QA-深度学习基础

1. 什么层（layer）或操作（operation）赋予了深度学习模型⾮线形关系学习能⼒

非线性激活函数的引入使深度学习模型具有非线性学习能力。

1. batch\_size 选择有什么讲究吗？

Batch\_size是指每一次进行训练的数据组的大小，这个数据组是整体数据中的一部分，当batchsize很大时会使在一个epoch内循环的次数减少，使得需要迭代更多次才能使损失函数收敛，数据量很大时对内存要求也很高；当batchsize很小的时候会导致损失函数很难收敛的情况。

1. Adam的优势是什么？

Adam梯度经过偏置校正后，每一次迭代学习率都有一个固定范围，使得参数比较平稳。结合了Adagrad善于处理稀疏梯度和RMSprop善于处理非平稳目标的优点

为不同的参数计算不同的自适应学习率，适用于大多非凸优化问题——适用于大数据集和高维空间。