# 第13章习题参考答案

1. “深度”学习“更深”的结构有哪些好处？同时带来了哪些问题？

好处：

首先，增加新的隐藏层便引入了新的激活函数，激活函数之间的反复嵌套使得模型的非线性性大幅度提升，能够描述出更加复杂的关系。

其次，深度网络可以视为单一隐藏层的浅层神经网络的叠加，前一个隐藏层的输出可以看作是后一个隐藏层的输入。这使得模型有着类似于决策树，或者说类似于人类的分层次逻辑。

问题：

参数过多，计算达到瓶颈。

梯度消失与梯度爆炸。

1. 怎样缓解“梯度消失”？

使用ReLU激活函数而不要使用Sigmoid

使用LSTM等RNN结构

1. 阐述CNN参数共享的思想和好处。

大幅度减少了参数量，极大的提升了运算效率；

解决了很多图像识别中目标的位置不固定等问题。

1. 举例说明几个序列化数据以及它们的序列化特征。

序列数据可以看作是若干个相同类别的数据连接组成的数据。例如一句话是由很多个字组成的，每个元素“字”是同一类别的。

序列数据一般是不固定长度的。以一句话为例，有多少个字是不固定的，可以很长也可以很短。

序列数据的前后顺序包含有重要信息，且有关联关系。继续以语言为例：“北京到上海”和“上海到北京”虽然包含的字完全一样，但因为顺序不同，意思完全不一样；对于“他看上去很高兴”这句话，如果直接问“他”指代谁，则无法给出回答，但如果你能读到上文“我见到了小明，他看上去很高兴”，那很明显“他”是小明。

1. 阐述为什么LSTM可以实现长期的记忆。

LSTM利用遗忘门、输入门和输出门来控制长期和短期的记忆。

1. 根据本章实践中使用CNN处理空间关系和LSTM处理时序关系，搭建同时结合CNN和LSTM的网络进行交通状态预测。

略

1. （选做）查阅资料，学习ResNet的结构，了解残差连接结构对于“更深”的深度学习的帮助。

Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778

残差连接结构可以使神经网络通过拟合残差的方式，实现一种类似“短路”的操作，使得网络部分层的输入=输出，避免了更深层次存在的过拟合和梯度问题

1. （选做）查阅资料，学习Transformer的结构，了解其如何解决了LSTM的非并行化、以及“长期记忆仍然不够长”的问题。以及基于Transformer的语言预训练模型BERT、GPT在NLP中表现的强大性能。

Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.

Transformer包括了encoder和decoder两个部分，但两个部分的核心都在于多头注意力机制（multi-head attention），其利用全连接神经网络将序列数据重映射到Query、Key、Value三个空间，并进行加权求和。其最大特点是不会因为序列过长而丢失长距离依赖性，且计算高度并行化。

Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

BERT是基于Transformer Encoder的预训练模型，其无监督（或者说自监督）地将词汇向量化，之后只需进行轻微的调整（fine-tuning）就可以表现出优良的性能。