# AIE6001 Assignment 1: Calculating Gradient for Deep Learning

*Due date: 11:59 PM, Monday, September 22, 2025.*

**Remark:** The Maximum point is 100. For instance, if your grade with bonus point is 95+10=105, your actual grade will be $\min(105, 100) = 100$.

**Question 1** (Chain Rule)**.** *Suppose I have a function $q : \mathbb{R}^d \to \mathbb{R}$ defined as*

$$q(x) = f(g(x)), \quad x \in \mathbb{R}^d,$$

*where $g : \mathbb{R}^d \to \mathbb{R}$ is defined as $g(x) = \sum_{i=1}^d x_i$ (namely, it takes summation over all entries of $x$), and $f : \mathbb{R} \to \mathbb{R}$ is defined as $f(y) = y^4$. Derive the expression of $\frac{\partial q(x)}{\partial x}$.* *(20 points)*

**Question 2** (Chain Rule for Vector-valued Function)**.** *Suppose I have a scalar function $q$ defined as*

$$q(x; a, b) = b \cdot \phi\big(a^\top \phi(x)\big), \quad x \in \mathbb{R}^d.$$

*In particular,*

- *The function $\phi(x) = \log(1 + e^x)$ operates componentwisely (namely, if $x$ is a scalar, $\phi(x) = \log(1 + e^x)$ is a scalar; if $x$ is a $d$-dimensional vector, $\phi(x) = (\log(1 + e^{x_1}), \ldots, \log(1 + e^{x_d}))^\top$ is also a $d$-dimensional vector).*
- *$a \in \mathbb{R}^d$ and $b \in \mathbb{R}$ are weight parameters.*

*Answer the following questions:*

- *For $x \in \mathbb{R}^d$, figure out the dimensions of $\phi(x)$, $a^\top \phi(x)$, $\phi\big(a^\top \phi(x)\big)$, and $q(x; a, b)$, respectively.* *(8 points)*
- *Calculate the expressions of $\frac{\partial q(x;a,b)}{\partial x}$, $\frac{\partial q(x;a,b)}{\partial a}$, $\frac{\partial q(x;a,b)}{\partial b}$.* *(12 points)*

**Question 3** (Chain Rule for Neural Networks)**.** *Suppose I have a scalar-valued function $F$ defined as*

$$F(x, y; W^{(1)}, W^{(2)}) = \left\| y - W^{(2)} \phi\big(W^{(1)} x\big) \right\|^2, \quad x \in \mathbb{R}^d, y \in \mathbb{R}^k.$$

*In particular,*

- *The function $\phi(x) = \log(1 + e^x)$ operates componentwisely.*
- *$W^{(1)} \in \mathbb{R}^{p \times d}, W^{(2)} \in \mathbb{R}^{k \times p}$ are weight parameters.*

*Answer the following questions:*

- *Calculate the expressions of $\frac{\partial F(x,y;W^{(1)},W^{(2)})}{\partial W^{(1)}}$ and $\frac{\partial F(x,y;W^{(1)},W^{(2)})}{\partial W^{(2)}}$.* *(20 points)*

- *Fill in the function* `compute_gradients.py` *shown in the python file* `Q3_AIE6001_A1`*.py that returns* $\frac{\partial F(x,y;W^{(1)},W^{(2)})}{\partial W^{(1)}}$ *and* $\frac{\partial F(x,y;W^{(1)},W^{(2)})}{\partial W^{(2)}}$*. Please provide a screenshot of the code in this part and also upload your code when you submit your homework. Note: You could only use numpy instead of other packages. (20 points)*
- *Run the python file* `Q3_AIE6001_A1`*.py. What do you observe from the output? (10 points)*

**Question 4** (Dimension)**.** *In each of the following, determine the dimension of the space (each small question weights 4 points):*

*1)*

$$\textit{Span}\left\{\begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \\ 4 \end{pmatrix}, \begin{pmatrix} -3 \\ 3 \\ 6 \end{pmatrix}\right\}$$

*(b) col(A), where* $A = \begin{bmatrix} 1 & -2 & 3 & 2 \\ -1 & 2 & -2 & -1 \\ 2 & -4 & 5 & 3 \end{bmatrix}$*;*

*(c)* $N(B)$*, where* $B = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 4 \\ 4 & 7 & 8 \end{bmatrix}$*;*

*(d) span* $\{(x-2)(x+2), x^2(x^4-2), x^6-8\}$*;*

*(e) span* $\{5, \cos 2x, \cos^2 x\}$ *as a subspace of* $C[-\pi, \pi]$.

$C[-\pi, \pi]$ *denotes the space of continuous functions defined on the domain* $[-\pi, \pi]$.