

深度学习笔试100题

文章目录

文章目录:

这个图片不能够显示，因为我的图床是在github上的，很烦，所以我把这个文档导出成pdf中了，你们可以直接在我上传的文档资源中找到，[下载](#)

1、梯度下降算法的正确步骤是什么？

- 1.计算预测值和真实值之间的误差
 - 2.重复迭代，直至得到网络权重的最佳值
 - 3.把输入传入网络，得到输出值
 - 4.用随机值初始化权重和偏差
 - 5.对每一个产生误差的神经元，调整相应的（权重）值以减小误差
- A、 1, 2, 3, 4, 5
B、 5, 4, 3, 2, 1
C、 3, 2, 1, 5, 4
D、 4, 3, 1, 5, 2

答案：D

解析：

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

答案：A

解析：

更多层意味着网络更深。没有严格的定义多少层的模型才叫深度模型，目前如果有超过2层的隐层，那么也可以叫做深度模型。

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，还是不对？

- A、对
B、不对

答案：A 这题选错

解析：

如寒sir所说，训练CNN时，可以进行这些操作。当然也不一定是必须的，只是data augmentation扩充数据后，模型有更多数据训练，泛化能力可能会变强。

4、下面哪项操作能实现跟神经网络中Dropout的类似效果？

- A、Boosting
- B、Bagging
- C、Stacking
- D、Mapping

答案：B 这题选错

解析：

Dropout可以认为是一种极端的Bagging，每一个模型都在单独的数据上训练，同时，通过和其他模型对应参数的共享，从而实现模型参数的高度正则化。

5、下列哪一项在神经网络中引入了非线性？

- 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

答案：B

解析：

修正线性单元是非线性的激活函数。

- A、学习率(learning rate)太低
- B、正则参数太高
- C、陷入局部最小值
- D、以上都有可能

答案：D

解析：

7、下列哪项关于模型能力（model capacity）的描述是正确的？（指神经网络模型能拟合复杂函数的能力）

- A、隐藏层层数增加，模型能力增加
- B、Dropout的比例增加，模型能力增加
- C、学习率增加，模型能力增加
- D、都不正确

答案：A

解析：

8、如果增加多层感知机（Multilayer Perceptron）的隐藏层层数，分类误差便会减小。这种陈述正确还是错误？

- A、正确
- B、错误

答案：B

解析：

并不总是正确。层数增加可能导致过拟合，从而可能引起错误增加。

下列哪一种架构有反馈连接？

- A、循环神经网络
- B、卷积神经网络
- C、限制玻尔兹曼机
- D、都不是

答案：A

解析：

10、在感知机中（Perceptron）的任务顺序是什么？

- 1 随机初始化感知机的权重
- 2 去到数据集的下一批（batch）
- 3 如果预测值和输出不一致，则调整权重
- 4 对一个输入样本，计算输出值

- A、1, 2, 3, 4
B、4, 3, 2, 1
C、3, 1, 2, 4
D、1, 4, 3, 2

答案：D

解析：

11、假设你需要调整超参数来最小化代价函数（cost function），会使用下列哪项技术？

- A、穷举搜索
B、随机搜索
C、Bayesian优化
D、都可以

答案：D 这题选错，选了C

解析：

答案：C

解析：

网络规模过大时，就可能学到数据中的噪声，导致过拟合

答案：B

解析：

PCA 提取的是数据分布方差比较大的方向，隐藏层可以提取有预测能力的特征

15、下列哪个函数不可以做激活函数？

- A、 $y = \tanh(x)$
B、 $y = \sin(x)$
C、 $y = \max(x, 0)$
D、 $y = 2x$

答案是：D 这题选错啦，选择了 B （激活函数非线性）

解析：

线性函数不能作为激活函数。

16、下列哪个神经网络结构会发生权重共享？

- A、卷积神经网络
B、循环神经网络
C、全连接神经网络
D、选项A和B

答案是：D

解析：

17、批规范化(Batch Normalization)的好处都有啥？

- A、让每一层的输入的范围都大致固定
- B、它将权重的归一化平均值和标准差
- C、它是一种非常有效的反向传播(BP)方法
- D、这些均不是

答案是：A

解析：

18、在一个神经网络中，下面哪种方法可以用来处理过拟合？

- A、Dropout
- B、分批归一化(Batch Normalization)
- C、正则化(regularization)
- D、都可以

答案是：D

解析：

对于选项A, Dropout 可以在训练过程中适度地删减某些神经元, 借此可以减小过拟合的风险.

对于选项B, 分批归一化处理过拟合的原理, 是因为同一个数据在不同批中被归一化后的值会有差别, 相当于做了data augmentatio.

对于选项C, 正则化(regularization)的加入,本身就是为了防止过拟合而做的操作.

