

《深度学习平台与应用》作业二 (20241018)

第一题:

- vanilla RNN的为什么会出现梯度消失或梯度爆炸的情况?
- vanilla RNN 出现梯度爆炸该如何解决?
- LSTM是如何缓解梯度消失的, 它能彻底解决梯度消失吗?

第二题:

- 问题1: 完成下面self-attention的代码填空, 并结合attention计算的数学公式解释你的做法。

```
1  import torch
2  import torch.nn.functional as F
3
4  # 假设输入是一个5个词, 4维的词向量矩阵
5  X = torch.tensor([[1.0, 0.0, 1.0, 0.0],
6                    [1.0, 2.0, 1.0, 0.0],
7                    [0.0, 1.0, 0.0, 1.0],
8                    [0.0, 1.0, 0.0, 0.0],
9                    [1.0, 0.0, 1.0, 0.0]])
10
11 # 初始化查询、键和值的权重矩阵
12 W_Q = torch.randn(4, 4)
13 W_K = torch.randn(4, 4)
14 W_V = torch.randn(4, 4)
15
16 #todo 计算查询、键、值
17
18
19 #todo 计算注意力得分
20
21
22 #todo 加权和计算最终output
23
24 #print(output)
```

- 问题2：对比并简要解释多头注意力的实现方式。
- 问题3：Transformer结构中encoder和decoder的区别有哪些

第三题：

- 请简要描述R-CNN模型的工作流程。具体来说，R-CNN是如何进行目标检测的？简述R-CNN模型的主要步骤
- 与传统的R-CNN相比，Fast R-CNN的最大改进是什么？请简要说明Fast R-CNN如何避免了R-CNN中存在的冗余计算问题。