

# 北京交通大学 2022-2023学年 秋季学期

## 本科生《深度学习》期末试题

出题教师：《深度学习》课程组

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

(注意：本卷共包含50道单项选择题，每题2分，共100分，请将每题的最佳答案直接填入题后括号中。)

1. “人工智能”概念诞生的标志性事件是 ( )  
(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. 以下关于机器学习模型的正确描述哪个是正确的 ( )  
(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
3. 以下关于Logistic回归的说法哪个是错误的? ( )  
(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回归属于广义线性模型  
(D) Logistic回归实际上是一种分类算法
4. 以下哪个符合对当前人工智能水平的描述? ( )  
(A) 能存会算 ( ) 能理解会思考 ( ) 能感知会识别 (D) 以上都符合
5. 深度学习快速崛起的三要素不包括以下哪个? ( )  
(A) 大数据 (B) 计算能力提高 (C) 资本关注 (D) 算法进步
6. 下列哪种函数可以用来解决多分类问题? ( )  
A softmax B sigmoid C ReLU D tanh
7. 单层感知机不能解决以下哪种问题? ( )  
A 与问题 B 或问题 C 非问题 D 异或问题
8. 以下哪个不是激活函数的性质要求? ( )  
(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
9. ReLU是一种常见的激活函数，以下对于它的描述不正确的是 ( )  
(A) ReLU函数具有单侧抑制、宽兴奋边界的生物学合理性。  
(B) ReLU函数有可能导致神经元的死亡 (即，永远不会被更新)。  
(C) ReLU 函数具有稀疏激活的特质。  
(D) ReLU函数的输出值域为[0, 1]。
10. 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

11. 以下哪种函数的值域在0到1之间? ( )

A sigmoid B tanh C ReLU D Leaky ReLU

$$\frac{1}{N} \sum_{i=1}^N l(f(x^i, \theta), y^i) + \frac{1}{2} \lambda \|\theta\|_2^2$$

12. 目标函数为  $\frac{1}{N} \sum_{i=1}^N l(f(x^i, \theta), y^i) + \frac{1}{2} \lambda \|\theta\|_2^2$ , 则模型最小化的是什么? ( )

(A) 平均风险 (B) 经验风险 (C) 期望风险 (D) 结构风险

13. 提前停止是防止过拟合的方法之一, 如果在什么集上的错误率不再下降, 就停止迭代? ( )

(A) 训练集 (B) 测试集 (C) 验证集 (D) 以上都可以

14. RMSprop算法中, 为什么将Gt的计算由累积方式变成了指数衰减移动平均? ( )

(A) 避免AdaGrad算法中学习率不断单调下降以至于过早衰减的缺点  
(B) 有助于帮助逃离局部极值点或鞍点  
(C) 优化过程中损失会呈现震荡的方式下降, 有助于缓解震荡  
(D) 以上都对

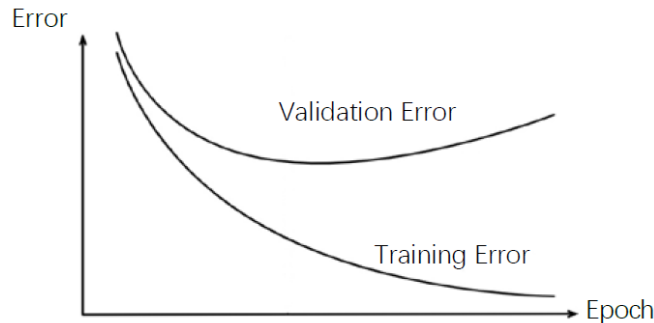
15. 下面哪个选项不是神经网络训练过程中防止过拟合的方法? ( )

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

16. 以下关于梯度下降法的描述哪个是错误的? ( )

(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

17. 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? ( )

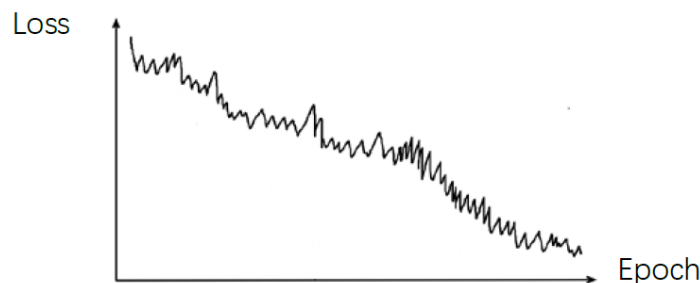


(A) Dropout (B) 提高模型复杂度 (C) 早停 (D) 正则化

18. 以下关于学习率的描述, 哪个是错误的? ( )

