Topics
My posts
More
CATEGORIES
General
Site Feedback
All categories
Skip to main content

2025 USA-NA-AIO Round 2, Problem 2, Part 1

USAAIO May 2025

Problem 2 (100 points)

Multi-head attention (MHA) is a big breakthrough in AI. Based on its original form, there are many variants that improved it.

In this problem, you are asked to study multi-head attention and its variants.

We use the following notation in this problem.

- B: batch size. b: index of a sample.
- L_1: length of an attending sequence. l_1: index of a position in this sequence.
- L_2: length of a being attended sequence. l_2: index of a position in this sequence.
- D_1: dimension of a hidden state/token in an attending sequence.
- D_2: dimension of a hidden state/token in a being attended sequence.
- H: number of heads. h: index of a head.
- D_v: dimension of a value vector.
- D_{qk}: dimension of a query/key vector.

```
# Run code in this cell

"""
DO NOT MAKE ANY CHANGE IN THIS CELL.
```

```
"""

import torch
import torch.nn as nn
import numpy as np
```

WARNING !!!

- Beyond importing libraries/modules/classes/functions in the preceeding cell, you are **NOT allowed to import anything else for the following purposes**:
  - **As a part of your final solution.** For instance, if a problem asks you to build a model without using sklearn but you use it, then you will not earn points.
  - **Temporarily import something to assist you to get a solution.** For instance, if a problem asks you to manually compute eigenvalues but you temporarily use `np.linalg.eig` to get an answer and then delete your code, then you violate the rule.

  Rule of thumb: Each part has its particular purpose to intentionally test you something. Do not attempt to find a shortcut to circumvent the rule.

Part 1 (5 points, non-coding task)

Do the following tasks (Reasoning is not required).

1. For each hidden state at position l_1 in an attending sequence, $\mathbf{x}_{l_1} \in \mathbb{R}^{D_1}$, we project it into a query vector for head h according to

$$\mathbf{q}_{l_1,h} = \mathbf{W}_h^{\mathbf{Q}} \mathbf{x}_{l_1}.$$

Skip to main content

What is the shape of $\mathbf{W}_h^{\mathbf{Q}}$?

2. For each hidden state at position l_2 in a being attended sequence $\mathbf{y}_{l_2} \in \mathbb{R}^{D_2}$, we project it into a key vector for head h according to

$$\mathbf{k}_{l_2,h} = \mathbf{W}_h^{\mathrm{K}} \mathbf{y}_{l_2}.$$

What is the shape of $\mathbf{W}_h^{\mathrm{K}}$?

3. For each hidden state at position l_2 in a being attended sequence $\mathbf{y}_{l_2} \in \mathbb{R}^{D_2}$, we project it into a value vector for head h according to

$$\mathbf{v}_{l_2,h} = \mathbf{W}_h^{\mathrm{V}} \mathbf{y}_{l_2}.$$

What is the shape of $\mathbf{W}_h^{\mathrm{V}}$?

USAAIO **May 2025**

Misplaced '#'

1. The shape of $\mathbf{W}_h^{\mathrm{Q}}$ is (D_{qk}, D_1).
2. The shape of $\mathbf{W}_h^{\mathrm{K}}$ is (D_{qk}, D_2).
3. The shape of $\mathbf{W}_h^{\mathrm{V}}$ is (D_v, D_2).

""" END OF THIS PART """

Skip to main content

Related topics

Topic	Replies	Activity
2025 USA-NA-AIO Round 2, Problem 2, Part 5	1	May 2025
2025 USA-NA-AIO Round 2, Problem 2, Part 2	2	Dec 2025
2025 USA-NA-AIO Round 2, Problem 2, Part 6	2	May 2025
2025 USA-NA-AIO Round 2, Problem 2, Part 4	1	May 2025
2025 USA-NA-AIO Round 2, Problem 2, Part 9	1	May 2025