因此答案是D

19、如果我们用了一个过大的学习速率会发生什么？

- A、神经网络会收敛
- B、不好说
- C、都不对
- D、神经网络不会收敛

答案是：D 这题选错啦，选择了 B

解析：

可能是我考虑的有些极端啦，大部分情况下，会出现其在震荡，梯度长时间不更新，就是因为学习率较大导致的

答案是：D

解析：

不知道神经网络的权重和偏差是什么，则无法判定它将会给出什么样的输出。

21、神经网络模型（Neural Network）因受人类大脑的启发而得名

神经网络由许多神经元（Neuron）组成，每个神经元接受一个输入，对输入进行处理后给出一个输出，如下图所示。请问下列关于神经元的描述中，哪一项是正确的？

- A、每个神经元可以有一个输入和一个输出
- B、每个神经元可以有多个输入和一个输出
- C、每个神经元可以有一个输入和多个输出
- D、每个神经元可以有多个输入和多个输出
- E、上述都正确

正确答案是：E，的选择是： B

解析：

每个神经元可以有一个或多个输入， 和一个或多个输出。

22、在一个神经网络中，知道每一个神经元的权重和偏差是最重要的一步。如果知道了神经元准确的权重和偏差，便可以近似任何函数，但怎么获知每个神经的权重和偏移呢？

- A、搜索每个可能的权重和偏差组合，直到得到最佳值
- B、赋予一个初始值，然后检查跟最佳值的差值，不断迭代调整权重
- C、随机赋值，听天由命
- D、以上都不正确的

正确答案是： B， 您的选择是： B

解析：

选项B是对梯度下降的描述。

23、基于二次准则函数的H-K算法较之于感知器算法的优点是()？

- A、计算量小
- B、可以判别问题是否线性可分
- C、其解完全适用于非线性可分的情况

正确答案是： B， 您的选择是： A

解析：

24、输入图片大小为 200×200 ，依次经过一层卷积（kernel size 5×5 ，padding 1，stride 2），pooling（kernel size 3×3 ，padding 0，stride 1），又一层卷积（kernel size 3×3 ，padding 1，stride 1）之后，输出特征图大小为

- A、95
- B、96
- C、97
- D、98

正确答案是： C， 您的选择是： C

解析：

首先我们应该知道卷积或者池化后大小的计算公式，其中，padding指的是向外扩展的边缘大小，而stride则是步长，即每次移动的长度。这样一来就容易多了，首先长宽一般大，所以我们只需要计算一个维度即可，这样，经过第一次卷积后的大小为: 本题 $(200 - 5 + 2 * 1) / 2 + 1$ 为99.5，取99， 向下取整

经过第一次池化后的大小为： $(99 - 3) / 1 + 1$ 为97

经过第二次卷积后的大小为： $(97 - 3 + 2 * 1) / 1 + 1$ 为97

25、深度学习是当前很热门的机器学习算法，在深度学习中，涉及到大量的矩阵相乘，现在需要计算三个稠密矩阵A,B,C的乘积ABC,假设三个矩阵的尺寸分别为 $m \times n$ ， $n \times p$ ， $p \times q$ ，且 $m < n < p < q$ ，以下计算顺序效率最高的是（）

- A、(AB)C
- B、AC(B)
- C、A(BC)
- D、所以效率都相同

正确答案是： A， 您的选择是： C

解析：

首先，根据简单的矩阵知识，因为 $A * B$ ， A 的列数必须和 B 的行数相等。因此，可以排除 B 选项，

然后，再看 A、C 选项。在 A 选项中， $m \times n$ 的矩阵 A 和 $n \times p$ 的矩阵 B 的乘积，得到 $m \times p$ 的矩阵 $A * B$ ，而 $A * B$ 的每个元素需要 n 次乘法和 $n - 1$ 次加法，忽略加法，共需要 $m * n * p$ 次乘法运算。同样情况分析 $A * B$ 之后再乘以 C 时的情况，共需要 $m * p * q$ 次乘法运算。因此，A 选项 (AB)C 需要的乘法次数是 $m * n * p + m * p * q$ 。同理分析，C 选项 A (BC) 需要的乘法次数是 $n * p * q + m * n * q$ 。

由于 $m * n * p$

26、当在卷积神经网络中加入池化层(pooling layer)时，变换的不变性会被保留，是吗？

- A、不知道
- B、看情况
- C、是
- D、否

正确答案是：C，您的选择是：C

解析：

池化算法比如取最大值/取平均值等，都是输入数据旋转后结果不变，所以多层叠加后也有这种不变性。

27、当数据过大以至于无法在RAM中同时处理时，哪种梯度下降方法更加有效？

- A、随机梯度下降法(Stochastic Gradient Descent)
- B、不知道
- C、整批梯度下降法(Full Batch Gradient Descent)
- D、都不是