(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

19. 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? ( )



(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

20. 全连接前馈网络的BP算法中，第L层的误差项等于？（ ）

- (A) 第L+1层的误差项的加权和，再乘以该层激活函数的梯度  
(B) 第L+1层的误差项的加权和，再乘以该层激活函数  
(C) 第L-1层的误差项的加权和，再乘以该层激活函数的梯度  
(D) 第L-1层的误差项的加权和，再乘以该层激活函数

21. 当数据过大以至于无法在RAM中同时处理时，可以采取以下哪种方法？（ ）

- (A) 随机梯度下降法  
(B) 批量梯度下降法  
(C) 增大学习率  
(D) 减小学习率

22. 在感知机中（Perceptron）的任务顺序是什么？（ ）

1. 初始化随机权重
  2. 去到数据集的下一批（batch）
  3. 如果预测值和输出不一致，改变权重
  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

23. 在写测试函数时，通常会在模型之后加上model.eval()函数，其作用是：（ ）

- A. 使梯度不再计算与更新，从而保持网络参数固定  
B. 初始化网络参数  
C. 不启用BN和Dropout，从而保持网络参数固定  
D. 与训练函数相区别

24. 过拟合是指模型在训练集上表现很好，到了验证和测试阶段就很差，即模型的泛化能力很差，通常可以使用的解决方法不包括：（ ）

- A. 早停机制 B. 增加正则项 C. Dropout D. 增加模型复杂度

25. 下面有关基于方差缩放的参数初始化的说法中正确的是（ ）：

- ① 是为了缓解梯度消失或爆炸的问题 ② 尽可能保证每个神经元的输入和输出一致 ③ 根据神经元的连接数量来自适应调整初始化问题 ④ 在一个区间 $[-r, r]$ 内采用均匀分布来初始化参数  
A. ①②③ B. ①④ C. ②③④ D. ①③

26. 下面哪种优化算法每次只是单个样本进行模型优化？（ ）

- A 批量梯度下降法 B 随机梯度下降法 C 小批量随机梯度下降法

27. 批量规范化（batch normalization）有用的原因是（ ）

- (A) 它将返回归一化的权重平均值和标准差  
(B) 在将所有的输入传递到下一层之前对其进行归一化（更改）

- (C) 它是一种非常有效的反向传播技术  
(D) 这些均不是

28. 对于分类任务，我们不是将神经网络中的随机权重初始化，而是将所有权重设为零。下列哪项是正确的？（ ）

- (A) 没有任何问题，神经网络模型将正常训练  
(B) 神经网络模型可以训练，但所有的神经元最终将识别同样的事情  
(C) 神经网络模型不会进行训练，因为没有净梯度变化  
(D) 这些均不会发生

29. 以下关于感知机收敛性叙述不正确的是？（ ）

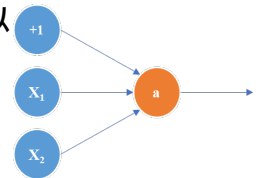
- (A) 在数据集线性可分时，感知器虽然可以找到一个超平面把两类数据分开，但并不能保证其泛化能力。  
(B) 感知器对样本顺序比较敏感，每次迭代的顺序不一致时，找到的分割超平面也往往不一致。  
(C) 在任何情况下，参数更新都能在次迭代内收敛。  
(D) 如果训练集不是线性可分的，就永远不会收敛。

30. 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) 全连接前馈网络  
(D) 卷积神经网络和循环神经网络

31. 假设在一个神经元上实现与函数（AND），其激活函数为如下，则以下哪组参数可以输出正确的结果？（ ）

- (A) Bias = -1.5,  $w_1 = 1$ ,  $w_2 = 1$   
(B) Bias = 1.5,  $w_1 = 2$ ,  $w_2 = 2$   
(C) Bias = 1,  $w_1 = 1.5$ ,  $w_2 = 1.5$   
(D) 以上都不正确



