

北京交通大学 2020-2021 学年暑假学期

## 硕士研究生《深度学习》期末试题 (A 卷)

出题教师:《深度学习》课程组

班级:\_\_\_\_\_ 姓名:\_\_\_\_\_ 学号:\_\_\_\_\_ 上课教师:\_\_\_\_\_

注意:

- 1、本试卷全部为单项选择题, 共 50 题, 每题 2 分, 共 100 分;
- 2、请将每道题的最佳答案填写在以下的表格中, 否则无效。

题号	答案	题号	答案	题号	答案	题号	答案	题号	答案
1		11		21		31		41	
2		12		22		32		42	
3		13		23		33		43	
4		14		24		34		44	
5		15		25		35		45	
6		16		26		36		46	
7		17		27		37		47	
8		18		28		38		48	
9		19		29		39		49	
10		20		30		40		50	

1. “人工智能”概念诞生的标志性事件是 ( C )

- (A) 1943 年第一个人工神经元模型 M-P 模型的提出
- (B) 1950 年阿兰·图灵发表《机器能思考吗? 》, 提出图灵测试
- (C) 1956 年达特茅斯会议召开
- (D) 1958 年罗森布拉特发明感知机



2. 以下哪个符合对当前人工智能水平的描述？（ C ）

- (A) 能存会算 (B) 能理解会思考 (C) 能感知会识别 (D) 以上都符合

3. 以下哪种不是符号主义的代表性方法？（ B ）

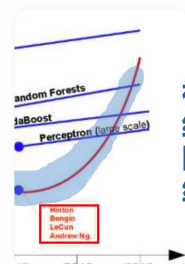
- (A) 专家系统 (B) 朴素贝叶斯 (C) 知识图谱 (D) 定理证明机

4. 深度学习属于以下哪个范畴？（ A ）

- (A) 联结主义 (B) 符号主义 (C) 贝叶斯学派 (D) 逻辑推理

5. 以下哪个不是神经网络第三次兴起的标志性事件？（ B ）

- (A) 逐层预训练方法的提出 (B) 反向传播被用于神经网络训练  
(C) ImageNet 准确率大幅提高 (D) AlphaGo 战胜李世石



6. 以下哪位学者对卷积神经网络的早期大规模实际应用做出了主要贡献？（ D ）

- (A) David Rumelhart (B) Geoffrey Hinton (C) ~~Feifei Li~~ (D) Yann LeCun

7. 深度学习快速崛起的三要素不包括以下哪个？（ C ）

- (A) 大数据 (B) 计算能力提高 (C) 资本关注 (D) 算法进步

8. 有关神经网络训练时使用的学习率参数说法错误的是？（ A ）

- (A) 学习率过大更容易导致训练陷入局部极小值  
(B) 随着训练误差动态调整学习率效果可能更好  
(C) 网络训练时刚开始学习率可以大一些，以便提高学习速度，随后应减少学习率  
(D) 学习率可以与其他网络参数一起训练，对降低代价函数是有利的

9. 下面哪个选项不是神经网络训练过程中防止过拟合的方法？（ D ）

- (A) Dropout (B) 提前终止 (C) L2 正则化 (D) 增加学习率

10. 对于下面一段 python 程序，以下哪个说法是正确的？（ C ）

```
import numpy as np
import scipy.stats
p=np.asarray([0.65,0.25,0.07,0.03])
q=np.array([0.6,0.25,0.1,0.05])
M=(p+q)/2
```

result=0.5\*scipy.stats.entropy(p, M)+0.5\*scipy.stats.entropy(q, M)

- (A) 最后一行是计算 p 和 q 之间的 KL 散度
- (B) 最后一行是计算 p 和 q 之间的条件熵
- (C) 最后一行是计算 p 和 q 之间的 JS 散度
- (D) 最后一行是计算 p 和 q 之间的交叉熵

11. 感知机通过使用激励函数 (activation function) 处理解释变量和模型参数的线性组合对样本分类。Rosenblatt 最初的感知机所使用的激励函数是下列哪一个? ( A )

- (A) 阶跃函数
- ~~(B) Tanh 函数~~
- ~~(C) Sigmoid 函数~~
- ~~(D) ReLU 函数~~

12. 以下关于梯度下降法的描述哪个是错误的? ( B )