正确答案是：A，您的选择是：A

解析：

梯度下降法分随机梯度下降(每次用一个样本)、小批量梯度下降法(每次用一小批样本算出总损失，因而反向传播的梯度折中)、全批量梯度下降法则一次性使用全部样本。这三个方法，对于全体样本的损失函数曲面来说，梯度指向一个比一个准确。但是在工程应用中，受到内存/磁盘IO的吞吐性能制约，若要最小化梯度下降的实际运算时间，需要在梯度方向准确性和数据传输性能之间取得最好的平衡。所以，对于数据过大以至于无法在RAM中同时处理时，RAM每次只能装一个样本，那么只能选随机梯度下降法。

28、在选择神经网络的深度时，下面哪些参数需要考虑？

- 1 神经网络的类型(如MLP,CNN)
 - 2 输入数据
 - 3 计算能力(硬件和软件能力决定)
 - 4 学习速率
 - 5 映射的输出函数
- A、1,2,4,5
 - B、2,3,4,5
 - C、都需要考虑
 - D、1,3,4,5

正确答案是：C，您的选择是：C

解析：

所有上述因素对于选择神经网络模型的深度都是重要的。特征抽取所需分层越多，输入数据维度越高，映射的输出函数非线性越复杂，所需深度就越深。另外为了达到最佳效果，增加深度所带来的参数量增加，也需要考虑硬件计算能力和学习速率以设计合理的训练时间。

29、考虑某个具体问题，你可能只有少量数据来解决这个问题。不过幸运的是你有一个类似问题已经预先训练好的神经网络。可以用下面哪种方法来利用这个预先训练好的网络？

- A、把除了最后一层外所有的层都冻结，重新训练最后一层
- B、对新数据重新训练整个模型
- C、只对最后几层进行调参(fine tune)
- D、对每一层模型进行评估，选择其中的少数来用

正确答案是：C，您的选择是：C

解析：

如果有个预先训练好的神经网络，就相当于网络各参数有个很靠谱的先验代替随机初始化。若新的少量数据来自于先前训练数据(或者先前训练数据量很好地描述了数据分布，而新数据采样自完全相同的分布)，则冻结前面所有层而重新训练最后一层即可；但一般情况下，新数据分布跟先前训练集分布有所偏差，所以先验网络不足以完全拟合新数据时，可以冻结大部分前层网络，只对最后几层进行训练调参(这也称之为fine tune)。

正确答案是：A，您的选择是：C

解析：

由于反向传播算法从后向前传播的过程中，学习能力降低，这就是梯度消失。换言之，梯度消失是梯度在反向传播中逐渐减为 0，按照图标题所说，四条曲线是 4 个隐藏层的学习曲线，那么最后一层梯度最高（损失函数曲线下降明显），第一层梯度几乎为零（损失函数曲线变成平直线）。所以 D 是第一层，A 是最后一层。

31、增加卷积核的大小对于改进卷积神经网络的效果是必要的吗？

- A
没听说过
B
是
C
否
D
不知道

正确答案是：C，您的选择是： B

解析：

增加核函数的大小不一定会提高性能。这个问题在很大程度上取决于数据集。我以为是卷积核的个数呢，哎！

32、假设我们已经在ImageNet数据集(物体识别)上训练好了一个卷积神经网络。然后给这张卷积神经网络输入一张全白的图片。对于这个输入的输出结果为任何种类的物体的可能性都是一样的，对吗？

- A、对的
B、不知道
C、看情况
D、不对

正确答案是：D，您的选择是：D

解析：

D，已经训练好的卷积神经网络，各个神经元已经精雕细作完工，对于全白图片的输入，其j层层激活输出给最后的全连接层的值几乎不可能恒等，再经softmax转换之后也不会相等，所以"输出结果为任何种类的等可能性一样"也就是softmax的每项均相等，这个概率是极低的。

33、对于一个分类任务，如果开始时神经网络的权重不是随机赋值的，而是都设成0，下面哪个叙述是正确的？

- A、其他选项都不对
B、没啥问题，神经网络会正常开始训练
C、神经网络可以训练，但是所有的神经元最后都会变成识别同样的东西
D、神经网络不会开始训练，因为没有梯度改变

正确答案是：C，您的选择是： B

解析：

令所有权重都初始化为0这个一个听起来还蛮合理的想法也许是一个我们假设中最好的一个假设了，但结果是错误的，因为如果神经网络计算出来的输出值都一个样，那么反向传播算法计算出来的梯度值一样，并且参数更新值也一样($w = w - \alpha * dw$)。更一般地说，如果权重初始化为同一个值，网络即是对称的，最终所有的神经元最后都会变成识别同样的东西。

