$ from part (b) to find the value of η that minimizes $f_2(x_1)$. Find this value of η.

Problem 3: Gradient Computation for Least Squares (13 points)

We will consider the least squares loss function and use both backward-mode and forward-mode autodiff to compute the gradient. In class, we already saw the backward mode autodiff for a very similar problem.

Consider the ordinary least squares loss function over a dataset of N examples with input matrix $X \in \mathbb{R}^{N \times D}$ and output vector $Y \in \mathbb{R}^N$:

$$L(\theta) = \frac{1}{N} \|X\theta - Y\|^2.$$

Recall that the computation graph can be broken into three operations:

$$\theta \xrightarrow{f_1(\theta) = X\theta - Y} v_1 \xrightarrow{f_2(v_1) = v_1^\top v_1} v_2 \xrightarrow{f_3(v_2) = \frac{1}{N} v_2} L$$

Part A: Computing the Local Jacobians (3 points) Show that the Jacobians for each operation are:

- $J_{f_1}(\theta) = X.$
- $J_{f_2}(v_1) = 2v_1^\top.$
- $J_{f_3}(v_2) = \frac{1}{N}.$

Note: The notes have the formulas for these, but explain why they are correct.

Part B: Backward-mode Autodiff (5 points) Compute the gradient $\nabla_\theta L(\theta)$ using backward-mode autodiff. Your solution should clearly state the computations for $J_L(L), J_L(v_2), J_L(v_1), J_L(\theta)$ by applying the chain rule.

Note: For all computations, you should only use the previous Jacobian and some local Jacobian.

Part C: Forward-mode Autodiff (5 points) An alternative way to compute the gradient is to propagate derivatives forward through the computation graph. Your solution should clearly state the computation for $J_\theta(\theta), J_{v_1}(\theta), J_{v_2}(\theta), J_L(\theta)$ by applying the chain rule.

Note: For all computations, you should only use the previous Jacobian and some local Jacobian.