- (A) 批量梯度下降法是最原始的形式, 每一次迭代时使用所有样本来进行梯度的更新
- ~~(B) 随机梯度下降法不同于批量梯度下降, 每次迭代时使用几个样本来对参数进行更新~~
- (C) 小批量梯度下降法是批量梯度下降法和随机梯度下降法的一个折中方案
- (D) 随机梯度下降也叫增量梯度下降

12. 以下关于回归分析的说法哪个是正确的? ( C )

- (A) 回归分析的自变量和因变量必须都是 ~~数值型变量~~
- ~~(B) 回归分析是一种拟合因变量和自变量之间关系的无监督学习方法~~
- (C) 可以用平方误差损失或交叉熵损失作为回归分析的学习准则
- ~~(D) 线性回归方程可以直接处理分类问题~~

13. 以下关于 Logistic 回归的说法哪个是错误的? ( A )

- ~~(A) Logistic 回归通常用于处理多分类问题~~
- (B) Logistic 回归的损失函数可以选用交叉熵, 然后采用梯度下降法调整其参数
- (C) Logistic 回归属于线性模型
- (D) Logistic 回归实际上是一种分类算法。

14. 以下关于机器学习模型描述哪个是正确的 ( B )

- ~~(A) 0-1 损失函数能够客观的评价模型的好坏且数学性质很好~~
- (B) 通过引入正则化来限制模型的能力, 从而实现结构风险最小化
- ~~(C) 正则化项可以通过增加参数空间来避免过拟合~~
- ~~(D) 随着模型复杂度的增加, 模型的拟合能力变强, 方差减小而偏差增大~~

15. 以下关于人工神经网络的描述哪个是错误的？（ A ）

- (A) 通过包含多个神经元的~~单层感知机~~就可以解决非线性问题
- (B) 节点、连边和连接方式是人工神经网络的三大要素
- (C) 人工神经网络的隐藏层具有特征的非线性变换功能
- (D) 神经元的偏置（bias）实际上起到的是一个阈值的作用

16. 以下关于深度学习的说法哪个是错误的？（ D ）

- (A) 根据万能近似定理，具有一个隐藏层的神经网络就可以逼近任意一个定义在实数空间中的有界闭集函数
- (B) 单隐层神经网络在最坏的情况下需要指数级的隐藏单元才能逼近某个函数
- (C) 更深层的网络具有更好的泛化能力
- (D) 随着深度的增加，神经网络的表示~~能力呈线性增长~~

17. 以下关于前馈神经网络的性质描述那个是错误的？（ C ）

- (A) 前馈网络中各个神经元按照接收信息的先后分成不同的组，每一组可看作一层
- (B) 前馈网络每一层中的神经元接收来自前一层神经元的输出，并输出给下一层神经元
- (C) 前馈网络中信息传播可以用一个~~无向~~无环图来表示
- (D) 前馈网络中的神经元不具有记忆功能

18. 以下哪种结构不属于记忆网络？（ A ）

- (A) 图卷积网络
- (B) Hopfield 网络
- (C) 受限玻尔兹曼机
- (D) 门控循环单元网络

19. 以下哪个不是激活函数的性质要求？（ B ）

- (A) 连续并可导（允许少数点上不可导）
- (B) 非线性且值域在~~-1 和 +1 之间~~
- (C) 函数及其导函数要尽可能简单
- (D) 导函数的值域要在一个合适的区间内

20. 以下关于前馈神经网络输出单元的描述哪个是错误的？（ D ）

- (A) 线性输出单元经常用于产生条件高斯分布的均值，适合连续值预测问题
- (B) 线性输出单元可以采用均方误差作为损失函数
- (C) Sigmoid 输出单元和 Softmax 输出单元均可采用交叉熵损失函数
- (D) ~~Sigmoid 输出单元和~~ Softmax 输出单元均适合于多分类问题

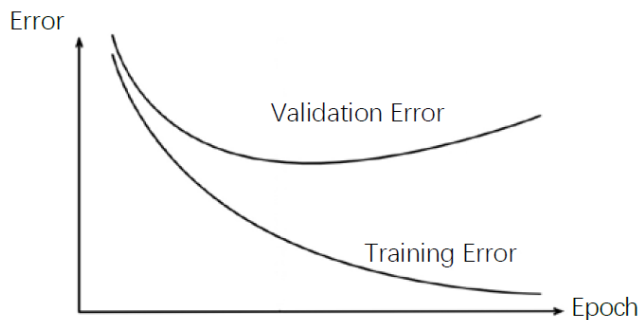
21. 计算机实现神经网络的反向传播算法常采用以下哪种微分方式？（ B ）

- (A) 手动微分 (B) 自动微分 (C) 数值微分 (D) 符号微分

22. 以下关于计算图的说法哪个是错误的？（ C ）

- (A) 在编译时构建计算图，运行不能改变，称之为静态计算图  
(B) 动态计算图的优点之一是灵活性较高  
(C) 静态计算图的缺点之一是~~并行能力差~~  
(D) PyTorch 和 TensorFlow 2.0 都支持动态计算图