正确答案是：A，您的选择是： B

解析：

选项A可以将陷于局部最小值的神经网络提取出来。总感觉题目出的有些不太严谨

5、对于一个图像识别问题(在一张照片里找出一只猫)，下面哪种神经网络可以更好地解决这个问题？

- A、循环神经网络
B、感知机
C、多层感知机
D、卷积神经网络

正确答案是：D，您的选择是：D

解析：

卷积神经网络将更好地适用于图像相关问题，因为考虑到图像附近位置变化的固有性质。

正确答案是：D，您的选择是：A

解析：

首先您将相关的数据去掉，然后将其置零。具体来说，误差瞬间降低，一般原因是多个数据样本有强相关性且突然被拟合命中，或者含有较大方差数据样本突然被拟合命中。所以对数据作主成分分析(PCA)和归一化能够改善这个问题。

正确答案是：B，您的选择是：B

解析：

选项B是正确的，为了减少这些“起伏”，可以尝试增加批尺寸(batch size)。具体来说，在曲线整体趋势为下降时，为了减少这些“起伏”，可以尝试增加批尺寸(batch size)以缩小batch综合梯度方向摆动范围。当整体曲线趋势为平缓时出现可观的“起伏”，可以尝试降低学习率以进一步收敛。“起伏”不可观时应该提前终止训练以免过拟合

39、对于神经网络的说法,下面正确的是：

- 1、 增加神经网络层数, 可能会增加测试数据集的分类错误率
- 2、 减少神经网络层数, 总是能减小测试数据集的分类错误率
- 3、 增加神经网络层数, 总是能减小训练数据集的分类错误率

- A、 1
 - B、 1 和 3
 - C、 1 和 2
 - D、 2
-

40、假定你在神经网络中的隐藏层中使用激活函数 X。在特定神经元给定任意输入，你会得到输出「-0.0001」。X 可能是以下哪一个激活函数？

- A、 ReLU
- B、 tanh
- C、 SIGMOID
- D、 以上都不是

正确答案是：B，您的选择是：B

解析：

答案为：B，该激活函数可能是 tanh，因为该函数的取值范围是 (-1,1)。

41、深度学习与机器学习算法之间的区别在于，后者过程中无需进行特征提取工作，也就是说，我们建议在进行深度学习过程之前要首先完成特征提取的工作。这种说法是：

- A、 正确的
- B、 错误的

正确答案是：B，您的选择是：B

解析：

正好相反，深度学习可以自行完成特征提取过程而机器学习需要人工来处理特征内容。

42、下列哪一项属于特征学习算法（representation learning algorithm）？

- A、 K近邻算法
- B、 随机森林
- C、 神经网络
- D、 都不属于

正确答案是：C，您的选择是：C

解析：

神经网络会将数据转化为更适合解决目标问题的形式，我们把这种过程叫做特征学习。

43、下列哪些项所描述的相关技术是错误的？

- A、AdaGrad使用的是一阶差分(first order differentiation)
- B、L-BFGS使用的是二阶差分(second order differentiation)
- C、AdaGrad使用的是二阶差分

正确答案是：C，您的选择是：A

解析： 记住吧，后面再去看看这些优化技术的原理

44、提升卷积核(convolutional kernel)的大小会显著提升卷积神经网络的性能，这种说法是

- A、正确的
- B、错误的

正确答案是： B，您的选择是： B

解析：

卷积核的大小是一个超参数(hyperparameter)，也就意味着改变它既有可能提高亦有可能降低模型的表现。

45、阅读以下文字：

假设我们拥有一个已完成训练的、用来解决车辆检测问题的深度神经网络模型，训练所用的数据集由汽车和卡车的照片构成，而训练目标是检测出每种车辆的名称（车辆共有10种类型）。现在想要使用这个模型来解决另外一个问题，问题数据集中仅包含一种车（福特野马）而目标变为定位车辆在照片中的位置。

- A、除去神经网络中的最后一层，冻结所有层然后重新训练
- B、对神经网络中的最后几层进行微调，同时将最后一层（分类层）更改为回归层
- C、使用新的数据集重新训练模型
- D、所有答案均不对

正确答案是： B，您的选择是： B

解析：

46、假设你有5个大小为7x7、边界值为0的卷积核，同时卷积神经网络第一层的深度为1。此时如果你向这一层传入一个维度为224x224x3的数据，那么神经网络下一层所接收到的数据维度是多少？

- A、218x218x5
- B、217x217x8
- C、217x217x3
- D、220x220x5

正确答案是： A，您的选择是： A

解析：

47、假设我们有一个使用ReLU激活函数(ReLU activation function)的神经网络，假如我们把ReLU激活替换为线性激活，那么这个神经网络能够模拟出同或函数(XNOR function)吗？

- A、可以
- B、不好说
- C、不一定
- D、不能

正确答案是：D，您的选择是：D

解析：

使用ReLU激活函数的神经网络是能够模拟出同或函数的。

但如果ReLU激活函数被线性函数所替代之后，神经网络将失去模拟非线性函数的能力。

48、考虑以下问题：

假设我们有一个5层的神经网络，这个神经网络在使用一个4GB显存显卡时需要花费3个小时来完成训练。而在测试过程中，单个数据需要花费2秒的时间。如果我们现在把架构变换一下，当评分是0.2和0.3时，分别在第2层和第4层添加Dropout，那么新架构的测试所用时间会变为多少？

