1. **梯度下降算法的正确步骤是什么？**

**方向传播，链式求导**

**2、已知：**

**- 大脑是有很多个叫做神经元的东西构成，神经网络是对大脑的简单的数学表达。**

**- 每一个神经元都有输入、处理函数和输出。**

**- 神经元组合起来形成了网络，可以拟合任何函数。**

**- 为了得到最佳的神经网络，我们用梯度下降方法不断更新模型**

**给定上述关于神经网络的描述，什么情况下神经网络模型被称为深度学习模型？**

**A.加入更多层，使神经网络的深度增加**

**B.有维度更高的数据**

**C.当这是一个图形识别的问题时**

**D.以上都不正确**

**3、训练CNN时，可以对输入进行旋转、平移、缩放等预处理提高模型泛化能力。这么说是对，还是不对？ A**

**A.对    B.不对**

**4、下面哪项操作能实现跟神经网络中Dropout的类似效果？ B**

**A.Boosting    B.Bagging    C.Stacking    D.Mapping**

**5、下列哪一项在神经网络中引入了非线性？ B**

**A.随机梯度下降**

**B.修正线性单元（ReLU）**

**C.卷积函数**

**D.以上都不正确**

**6.CNN的卷积核是单层的还是多层的？**

**7.什么是卷积？**

图像和滤波矩阵做内积（逐个元素相乘再求和）的操作就是所谓的『卷积』操作，也是卷积神经网络的名字来源。

**8.什么是CNN的池化pool层？**

池化的作用最可靠的解释是在尽量保持原数据关系的基础上进行变化降低指定维度的长度，常用的是平均和最大池化法。

**9.简述下什么是生成对抗网络。**

GAN之所以是对抗的，是因为GAN的内部是竞争关系，一方叫generator，它的主要工作是生成图片，并且尽量使得其看上去是来自于训练样本的。另一方是discriminator，其目标是判断输入图片是否属于真实训练样本。

这两个网络进行连续的博弈，生成器学习产生越来越多的现实样本，鉴别器正在学习越来越好地区分生成的数据和实际数据。 这两个网络同时进行训练，最后的希望是竞争能够使生成器生成的样本与实际数据不可区分。

**10.学梵高作画的原理是什么？**

风格迁移：

**11.请简要介绍下tensorflow的计算图。**

Tensorflow是通过 计算图 的形式来表述计算的编程系统，**计算图也叫数据流图，**可以把计算图看做是一种**有向图**，Tensorflow中的**每一个节点都是**计算图上的一个**Tensor**, 而**节点之间的边**描述了**计算**之间的依赖**关系**(定义时)和数学操作(运算时)。

**12.你有哪些deep learning（rnn、cnn）调参的经验？**

1、参数初始化，参数初始化影响收敛速度和收敛结果甚至造成Nan等问题，

uniform均匀分布初始化：w = np.random.uniform

normal高斯分布初始化方法如下：w = np.random.randn

2、数据预处理

X -= np.mean(X, axis = 0) # zero-center 减去均值就是 zero-center

X /= np.std(X, axis = 0) # normalize 标准化（zero-center 除以标准差）

**★★★不会 PCA whitening,这个用的比较少.**

暗通道去雾法

Tanh\sigmoid后半部增亮法

3、训练技巧：

1）dropout， 近输入端 和 近输出端

2) batchsize，不能太小

3) adam、sgd 关于学习率， 动量参数

4）空间金字塔

5）adam,adadelta等,在小数据上,我这里实验的效果不如sgd, sgd收敛速度会慢一些，但是最终收敛后的结果，一般都比较好。

6、尽量对数据做shuffle：

**13.CNN最成功的应用是在CV，那为什么NLP和Speech的很多问题也可以用CNN解出来？为什么AlphaGo里也用了CNN？这几个不相关的问题的相似性在哪里？CNN通过什么手段抓住了这个共性？**

不相关问题的相关性在于，都存在局部与整体的关系，由低层次的特征经过组合，组成高层次的特征，并且得到不同特征之间的空间相关性。

CNN抓住此共性的手段主要有四个：**局部连接／权值共享／池化操作／多层次结构。**局部连接使网络可以提取数据的局部特征；权值共享大大降低了网络的训练难度，一个Filter只提取一个特征，在整个图片（或者语音／文本） 中进行卷积；池化操作与多层次结构一起，实现了数据的降维，将低层次的局部特征组合成为较高层次的特征，从而对整个图片进行表示。