23. 某分类网络在训练集和验证集上的误差变化如下图所示，则以下哪个尝试是不能改善此网络的分类效果的？（ B ）



- (A) Dropout (B) ~~提高模型复杂度~~ (C) 早停 (D) 正则化

24. 在神经网络中，优化超参数来最小化代价函数，可以使用下列哪项技术？（ D ）

- (A) 网格搜索 (B) 随机搜索 (C) Bayesian 优化 (D) 以上都可以

25. 神经网络中 Dropout 实现了与下列哪种操作类似的效果？（ A ）

- (A) 集成学习 (B) 对抗学习 (C) 强化学习 (D) 以上都不正确

26. 当选取的小批量样本数量较小时，可以采用以下哪种逐层归一化方式？（ B ）

- (A) 批量归一化 (B) 层归一化 (C) 两者都可以 (D) 两者都不可以

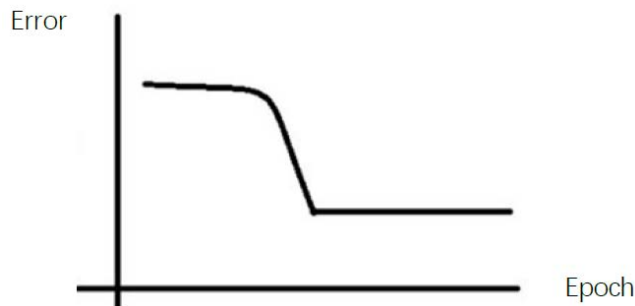
27. 以下关于学习率的描述，哪个是错误的？（ C ）

- (A) 学习率过大，不收敛  
(B) 学习率过小，收敛太慢  
(C) 学习率在一开始要~~小些~~以避免振荡，收敛到最优点附近时要大些以保证快速收敛  
(D) 为了逃离局部最小值或鞍点，可在训练过程中周期性地增大学习率

28. 以下哪个正确地解释了梯度下降法中动量项的作用？（ D ）

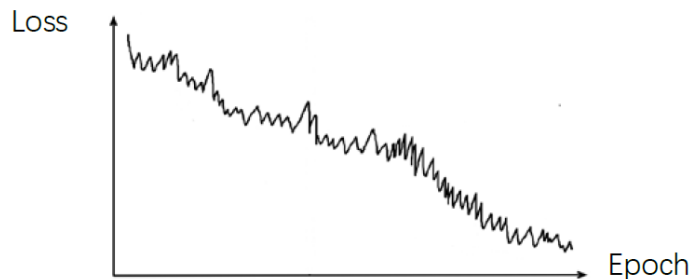
- (A) 在迭代初期，梯度方向比较一致，动量法会起到加速作用，可以更快地到达最优点
- (B) 在迭代后期，梯度方向会不一致，动量法利用之前迭代时的梯度值，减小震荡
- (C) 动量法虽然不能保证收敛到全局最优，但有一定可能跳出局部极值点
- (D) 以上都正确

29. 某神经网络的训练过程如下图所示，开始时有一个停滞期，这是因为神经网络在进入全局最小值之前陷入局部最小值，为了避免这种情况，下面哪个策略是有效的？（ D ）



- (A) 多次初始化模型，选取最优结果
- (B) 采用动量法或 Adam 算法利用之前迭代时的梯度值，努力跳出局部极小值点
- (C) 周期性地增大学习率
- (D) 以上都可以

30. 如果一个网络在训练中损失函数随着迭代次数的增加呈现下图所示的变化曲线，则以下哪个说法是正确的？（ B ）



- (A) 如果使用的是 mini-batch 梯度下降，那可能有问题；如果使用的批量梯度下降，那应该是合理的
- (B) 如果使用的是 mini-batch 梯度下降，那应该是合理的；如果使用的是批量梯度下降，那可能有问题

- (C) 无论使用的是 mini-batch 梯度下降还是批量梯度下降，都应该是合理的  
 (D) 无论使用的是 mini-batch 梯度下降还是批量梯度下降，都可能有问题