- A、少于2s
- B、大于2s
- C、仍是2s
- D、说不准

正确答案是：C，您的选择是：A

解析：

在架构中添加Dropout这一改动仅会影响训练过程，而并不影响测试过程。

9、下列的哪种方法可以用来降低深度学习模型的过拟合问题？

- 1 增加更多的数据
- 2 使用数据扩增技术(data augmentation)
- 3 使用归纳性更好的架构
- 4 正规化数据
- 5 降低架构的复杂度

- A、1 4 5
- B、1 2 3
- C、1 3 4 5
- D、所有项目都有用

正确答案是：D，您的选择是：D

解析：

上面所有的技术都会对降低过拟合有所帮助。

50、混沌度(Perplexity)是一种常见的应用在使用深度学习处理NLP问题过程中的评估技术，关于混沌度，哪种说法是正确的？

- A、混沌度没什么影响
- B、混沌度越低越好
- C、混沌度越高越好
- D、混沌度对于结果的影响不一定

正确答案是：B，您的选择是：D

解析：对NLP没有怎么学过，也不了解这个概念，有感兴趣的自己去查吧

51、假设下方是传入最大池化层的一个输入，该层中神经元的池化大小为(3,3)。

3	4	5
4	5	6

那么，该池化层的输出是多少？

- A、3
- B、7
- C、5
- D、5.5

正确答案是： B，您的选择是： B

解析：

最大池化是这样运作的：它首先使用我们预先定义的池化尺寸来获取输入内容，然后给出有效输入中的最大值。

正确答案是： B，您的选择是： B

解析：

Affine是神经网络中的一个全连接层，仿射（Affine）的意思是前面一层中的每一个神经元都连接到当前层中的每一个神经元。即当前层的每一个神经元输入均为增广权值向量和前一层神经元向量的内积，**本质是一个线性变换**。而relu是一个常用的非线性激活函数，其表达为 $\max(0, x)$ ，如果去掉Relu层，该神经网络将失去非线性表达能力，即无法拟合复杂世界中的非线性函数规律，故答案选B。

53、深度学习可以用在下列哪些NLP任务中？

- A、情感分析
- B、问答系统
- C、机器翻译
- D、所有选项

正确答案是： D，您的选择是： D

解析：

深度学习可以用来解决上述所有NLP问题。

正确答案是： B，您的选择是： A

解析：

情景1基于欧几里得数据(Euclidean data)而情景2基于图形数据，这两种类型的数据深度学习均可处理。

55、下列哪些项目是在图像识别任务中使用的数据扩增技术(data augmentation technique)？

- 1 水平翻转(Horizontal flipping)
 - 2 随机裁剪(Random cropping)
 - 3 随机放缩(Random scaling)
 - 4 颜色抖动(Color jittering)
 - 5 随机平移(Random translation)
 - 6 随机剪切(Random shearing)
- A、1, 3, 5, 6
 - B、1, 2, 4
 - C、2, 3, 4, 5, 6
 - D、所有项目

正确答案是： D，您的选择是： D

解析：

56、给定一个长度为n的不完整单词序列，我们希望预测下一个字母是什么。比如输入是“predictio” (9个字母组成)，希望预测第十个字母是什么。下面哪种神经网络结构适用于解决这个问题？

- A、循环神经网络
- B、全连接神经网络
- C、受限波尔兹曼机
- D、卷积神经网络

正确答案是：A，您的选择是：A

解析：

循环神经网络对于序列数据最有效，因此适用于这个问题。

57、当构建一个神经网络进行图片的语义分割时，通常采用下面哪种顺序？

- A、先用卷积神经网络处理输入，再用反卷积神经网络得到输出
- B、先用反卷积神经网络处理输入，再用卷积神经网络得到输出
- C、不能确定

正确答案是：A，您的选择是：A

解析：

58、Sigmoid是神经网络中最常用到的一种激活函数，除非当梯度太大导致激活函数被弥散，这叫作神经元饱和。

这就是为什么ReLU会被提出来，因为ReLU可以使得梯度在正向时输出值与原始值一样。

这是否意味着在神经网络中ReLU单元永远不会饱和？

- A、正确的
- B、错误的

正确答案是：B，您的选择是：B

解析：

ReLU也可能造成饱和，当输出为负的时候。

59、Dropout率和正则化有什么关系？

提示：我们定义Dropout率为保留一个神经元为激活状态的概率

- A、Dropout率越高，正则化程度越低
- B、Dropout率越高，正则化程度越高

正确答案是：A，您的选择是：B

解析：

高dropout率意味着更多神经元是激活的，所以这亦为之正则化更少。

(这个解释我也是不太明白，它说的可能是和API有关系吧，tensorflow和pytorch对于droupout的大小是相相反的)