**14.LSTM结构推导，为什么比RNN好？**

LSTM 相对于RNN长期记忆更好。

因为LSTM有进有出且当前的cell informaton是通过input gate控制之后叠加的，RNN是叠乘，因此LSTM可以防止梯度消失或者爆炸。

**15.Sigmoid、Tanh、ReLu这三个激活函数有什么缺点或不足，有没改进的激活函数。**

Sigmoid 、tanh 都是饱和函数， 输入必须限制在非饱和区， -1 至 1.

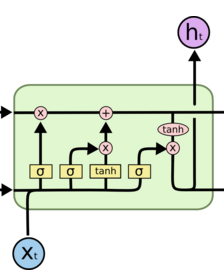
Relu 缺点，

改进： swish smooth hardswish。

**16、为什么引入非线性激励函数？**

**17、请问人工神经网络中为什么ReLu要好过于tanh和sigmoid function？**

**18、为什么LSTM模型中既存在sigmoid又存在tanh两种激活函数，而不是选择统一一种sigmoid或者tanh？这样做的目的是什么？**



Sigmoid 要的是 0 – 1的输出数据。

从激活的角度 tanh 优于 sigmoid。

**19、如何解决RNN梯度爆炸和弥散的问题？**

**20、什麽样的资料集不适合用深度学习？**

**21.广义线性模型是怎被应用在深度学习中？**

**22.如何解决梯度消失和梯度膨胀？**

**23.简述神经网络的发展历史。**

**24.深度学习常用方法。**

**25.请简述神经网络的发展史。**

**26.神经网络中激活函数的真正意义？一个激活函数需要具有哪些必要的属性？还有哪些属性是好的属性但不必要的？**

**27.梯度下降法的神经网络容易收敛到局部最优，为什么应用广泛？**

**28.简单说说CNN常用的几个模型。**

**29.为什么很多做人脸的Paper会最后加入一个Local Connected Conv？**

**30.什么是梯度爆炸？**

**31.梯度爆炸会引发什么问题？**

**32.如何确定是否出现梯度爆炸？**

**33.如何修复梯度爆炸问题？**

**34.LSTM神经网络输入输出究竟是怎样的？**

**35.什么是RNN？**

**36、简单说下sigmoid激活函数**

**37、rcnn、fast-rcnn和faster-rcnn三者的区别是什么**

**38、在神经网络中，有哪些办法防止过拟合？**

**39、CNN是什么，CNN关键的层有哪些？**

**40、GRU是什么？GRU对LSTM做了哪些改动？**

**41、请简述应当从哪些方向上思考和解决深度学习中出现的的over fitting问题？**

**42、神经网络中，是否隐藏层如果具有足够数量的单位，它就可以近似任何连续函数？**

**43、为什么更深的网络更好？**

**44、更多的数据是否有利于更深的神经网络？**

**45、不平衡数据是否会摧毁神经网络？**

**46、你如何判断一个神经网络是记忆还是泛化?**

**47、无监督降维提供的是帮助还是摧毁？**

**48、是否可以将任何非线性作为激活函数?**

**49、批大小如何影响测试正确率？**

**50、损失函数重要吗？**

**51、初始化如何影响训练?**

**52、不同层的权重是否以不同的速度收敛？**

**53、正则化如何影响权重？**

**54、什么是fine-tuning？**

**55、请简单解释下目标检测中的这个IOU评价函数（intersection-over-union）**

**56、什么是边框回归Bounding-Box regression，以及为什么要做、怎么做**

**57、请阐述下Selective Search的主要思想**

**58、什么是非极大值抑制（NMS）？**

**59、什么是深度学习中的anchor？**

**60、CNN的特点以及优势**

**61、深度学习中有什么加快收敛/降低训练难度的方法？**

**62、请简单说下计算流图的前向和反向传播**

**63、请写出链式法则并证明**

**64、请写出Batch Normalization的计算方法及其应用**

**65、神经网络中会用到批量梯度下降（BGD）吗？为什么用随机梯度下降（SGD）?**

**66、当神经网络的调参效果不好时，从哪些角度思考？（不要首先归结于overfiting）**