



山东大学
SHANDONG UNIVERSITY

信息基础2 机器学习与深度学习 测试题

陈雷

山东大学信息科学与工程学院

单选：

1. 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
 - D. 以上都不正确

答案： B



2. 对于多层感知器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

答案： A



3. 对于一个图像识别问题(如：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？（ ）

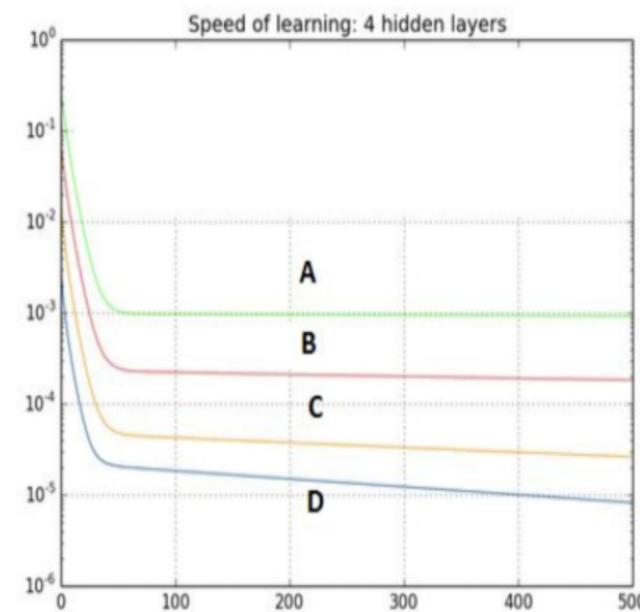
- A. 循环神经网络
- B. 感知机
- C. 多层感知机
- D. 卷积神经网络

答案： D



4. 下图是一个利用sigmoid函数作为激活函数的含四个隐藏层的神经网络训练的梯度下降图。这个神经网络遇到了梯度消失的问题。下面哪个叙述是正确的？（ ）

- A. 第一隐藏层对应D，第二隐藏层对应C，第三隐藏层对应B，第四隐藏层对应A
- B. 第一隐藏层对应A，第二隐藏层对应C，第三隐藏层对应B，第四隐藏层对应D
- C. 第一隐藏层对应A，第二隐藏层对应B，第三隐藏层对应C，第四隐藏层对应D
- D. 第一隐藏层对应B，第二隐藏层对应D，第三隐藏层对应C，第四隐藏层对应A



答案：A



5. 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

答案： C



6. 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. 所有效率都相同

答案： A



7. 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. 不确定
- B. 根据其他层决定
- C. 是
- D. 否

答案： C



8. 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

答案： B



9. 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. 各个输出的平方差之和
- B. 各个输入的平方差之和
- C. 各个权重的平方差之和
- D. 都不正确

答案： A



10. 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. 不确定
- C. 根据神经网络结构决定
- D. 不能

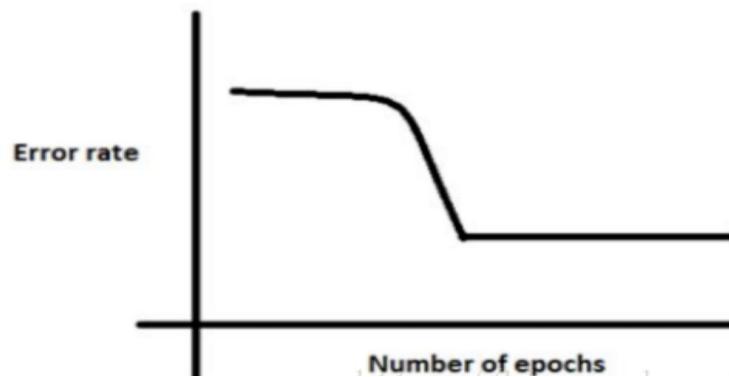
答案： D



11. 下图显示，当开始训练时，误差一直很高，这是因为神经网络在往全局最小值前进之前一直被卡在局部最小值里。为了避免这种情况，我们可以采取下面哪种策略？（ ）

- A. 改变学习速率，比如一开始的几个训练周期不断更改学习速率
- B. 一开始将学习速率减小10倍，然后用动量项(momentum)
- C. 增加参数数目，这样神经网络就不会卡在局部最优处
- D. 其他都不对