60、普通反向传播算法和随时间的反向传播算法（BPTT）有什么技术上的不同？

- A、与普通反向传播不同的是，BPTT会在每个时间步长内减去所有对应权重的梯度
- B、与普通反向传播不同的是，BPTT会在每个时间步长内叠加所有对应权重的梯度

正确答案是：B，您的选择是：A

解析：

与普通反向传播不同的是，BPTT会在每个时间步长内叠加所有对应权重的梯度。

下面的题目错的有点过分啦（捂脸）

61、梯度爆炸问题是指在训练深度神经网络的时候，梯度变得过大而损失函数变为无穷。在RNN中，下面哪种方法可以较好地处理梯度爆炸问题？

- A、用改良的网络结构比如LSTM和GRUs
- B、梯度裁剪
- C、Dropout
- D、所有方法都不行

正确答案是：B，您的选择是：A

解析：

为了处理 梯度爆炸 问题，最好 让权重的梯度更新限制在一个合适的范围。（梯度裁剪）

62、有许多种梯度下降算法，其中两种最出名的方法是I-BFGS和SGD。I-BFGS根据二阶梯度下降而SGD是根据一阶梯度下降的。

在下述哪些场景中，会更加偏向于使用I-BFGS而不是SGD？

场景1： 数据很稀疏

场景2： 神经网络的参数数量较少

A、场景1

B、场景2

C、两种情况都是

D、都不会选择I-BFGS

正确答案是：C，您的选择是：A

解析：

在这两种情况下，I-BFGS的效果都是最好的

63、下面哪种方法没办法直接应用于自然语言处理的任务？

A、去语法模型

B、循环神经网络

C、卷积神经网络

D、主成分分析(PCA)

正确答案是：D，您的选择是：C

解析：我以为卷积主要是应用在图像处理，NLP应该是RNN，结果。。。

64、对于*非连续*目标在深度神经网络的优化过程中，下面哪种梯度下降方法是最好的？

A、SGD

B、AdaGrad

C、I-BFGS

D、拉格朗日松弛Subgradient method

正确答案是：D，您的选择是：A

解析：

优化算法无法作用于非连续目标。

65、下面哪个叙述是对的？

Dropout对一个神经元随机屏蔽输入权重

Dropconnect对一个神经元随机屏蔽输入和输出权重

A、1是对的，2是错的

B、都是对的

C、1是错的，2是对的

D、都是错的

正确答案是：D，您的选择是：A

解析：

在dropout的过程中，神经元被失活，在dropconnect的过程中，失活的是神经元之间的连接。

所以dropout会使输入和输出权重都变为无效，而在dropconnect中，只有其中一种会被失活。

在上图中，最好在哪个时间停止训练？

A、D

B、A

C、C

D、B

正确答案是：C，您的选择是：C

解析：

你最好在模型最完善之前「提前终止」，所以C点是正确的。

67、图片修复是需要人类专家来进行修复的，这对于修复受损照片和视频非常有帮助。下图是一个图像修复的例子。

正确答案是：C，您的选择是： B

解析：看错啦，以为A是交叉熵函数，一般交叉熵损失函数应用于

68、反向传播算法一开始计算什么内容的梯度，之后将其反向传播？

- A、预测结果与样本标签之间的误差
- B、各个输入样本的平方差之和
- C、各个网络权重的平方差之和
- D、都不对

正确答案是：A，您的选择是：C

解析：思维絮乱，导致选错

69、在构建一个神经网络时，batch size通常会选择2的次方，比如256和512。这是为什么呢？

- A、当内存使用最优时这可以方便神经网络并行化
- B、当用偶数是梯度下降优化效果最好
- C、这些原因都不对
- D、当不用偶数时，损失值会很奇怪

正确答案是：A，您的选择是：A

Xavier初始化是用来帮助信号能够在神经网络中传递得更深，下面哪些叙述是对的？

- 1 如果权重一开始很小，信号到达最后也会很小
- 2 如果权重一开始很大，信号到达最后也会很大
- 3 Xavier初始化是由高斯分布引出的
- 4 Xavier初始化可以帮助减少梯度弥散问题

- A、234
- B、1234
- C、124
- D、134

正确答案是：B，您的选择是：B

解析：

所有项目都是正确的

71、随着句子的长度越来越多，神经翻译机器将句意表征为固定维度向量的过程将愈加困难，为了解决这类问题，下面哪项是我们可以采用的？

- A、使用递归单元代替循环单元
 - B、使用注意力机制(attention mechanism)
 - C、使用字符级别翻译(character level translation)
 - D、所有选项均不对
-

72、一个循环神经网络可以被展开成为一个完全连接的、具有无限长度的普通神经网络，这种说法是

- A、正确的
- B、错误的

正确答案是：A，您的选择是： A

解析：

循环神经元可以被认为是一个具有无限时间长度的神经元序列。

根据以上描述，Dropout技术在下列哪种神经层中将无法发挥显著优势？

- A、仿射层
- B、卷积层
- C、RNN层
- D、均不对

正确答案是：C，您的选择是： A

解析：

Dropout对于循环层效果并不理想，您可能需要稍微修改一下dropout技术来得到良好的结果。

仿射层：Affine 仿射层, 又称 Linear 线性变换层, 常用于神经网络结构中的全连接层.

