CS-5365 Deep Learning: Homework 1 (written questions)

<u>Your Name:</u> (your email address)

January 30, 2026

Total: 50 points.

1 Derive $\frac{\partial f}{\partial x_i}$ Using the Chain Rule (20 pts)

Instructions. For each problem below, derive the partial derivative $\frac{\partial f}{\partial x_i}$ with respect to a *single coordinate* x_i. Show your chain rule steps by introducing intermediate variables whenever helpful.

Question 1 (5 pts): Elementwise nonlinearity then sum. Let $\mathbf{x} \in \mathbb{R}^n$ and define

$$f(\mathbf{x}) = \sum_{j=1}^{n} \tanh(x_j)^2.$$

Derive $\frac{\partial f}{\partial x_i}$ and justify why only one term in the sum contributes.

Question 2 (7 pts): Mistake finder (sign and dependency check). Consider

$$f(\mathbf{x}) = \exp\left(-\sum_{j=1}^{n} x_j^2\right).$$

A student claims $\frac{\partial f}{\partial x_i} = 2x_i f(\mathbf{x})$. Determine whether the claim is correct; if not, provide the corrected derivative and show the chain rule steps.

Question 3 (8 pts): Self-power function. Let $x > 0$ and define

$$f(x) = x^x.$$

Calculate $\frac{df}{dx}$. (Hint: rewrite $f(x) = \exp(x \log x)$ and apply the chain rule.)

2 Matrix–Vector Calculations (30 pts)

Instructions. Show all key steps. Unless stated otherwise, vectors are column vectors.

Question 4 (8 pts): Shapes and sanity checks. Let $W \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and define $y = Wx + b$.

1. State the shapes of y, Wx, and b.

2. Suppose x is a batch of B examples arranged as $X \in \mathbb{R}^{B \times n}$. Write the batched affine map producing $Y \in \mathbb{R}^{B \times m}$.

3. In the batched setting, explain one common broadcasting convention for adding b.

Question 5 (7 pts): Matrix–vector product (shape + cost). Let $W \in \mathbb{R}^{m \times n}$ and $x \in \mathbb{R}^n$.

1. Write the i-th entry of $y = Wx$ as an explicit summation.

2. State the shape of y.

3. Count the number of scalar multiplications and additions required to compute y (as a function of m, n).

Question 6 (7 pts): Broadcasting practice. Let $X \in \mathbb{R}^{B \times d}$, $b \in \mathbb{R}^d$, and $\gamma \in \mathbb{R}^d$. Consider the computation

$$Y = (X + b) \odot \gamma,$$

where $\odot$ denotes elementwise multiplication.

1. State the shape of Y.

2. Explain (in one sentence) how b and γ are broadcast to match X.

3. Write $Y_{i,j}$ explicitly in terms of $X_{i,j}$, b_j, and γ_j.

Question 7 (8 pts): Outer product and rank. Let $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$. Define $A = uv^\top$.

1. State the shape of A.

2. Write $A_{i,j}$ as a product of scalars.

3. Show that $Ax = u(v^\top x)$ for any $x \in \mathbb{R}^n$ and explain why this can reduce computation.