1. 填空题
2. 强化学习是机器人导航、避障领域常用的算法。
3. 深度学习等同于采用神经网络的方法寻找映射函数。
4. 深度学习中损失函数用于估量模型的预测值f(x)与真实值y的不一致程度。
5. 多层感知机模型可以分为输入层、隐藏层和输出层。（第3章）
6. 在深度学习中，减少数据特征、使用更简单的模型结构和引入更多的训练技巧，都是缓解过拟合现象的常见方法之一。
7. 机器学习中的正则化方法有两种主要类型，分别是L1正则化和L2正则化。
8. 使用PyTorch创建一个形状为6×5的全零张量y的代码表达式为y=torch.zeros(6,5)。
9. 对于一张200×200×3的图片，经过一次平均池化操作（池化核大小为2×2），输出特征图大小为100×100×3。
10. AlexNet网络有5个卷积层。
11. 长短时记忆网络（LSTM）包括输入门、遗忘门和输出门。
12. VGG-19网络有 16 个卷积层。
13. 常见的两种池化方法为最大池化和 平均池化 。
14. 减少训练轮次，增加正则化参数，减少网络复杂度，可解决 过拟合 现象。
15. 递归神经网络（RNN）被广泛应用于处理时间序列数据，其中隐藏层的状态可以捕捉到输入数据的时序信息。（第12章）
16. 多层感知器通过反向传播算法进行训练，不断调整权重以最小化损失函数，从而实现更准确的输出。（第3章）
17. 在卷积神经网络中，用于提取图像局部特征的是 卷积 层。
18. 在循环神经网络（RNN）中，为了避免梯度消失问题，通常会使用 LSTM网络结构。
19. 在深度学习中，优化算法中常用的梯度下降法包括随机梯度下降法、批量梯度下降法和 小批量梯度下降法 。
20. 在深层神经网络中，层数的增加通常会导致模型复杂度的提升，这可能会引起 过拟合 问题。
21. 在激活函数中，ReLU函数由于其单侧抑制的特点，在深度学习中得到了广泛应用。
22. 一份包含50000张图片的训练集，一个epoch的训练次数为80，其设置的batch-size大小为 625。
23. 使用PyTorch创建一个形状为3×4的全零张量y的代码表达式为y= torch.zeros(3,4) 。
24. 8.在卷积神经网络中，通过设置合适的卷积核、 步幅 和填充，可以控制输出特征图的尺寸。
25. 9.强化学习是一种通过与环境互动来学习策略的方法，其中 Q-learning 算法是一种基于价值的、通过维护Q表的强化学习算法。
26. 单选题