32. 关于空洞卷积的优点，正确的是（ ）

- (A) 在增加参数的情况下增大感受野，适用于图片size较大，或需要快速感受全局信息的情况。  
(B) 通过设置不同的孔洞率捕获多尺度上下文信息，适用于需要捕获图片的多层次的情况，如语义分割等。  
(C) 空洞卷积存在网格效应。  
(D) 与具有相同参数数量的卷积相比，具有更小的感受野。

33. 在使用一个卷积神经网络层处理图像时，为了保证输出图像的大小即长、宽与输入图像相等，且已知步长 $T=1$ ，卷积核大小为 $K=3$ ，则图像中每行像素的两端应该分别填充（ ）个0值。

- (A) 0 (B) 1 (C) 2 (D) 3

34. 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（此处为传统卷积操作，非互相关操作），填充为0，步长为1，则卷积后的序列为（ ）

- (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)  $[0, 0, 6, 2, 3, 5, -2]$

35. 下面哪个不属于卷积网络出现的动机？（ ）

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

36. 下面哪个不属于池化层的作用？（ ）

- (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

37. 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？（ ）

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

38. 下列关于卷积神经网络说法错误的是（ ）

- (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

39. 在不增加参数量的前提下增加输出单元的感受野，不能采取以下哪种方式：（ ）

- (A) 在卷积之前进行池化操作
- (B) 增加卷积核的大小
- (C) 空洞卷积
- (D) 膨胀卷积

40. 在卷积神经网络中使用 $1 \times 1$ 卷积时，下列哪一项是正确的？（ ）

- A.  $1 \times 1$ 卷积核可以帮助降低维度
- B.  $1 \times 1$ 卷积核可以用于特征池表示
- C.  $1 \times 1$ 卷积核可以降低过拟合的风险
- D. 上述都正确

41. 若要考虑上下文完成语句填空“我的手机坏了，我打算\_\_\_\_一部新手机。”选用哪个网络更好？（ ）

- (A) RNN (B) LSTM (C) 深层RNN (D) 双向RNN

42. RNN网络可以采用以下哪种逐层归一化方式？（ ）

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

43. 以下循环神经网络结构与实际任务对应错误的是（ ）

- (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

44. LSTM的门控单元不包括下列哪个门？（ ）

- (A) 更新门 (B) 遗忘门 (C) 输入门 (D) 输出门

45. 给定一条长度为N的长序列，使用长度为M的滑动窗口进行采样（ $M < N$ ），可以得到多少条子序列？（ ）

- (A)  $N-M$
- (B)  $N+M$
- (C)  $N-M+1$
- (D)  $N-M-1$

46. 使用PyTorch实现循环神经网络时，一定不会用到下面哪个模块？（ ）

- (A) nn.Parameter (B) nn.Tanh (C) nn.Conv1d (D) nn.Linear

47. 有关LSTM中激活函数选择，说法正确的是：（ ）

- (A) sigmoid 用在了各种gate上，产生0~1之间的值，tanh 用在了状态和输出上，是对数据的处理。
- (B) tanh 用在了各种gate上，产生0~1之间的值，sigmoid 用在了状态和输出上，是对数据的处理。
- (C) relu 用在了各种gate上，产生0~1之间的值，tanh 用在了状态和输出上，是对数据的处理。

(D) 以上都不正确

48. 使用nn.linear()实现一个线性分类模型，代码如下：

```
class LinearNet(nn.Module):
    def __init__(self, n_feature):
        super(LinearNet, self).__init__()
        self.linear = nn.Linear(n_feature, 1)
    def forward(self, x):
        y = self.linear(x)
        return y
```

则选用下列哪一个loss函数最合适？（ ）

- (A) nn.L1Loss() (B) nn.BCEWithLogitsLoss() (C) nn.L2Loss() (D) nn.L1Loss()或nn.L2Loss()

49. 在PyTorch中使用SGD优化方法进行实验时，下列5行代码顺序正确的是（ ）

```
opt = torch.optim.SGD(model.parameters(), lr)
for input, target in dataset:
    ( )
```

- ①opt.step() ②opt.zero\_grad() ③output = model(input) ④loss = loss\_fn(output, target) ⑤loss.backward()  
(A) ③⑤④①② (B) ③④⑤②① (C) ②③④⑤① (D) ①③④⑤②
50. RNN返回的输出状态为三维张量out=[batch\_size, seq\_len, hidden\_size], 若想取出最后一步的输出状态, 下列语句正确的是 ( )  
(A) out[-1, :, :] (B) out[:, -1, :] (C) out[:, :, -1] (D) out