74、假设你的任务是使用一首曲子的前半部分来预测乐谱的未来几个小节，比如输入的内容如下的是包含音符的一张乐谱图片：

正确答案是： B，您的选择是： B

解析：

75、当在内存网络中获得某个内存空间时，通常选择读取矢量形式数据而不是标量，这里需要的哪种类型的寻址来完成？

- A、基于内容的寻址
- B、基于位置的寻址
- C、都不行
- D、都可以

正确答案是：A，您的选择是： B

解析：不明白这一题考察的是什么，反正我是不懂是为什么

76、一般我们建议将卷积生成对抗网络（convolutional generative adversarial nets）中生成部分的池化层替换成什么？

- A、跨距卷积层(Strided convolutional layer)
 - B、ReLU层
 - C、局部跨距卷积层(Fractional strided convolutional layer)
 - D、仿射层(Affine layer)
-

正确答案是： B，您的选择是： B

解析：

二维平面没有直线可以分开两类样本点，线性不可分。

78、以下哪些是通用逼近器？

- A、Kernel SVM
- B、Neural Networks
- C、Boosted Decision Trees
- D、以上所有

正确答案是：D，您的选择是： D

解析：

79、在下列哪些应用中，我们可以使用深度学习来解决问题？

- A、蛋白质结构预测
- B、化学反应的预测
- C、外来粒子的检测
- D、所有这些

正确答案是：D，您的选择是：A

解析：

我们可以使用神经网络来逼近任何函数，因此理论上可以用它来解决任何问题。

这个解析我也是无力反驳，这么说它也是除了不能生小孩，什么都能做啦！

80、在CNN中使用 1×1 卷积时，下列哪一项是正确的？

- A、它可以帮助降低维数
- B、可以用于特征池
- C、由于小的内核大小，它会减少过拟合
- D、所有上述

正确答案是：D，您的选择是：C

解析：

1×1 卷积在CNN中被称为bottleneck structure（瓶颈结构）。更多关于 1×1 卷积的作用和好处，可以参考

81、声明1：可以通过将所有权重初始化为0 来训练网络。

声明2：可以通过将偏差初始化为0来很好地训练网络

以上哪些陈述是真实的？

- A、1对2错
- B、1错2对
- C、1和2都对
- D、1和2都错

正确答案是：B，您的选择是：B

解析：

即使所有的偏差都为零，神经网络也有可能学习。另一方面，如果所有的权重都是零；神经网络可能永远不会学习执行任务，因为梯度从一开始就不能达到更新。

82、对于MLP，输入层中的节点数为10，隐藏层为5.从输入层到隐藏层的最大连接数是

- A、50
- B、小于50
- C、超过50
- D、这是一个任意值

正确答案是：A，您的选择是：A

解析：

由于MLP是完全连通的有向图，因此连接数是输入层和隐藏层中节点数的乘积。

83、输入图像已被转换为大小为 28×28 的矩阵和大小为 7×7 的步幅为1的核心/滤波器。卷积矩阵的大小是多少？

- A、 22×22
- B、 21×21
- C、 28×28
- D、 7×7

正确答案是：A，您的选择是：A

解析：

解决方案：A

卷积矩阵的大小由 $C = ((I-F + 2P) / S) + 1$ 给出，其中C是卷积矩阵的大小，I是输入矩阵的大小，F是滤波器的大小矩阵和P填充应用于输入矩阵。这里 $P = 0$ ， $I = 28$ ， $F = 7$ 和 $S = 1$ 。答案是22。

84、在一个简单的MLP模型中，输入层有8个神经元，隐藏层有5个神经元，输出层有1个神经元。隐藏输出层和输入隐藏层之间的权重矩阵的大小是多少？

- A、[1 X 5]，[5 X 8]
- B、[8×5]，[1×5]
- C、[5×8]，[5×1]
- D、[5×1]，[8×5]

正确答案是：D，您的选择是：D

解析：

任何层1和层2之间的权重的大小由[层1中的节点X 2层中的节点]

85、如果我们希望预测n个类 ($p_1, p_2 \dots p_k$) 的概率使得所有n的p的和等于1，那么下列哪个函数可以用作输出层中的激活函数？

- A、Softmax
- B、ReLU
- C、Sigmoid
- D、Tanh

正确答案是：A，您的选择是：A

解析：

Softmax函数的形式是所有k的概率之和总和为1。

86、假设一个具有 3 个神经元和输入为 [1, 2, 3] 的简单 MLP 模型。输入神经元的权重分别为 4, 5 和 6。假设激活函数是一个线性常数值 3 (激活函数为: $y = 3x$)。输出是什么？