答案： A



12. 假设你有5个大小为 7×7 、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 $224 \times 224 \times 3$ 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. $218 \times 218 \times 5$
- B. $217 \times 217 \times 8$
- C. $217 \times 217 \times 3$
- D. $220 \times 220 \times 5$

答案： A



13. 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（如：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

答案： B



14. 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

答案: C



15. 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. 是
- B. 否

答案： B



16. 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. $[1 \times 5]$, $[5 \times 8]$
- B. $[8 \times 5]$, $[1 \times 5]$
- C. $[5 \times 8]$, $[5 \times 1]$
- D. $[5 \times 1]$, $[8 \times 5]$

答案： D



17. 梯度下降算法的正确步骤是什么？（ ）
- a.计算预测值和真实值之间的误差 b.重复迭代，直至得到网络权重的最佳值 c.把输入传入网络，得到输出值 d.用随机值初始化权重和偏差 e.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. abcde
B. edcba
C. cbaed
D. dcaeb

答案： D



18. 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. 不对
- B. 对

答案： B



19. 考虑某个具体问题时，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

答案： C



20. 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. 不对
- C. 不确定
- D. 没有输出结果

答案： B



21. 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. 整批梯度下降法(Full Batch Gradient Descent)
- C. 两种方法都可以
- D. 两种方法都不可以

答案： A



22. 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. 不确定

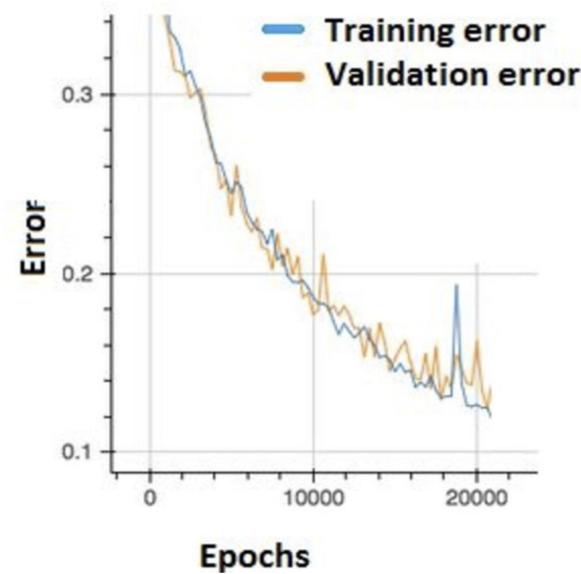
答案： C



23. 在下图中，我们可以观察到误差出现了许多小的"涨落"。这种情况我们应该担心吗？（ ）

- A. 需要，这也许意味着神经网络的学习速率存在问题
- B. 不需要，只要在训练集和交叉验证集上有累积的下降就可以了
- C. 不确定
- D. 不会出现这种情况

答案： B



24. 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. 神经网络会正常开始训练
- B. 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
- C. 神经网络不会开始训练，因为没有梯度改变
- D. 其他选项都不对

答案： B



多选

25. 关于神经网络模型（**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。请问下列关于神经元的描述中，哪项是正确的？（ ）

- 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

答案： **ABCD**



26. 在CNN中使用 1×1 卷积时，下列哪项是正确的？（ ）

- 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. 以上都不正确

答案：ABC



27. 下列哪些函数可以做激活函数? ()

