

深度学习参考习题()

1. 什么情况下神经网络模型被称为深度学习模型? (A)
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
2. 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
3. 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
4. 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)
A. $(AB)C$ B. $AC(B)$ C. $A(BC)$ D. 所有效率相同
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 (C)
A. 95 B. 96 C. 97 D. 98
解答 输出大小为 $\lfloor (n - k + 2p + s)/s \rfloor$, 故 $\lfloor (200 - 5 + 2 \times 1 + 2)/2 \rfloor = 99$,
 $\lfloor (99 - 3 + 2 \times 0 + 1)/1 \rfloor = 97$, $\lfloor (97 - 3 + 2 \times 1 + 1)/1 \rfloor = 97$ □
6. 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? (E)
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
7. 如果我们用了一个过大的学习速率会发生什么? (D)
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
8. 在一个神经网络中, 下面哪种方法可以用来处理过拟合? (D)
A. Dropout B. 分批归一化 (Batch Normalization)
C. 正则化 (regularization) D. 都可以
9. 批规范化 (Batch Normalization) 的好处 (A)
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

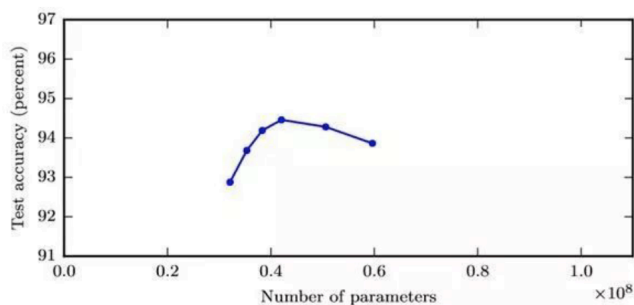
10. 下列哪个神经网络结构会发生权重共享? (D)

- 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

11. 下列哪个函数不可以做激活函数? (线性函数) (D)

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

12. 下图显示了训练过的 3 层卷积神经网络准确度, 与参数数量 (特征核的数量) 的关系。从图中



趋势可见, 如果增加神经网络的宽度, 精确度会增加到一个特定阈值后, 便开始降低。造成这一现象的可能原因是什么? (C)

- A. 即使增加卷积核的数量, 只有少部分的核会被用作预测
B. 当卷积核数量增加时, 神经网络的预测能力会降低
C. 当卷积核数量增加时, 导致过拟合
D. 以上都不正确

13. 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? (D)

- A. 穷举搜索 B. 随机搜索 C. Bayesian 优化 D. 都可以

14. 构建一个神经网络, 将前一层的输出和它自身作为输入。下列哪一种架构有反馈连接?

(A)

- 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

15. 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)

- 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

16. 在训练神经网络时, 损失函数 (loss) 在最初的几个 epochs 时没有下降, 可能的原因是?

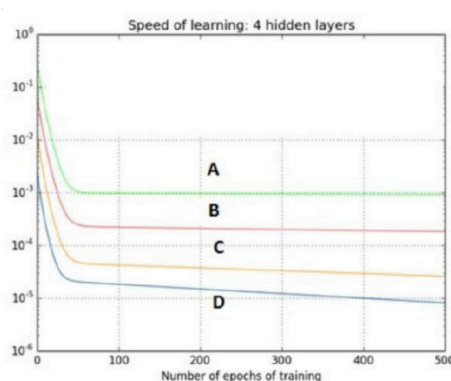
(D)

- 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

17. 下列哪一项属于特征学习算法 (representation learning algorithm)? (C)

- A. K 近邻算法 B. 随机森林 C. 神经网络 D. 都不属于

18. 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。（ B ）
- 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
19. 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 ）
- A. $218 \times 218 \times 5$ B. $217 \times 217 \times 8$ C. $217 \times 217 \times 3$ D. $220 \times 220 \times 5$
20. 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？（ D ）
- A. 可以 B. 不好说 C. 不一定 D. 不能
21. 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 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（ D ）
- A. 1 4 5 B. 1 2 3 C. 1 3 4 5 D. 都有用
22. 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 ）



- 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

23. 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？（ C ）

- 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

24. 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

(C)

- A. 1, 2, 4, 5
- B. 2, 3, 4, 5
- C. 都需要考虑
- D. 1, 3, 4, 5

解答 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。

□

25. 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 ）

- A. 随机梯度下降法
- B. 不知道
- C. 整批梯度下降法
- D. 都不是

26. 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？（ B ）

- 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

解答 HK 算法思想很朴实，就是在最小均方误差准则下求得权矢量。他相对于感知器算法的优点在于，他适用于线性可分和非线性可分的情况，对于线性可分的情况，给出最优权矢量，对于非线性可分得情况，能够判别出来，以退出迭代过程。

□

27. 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？（ B ）

- 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（梯度下降）
- C. 随机赋值，听天由命
- D. 以上都不正确的

28. 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？（ B ）

- 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

29. 训练神经网络过程中，损失函数在一些时期（Epoch）不再减小，原因可能是：
1. 学习率太低 2. 正则参数太大 3. 卡在了局部最小值 (D)
A. 1, 2 B. 2, 3 C. 1, 3 D. 都是
30. 我们不是想要绝对零误差，而是设置一个称为贝叶斯（bayes）误差（我们希望实现的误差）的度量。使用贝叶斯（bayes）误差的原因是什么？ (D)
A. 输入变量可能不包含有关输出变量的完整信息
B. 系统（创建输入-输出映射）可以是随机的
C. 有限的训练数据
D. 所有