31. 给定一个信号序列[-1, 0, 2, 0, 2, 2, -1, 1, 0]，使用一维卷积核[1, 0, 2]从左往右对其进行卷积，填充为 0，步长为 1，则卷积后的序列为 ( D )

- (A) [3, 0, 6, 4, 0, 4, -1]      (B) [3, 0, 0, 2, -3, 1, 1]  
 (C) [0, 0, 1, 3, 5, 4, -2]      (D) [0, 0, 6, 2, 3, 5, -2]

32. 现有一张灰度图像，大小为  $227 \times 227$ ，经过步长为 2、填充为 1、卷积核大小为  $3 \times 3$  的卷积操作之后，其输出维度是多少？ ( B )

- (A)  $113 \times 113$       (B)  $114 \times 114$       (C)  $115 \times 115$       (D)  $116 \times 116$

$$(227+2-3) / 2 + 1$$

33. 给定一个二维矩阵：

1	1	1	1	1
-1	0	-3	0	1
2	1	1	-1	0
0	-1	1	2	1
1	2	1	1	1

$$1 + 1 = 2$$

使用如下卷积核对其进行步长为 1，填充为 0 的卷积操作：

-1	0	0
0	0	0
0	0	<u>1</u>

则卷积之后的输出结果为 ( B )

0	<del>-2</del>	-1
2	2	4
-1	0	0

(A)

0	<u>2</u>	1
-2	-2	-4
1	0	0

(B)

<del>2</del>	0	1
0	2	-2
3	2	2

(C)

<del>-2</del>	0	-1
0	-2	2
-3	-2	-2

(D)

34. 下面哪个不属于卷积网络出现的动机？ ( C )

- (A) 稀疏交互      (B) 参数共享      (C) ~~尺寸不变性~~      (D) 平移不变性

35. 下面哪个不属于池化层的作用？（ B ）

- (A) 缓解模型过拟合 (B) ~~改变输出通道数~~ (C) 降低特征维度 (D) 减少参数数量

36. 下列关于池化操作的描述，哪个是错误的？（ D ）

- (A) 常用的池化操作包括最大池化和平均池化  
(B) 池化操作可以更好地获取平移不变性  
(C) 池化操作可以提高网络的计算效率  
(D) 在计算机视觉领域，~~平均池~~  
最大化对于纹理的提取效果较好

37. 给定一张  $1000 \times 1000$  像素的图像，分别采用一个具有 1 百万个神经元的全连接层和一个具有 100 个  $10 \times 10$  卷积核的卷积层对其进行特征转换，请问二者的参数量大概相差多少倍？（ A ）

- (A)  $10^8$  (B)  $10^9$  (C)  $10^{10}$  (D)  $10^{11}$

38. 某卷积层输入的图片大小为  $32 \times 32 \times 3$ ，该层共有 10 个卷积核，大小为  $5 \times 5$ ，请问该卷积层共有多少个需要学习的参数？（ D ）  $5 \times 5 \times 3 + 1 \times 10$

- (A) 250 (B) 260 (C) 750 (D) 760

39. 下列关于卷积网络说法正确的是（ A ）

- (A) 训练过程中，大部分池化层不需要进行参数更新  
(B) 大部分卷积网络需要学习的参数包括卷积层的参数和~~池化层参数~~  
(C) 卷积网络的卷积核是否进行翻转和其特征抽取的能力有关  
(D) 训练过程中，训练参数（超参）设置不合理会导致过拟合，但不会导致欠拟合

40. 下列关于卷积操作说法正确的是（ B ）

- (A) 通过卷积操作只能实现高维特征到低维特征的转换 随着深度可以提取更高维度的特征  
(B) 增加卷积层数和卷积核大小都可以增加输出单元的感受野  
(C) 空洞卷积通过给卷积核插入空洞来变相地增加其大小，所以会增加参数数量  
(D) 在卷积之前进行池化操作和空洞卷积都可以增加输出单元的感受野并且不会损失信息

