

1. Geoffrey Hinton 既获得了诺贝尔奖也获得了图灵奖
2. John Hopfield 仅获得了诺贝尔奖，没有获得图灵奖
3. 深度学习中的“深度”指的是神经网络的深度（层数）
4. 卷积网络包括卷积层、池化层和全连接层
5. 前馈神经网络的特点是相邻层间的神经元密集连接
6. 欠拟合表现为训练误差和测试误差都偏大
7. 训练集效果良好但不能推广到测试集为过拟合
8. 如果收敛网络出现过拟合，则表现为训练误差小，测试误差大
9. 卷积神经网络 CNN 中，卷积层的作用是提取局部特征
10. CNN 中，池化层的作用是降低特征维度
11. GRU/LSTM 网络相比于 RNN 的改进是更好地解决梯度消失问题（但也不能完全解决）
12. LSTM 中，遗忘门、输入门、输出门用的激活函数是 sigmoid
13. 交叉熵损失更适用于分类问题
14. 衡量模型预测的概率分布与真实概率分布之间的差异的损失函数是交叉熵损失
15. 网络权重不是超参数，网络层数、学习率、批量大小是超参数
16. 二阶算法牛顿法优化算法在深度学习中不常用，SGD/Adam/AdamW 优化算法在深度学习中常用。
17. 用数据的输入的一部分预测另一部分，输入的一部分有标签，无需人工标注标签的方法叫做自监督学习。
18. Google Net, AlexNet, ResNet 均使用了 ReLU, LeNet 没有使用 ReLU。
19. 注意力机制包括键、值、查询，不包括池化。
20. AlexNet 最早使用了 GPU
21. 常用的激活函数有哪些，各自有什么优缺点

22. Relu 相比于 sigmoid 有什么优点? sigmoid 在很多网络中仍被使用, 哪些优点是 relu 不具备的

23. 卷积和池化的计算过程, 填充和步幅的作用分别是什么

24. 批量规范化有什么作用

25. 丢弃法有什么作用

26. 1×1 卷积的作用是什么

27. 什么是梯度消失和梯度爆炸，二者在深度学习中分别会带来什么问题，分别有哪些改善方法

28. 解释多头自注意力的计算过程

29. 在注意力机制中为什么使用位置编码

30. 在 transformer 的解释器中为什么使用掩蔽注意力

31. 大模型中预训练和微调分别指什么，为什么分两步？
什么是自监督学习，为什么采用自监督学习的预训练方法？