- A、32
- B、643
- C、96
- D、48

正确答案是：C，您的选择是：C

解析：

输出将被计算为 $3 (1 * 4 + 2 * 5 + 6 * 3) = 96$

87、在输出层不能使用以下哪种激活函数来分类图像？

- A、sigmoid
- B、Tanh
- C、ReLU
- D、If ($x > 5, 1, 0$)

正确答案是：C，您的选择是： B

解析：

解答：C

ReLU在0到无限的范围内提供连续输出。但是在输出层中，我们需要一个有限范围的值。所以选项C是正确的。

88、在神经网络中，每个参数可以有不同的学习率。这句话是对还是错

- A、对
- B、错

正确答案是：A，您的选择是： B

解析：

是的，我们可以定义每个参数的学习率，并且它可以与其他参数不同。

89、使用批量归一化可以解决以下哪种神经网络的训练？

- A、过拟合Overfitting
- B、Restrict activations to become too high or low
- C、训练过慢
- D、B和C都有

正确答案是：D，您的选择是： A

解析：

Batch normalization restricts the activations and indirectly improves training time.

- A、1
- B、2
- C、任何一个
- D、都不用

正确答案是：C，您的选择是： C

解析：

我们可以使用一个神经元作为二值分类问题的输出或两个单独的神经元。

- A、绿色曲线
- B、蓝色曲线

正确答案是： B，您的选择是： B

解析：

蓝色曲线表示过拟合overfitting,绿色曲线表示泛化generalized.

92、使用深度学习的情感分析是多对一的预测任务

- A、对
- B、错

正确答案是：A，您的选择是： B

解析：

选项A是正确的。这是因为从一系列单词中，你必须预测情绪是积极的还是消极的。

93、我们可以采取哪些措施来防止神经网络中的过拟合？

- A、数据增强
- B、权重共享
- C、提前停止
- D、Dropout
- E、以上全部

正确答案是：E，您的选择是： E

解析：

解决方案：E

上述所有方法都可以帮助防止过度配合问题。

94、Gated Recurrent units的出现可以帮助防止在RNN中的梯度消失问题。

- A、对
- B、错

正确答案是：A，您的选择是：A

解析：

选项A是正确的。This is because it has implicit memory to remember past behavior.

95、What does a neuron compute?

- A、A neuron computes the mean of all features before applying the output to an activation function
- B、A neuron computes a function g that scales the input x linearly ($Wx + b$)
- C、A neuron computes an activation function followed by a linear function ($z = Wx + b$)
- D、A neuron computes a linear function ($z = Wx + b$) followed by an activation function

正确答案是：D，您的选择是：D

解析：

一个神经元先计算线性函数（linear function），然后计算激活函数（activation function）。

也就是说，输入是 x 的话，先计算 $z = Wx + b$ ，再把 z 作为输入计算sigmoid (z)，显然这里我们假设激活函数是sigmoid。

96、You are building a binary classifier for recognizing cucumbers ($y=1$) vs. watermelons ($y=0$). Which one of these activation functions would you recommend using for the output layer?

- A、ReLU
- B、Leaky ReLU
- C、sigmoid
- D、tanh

正确答案是：C，您的选择是：C

解析：

对于逻辑回归问题，sigmoid作为激活函数更加合适。答案是c，relu函数不可以，因为它不可以让最后的输出值规范化到某一个范围，tanh的值范围有负值。

97、Which of the following are reasons for using feature scaling?

- A、It prevents the matrix XTX (used in the normal equation) from being non-invertible (singular/degenerate).
 - B、It speeds up gradient descent by making it require fewer iterations to get to a good solution.
 - C、It speeds up gradient descent by making each iteration of gradient descent less expensive to compute.
 - D、It is necessary to prevent the normal equation from getting stuck in local optima.
-

98、在CNN，拥有最大池总是减少参数？

- A、对
- B、错

正确答案是：B，您的选择是：B

解析：

这并非总是如此。如果我们将池大小的最大池层设置为1，则参数将保持不变。

99、将Sigmoid激活函数改为ReLU，将有助于克服梯度消失问题？

- A、对
- B、错

正确答案是：A，您的选择是：A

解析：

ReLU可以帮助解决梯度消失问题

100、多义现象可以被定义为在文本对象中一个单词或短语的多种含义共存。

下列哪一种方法可能是解决此问题的最好选择？

- A、随机森林分类器
- B、卷积神经网络
- C、梯度爆炸
- D、上述所有方法

正确答案是： B，您的选择是： B

解析：

CNN 是文本分类问题中比较受欢迎的选择，因为它们把上下文的文本当作特征来考虑，这样可以解决多义问题