41. 下列关于卷积神经网络说法错误的是（ D ）

- (A) 用多层小卷积核来代替大卷积核，可以减少计算量和参数量

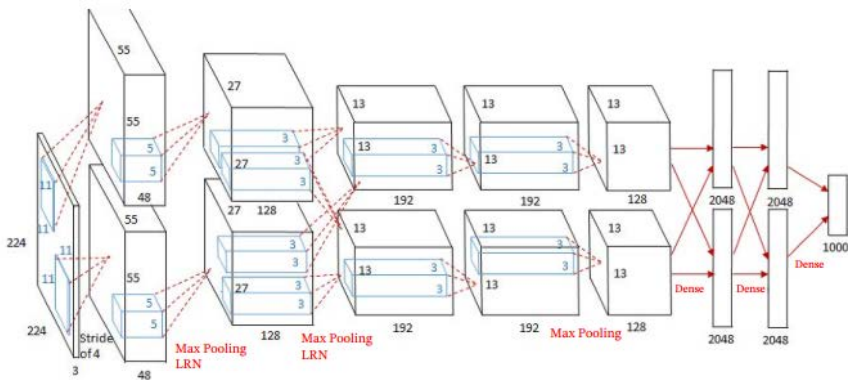


- (B) Inception module 的优点包括减少网络参数，降低运算量和实现多尺度多层次滤波
- (C) 残差网络可以解决由于网络加深而引起的梯度消失和爆炸以及网络退化等问题，并且没有带来额外的参数和计算开销
- (D) 深度残差网络能够缓解网络退化等问题，~~但是模型精度会因此而降低~~

42. 以下关于卷积核的描述，哪个是错误的？（ C ）

- (A) 卷积核和滤波器是一个概念
- (B) 在图像处理领域，卷积核代表图像的某种特征
- (C) 卷积核越大，感受野越大，特征提取能力越强 并不是，平衡最好
- (D) 网络层数越深，卷积核提取的特征越抽象

43. 以下结构是哪个经典的卷积神经网络？（ B ）



- (A) LeNet
- (B) AlexNet
- (C) VGG16
- (D) GoogLeNet

44. 以下循环神经网络结构与实际任务对应错误的是（ D ）

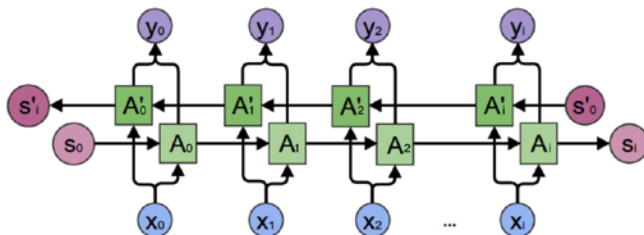
- (A) One-to-Many 结构：看图作文
- (B) Many-to-One 结构：机器翻译
- (C) Many-to-Many（同步）结构：实体识别
- (D) Many-to-Many（异步）结构：文档摘要

45. 以下关于循环神经网络训练方法的描述，哪个是正确的？（ D ）

- (A) 循环神经网络的训练需要基于时间反向传播误差
- (B) BPTT 算法的核心思想和 BP 算法是相同的
- (C) BPTT 算法与 BP 算法不同之处在于参数的寻优过程需要追溯历史数据

(D) 以上都正确

46. 对于如下图所示的双向循环神经网络，那个说法是错误的？（ C ）



(A) 双向循环神经网络的两层循环网络的输入是相同的

(B) 双向循环神经网络的两层循环网络的信息传递方向不同

☒ (C) 双向循环神经网络的正向计算和反向计算共享权重

(D) 以上都正确

47. 循环神经网络不能向前看序列中任意多个点的原因是（ C ）

(A) 网络深度问题 (B) 短程依赖问题 (C) 长程依赖问题 (D) 计算能力问题

48. 以下关于梯度消失和梯度爆炸的描述，哪个是错误的？（ D ）

(A) 梯度消失或者梯度爆炸会导致梯度为 0 或 NaN，无法继续训练更新参数

(B) 通常来说，梯度爆炸比梯度消失更容易处理一些

(C) 合理的参数初始化和使用 ReLU 函数可以用来解决梯度消失问题

(D) 通过优化网络结构无法解决梯度消失问题

49. LSTM 的门控单元不包括下列哪个门？（ A ） GRU的包含更新门和重置门

(A) 更新门 (B) 遗忘门 (C) 输入门 (D) 输出门

50. 以下关于 GRU 和 LSTM 的说法，哪个是错误的？（ D ）

(A) GRU 是 LSTM 的一种变体，但是比 LSTM 结构更简单

(B) 更新门用于控制前一时刻的状态信息被带入到当前状态中的程度，更新门的值越大说明前一时刻的状态信息带入越多

(C) 重置门控制前一状态有多少信息被写入到当前的候选集上，重置门越小，前一状态的信息被写入的越少

(D) ~~GRU 比 LSTM 需要更多的数据来进行训练~~