A. $y = \tanh(x)$

B. $y = \sin(x)$

C. $y = \max(x, 0)$

D. $y = 2x$

答案: ABC



28. 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. 图像的风格转换

答案： ABCD



29. 下面哪个决策边界是神经网络生成的？（ ）

A. A

A

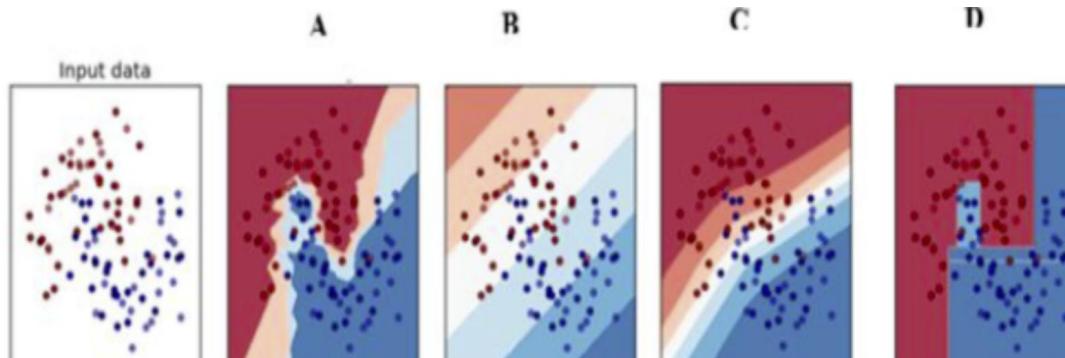
B. D

D

C. C

C

D. B



答案：ABCD



30. 使用批量归一化可以解决以下哪种神经网络的训练?
()

- A. Overfitting
- B. Restrict activations to become too high or low
- C. 训练过慢
- D. 欠拟合

答案： BC



30. 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 减少更多的数据 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 提高架构的复杂度

A. 1

B. 2 3

C. 4

D. 5

答案： BC



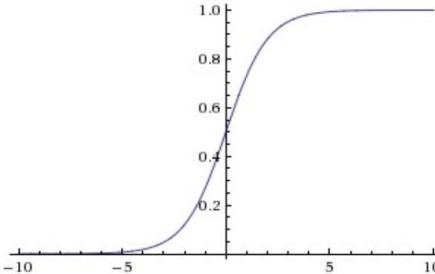
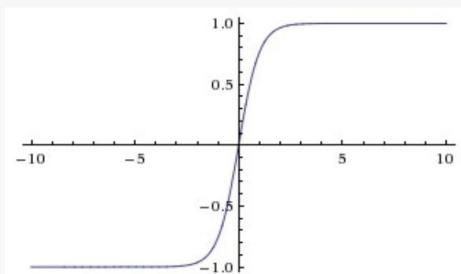
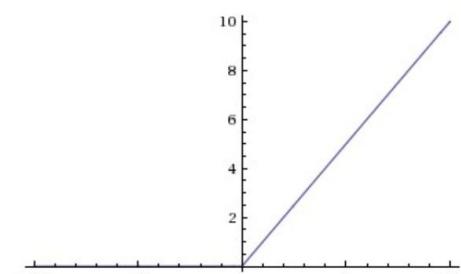
简答：

1. 在神经网络中，为什么要使用激活函数，并简单对比下列常用激活函数的优缺点。

如果不使用激活函数，每一层输出都是上层输入的线性函数，无论神经网络有多少层，输出都是输入的线性组合。

激活函数给神经元引入了非线性因素，使得神经网络可以任意逼近任何非线性函数，这样神经网络就可以应用到众多的非线性模型中。



激活函数	公式	缺点	优点
Sigmoid	$\sigma(x) = 1/(1 + e^{-x})$ 	1、会有梯度弥散 2、不是关于原点对称 3、计算exp比较耗时	-
Tanh	$\tanh(x) = 2\sigma(2x) - 1$ 	梯度弥散没解决	1、解决了原点对称问题 2、比sigmoid更快
ReLU	$f(x) = \max(0, x)$ 	梯度弥散没完全解决，在(-)部分相当于神经元死亡而且不会复活	1、解决了部分梯度弥散问题 2、收敛速度更快



2. 简述Batch Normalization 批标准化的作用。

BN就是通过一定的规范化手段，对于每个隐层神经元，把逐渐向非线性函数映射后取值区间极限饱和区靠拢的输入分布，强制拉回到均值为0方差为1的比较**标准的正态分布**，使得非线性变换函数的输入值落入对输入比较敏感的区域，以此**避免梯度消失**问题。梯度变大意味着学习收敛速度快，能大大加快训练速度。



3. 图示简述残差神经网络ResNet的残差块，以及ResNet-18的神经网络结构。 简述略