1.神经元之间通过什么结构进行信息传递？（ C ）

A. 树突 B. 轴突 C. 突触 D. 髓鞘

2.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. 全批量梯度下降法

C．小批量梯度下降法

D．以上都不正确

3.ANN 是一种模仿生物神经网络，其中的（ B ）扮演了生物神经模型中突触的角色，用于调节一个神经元对另一个神经元的影响程度。

A. 细胞体 B. 权重 C. 输入通道 D. 输出通道

4.下面哪个不是深度学习的常见应用领域？（ D ）（第1章）

A. 图像识别 B. 语音识别 C. 自然语言处理 D. 数据库管理

5.以下关于深度学习和神经网络的描述，正确的是（ C ）。

A.神经网络等同于深度学习

B.神经网络是一种浅层学习

C.目前深度学习以神经网络为主要模型

D.神经网络不可以解决贡献度分配问题

6. 以下哪个语言被广泛用于深度学习开发？（ B ）

A. Java B. Python C. C++ D. JavaScript

7.以下关于深度学习和神经网络的描述，正确的是（ C ）。

A.神经网络等同于深度学习

B.神经网络是一种浅层学习

C.目前深度学习以神经网络为主要模型

D.神经网络不可以解决贡献度分配问题

8.以下关于感知机模型的说法，正确的是（ D ）。

A.单层感知机只由单层神经元组成

B.多层感知机由两层神经元组成，分别为输入层和输出层

C.损失函数是感知机模型优化的目标函数，拟合越差损失函数值越小

D.Sigmoid、tanh、ReLU都是感知机模型中常见的激活函数

9.下列关于多层感知机MLP的说法，不正确的是（ B ）。

A.MLP各层之间的神经元连接方式为全连接

B.增加MLP的隐藏层层数会增强MLP的表示能力，因此测试误差便一定会减小，预测能力也会更强

C.MLP通常使用反向传播算法来进行监督学习

D.MLP的常用激活函数有Sigmoid函数，tanh函数，ReLU函数等

10.监督学习的主要类型是（ A ）。 （第1章）

A. 分类和回归 B. 聚类和回归 C. 分类和降维 D. 聚类和降维

11.什么是前馈神经网络？（ A ）（第6章）

A、一种只有正向传递信息的神经网络

B、一种只有反向传递信息的神经网络

C、一种既有正向传递又有反向传递信息的神经网络

D、一种只有横向传递信息的神经网络

12.ResNet-50 有多少个卷积层（ B ）。

A.48 B.49

C.50 D.51

13.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，但如何获知较好的神经网络权重和偏移（ B ）。（第8章）

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.以上均不正确

14.相比于传统的神经网络激活函数Sigmoid函数和tanh函数，使用ReLU函数作为激活函数的优势不包括的是（ D ）。

A.ReLU函数计算简单，并且能加快网络的训练速度

B.可以有效缓解梯度消失的问题

C.ReLU的单侧抑制提供了网络的稀疏表达能力，模拟了人类大脑神经元的实际活动

D.ReLU函数具有连续，可导等比较好的数学性质

15.下面关于随机梯度下降法（SGD）的叙述，不正确的是（ B ）。

A.使用参数的梯度，沿梯度方向更新参数，并重复这个步骤多次

B.参数更新公式中的学习率需要随机选取

C.SGD形式简单，并且容易实现

D.SGD在收敛过程中波动较大

16.下列哪一项在神经网络中引入了非线性（ B ）。

A.随机梯度下载 B.激活函数

C.卷积函数 D.以上均正确

17.以下关于神经网络的训练过程，不正确的是（ D ）。

A.mini-batch从训练数据中随机选出一部分数据，这部分数据称为mini-batch，我们的目标是减少mini-batch损失函数的值

B.计算梯度为了减小mini-batch的损失函数，需要求出各个权重参数的梯度

C.更新参数梯度反方向表示损失函数的值减小最多的方向，将权重参数沿梯度反方向进行微小更新

D.随机梯度下降stochastic gradient descent：“随机”代表在梯度下降中随机初始一个学习率，并不断尝试多个值，寻求最好的结果

18.构建一个神经网络，将前一层的输出和它自身作为输入，是属于哪一种架构（ A ）。（第12章）

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

11.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.学习率太低 B.正则参数太高

C.陷入局部最小值 D.都有可能

19.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.批归一化

C.正则化 D.都可以

20.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，哪一项是错误的（ D ）。

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.每个神经元必须有多个输入和多个输出

21.假设有5个大小为7x7、步长为1的卷积核，同时卷积神经网络第一层的深度为1。此时如果向这一层传入一个维度为224x224x3的数据，边界不填充，那么神经网络下一层所接收到的数据维度是（ A ）。

A.218x218x5 B.217x217x8

C.217x217x3 D.220x220x5

22.相比于传统的神经网络激活函数Sigmoid函数和tanh函数，使用ReLU函数作为激活函数的优势不包括的是（ D ）。

A.ReLU函数计算简单，并且能加快网络的训练速度

B.可以有效缓解梯度消失的问题

C.ReLU的单侧抑制提供了网络的稀疏表达能力，模拟了人类大脑神经元的实际活动

D.ReLU函数具有连续，可导等比较好的数学性质

23.在一个神经网络里，知道每一个神经元的权重和偏差是最重要的一步。如果以某种方法知道了神经元准确的权重和偏差，就可以近似任何函数。实现这个最佳的办法是（ C ）。

A.随机赋值，祈祷它们是正确的

B.搜索所有权重和偏差的组合，直到得到最优值

C.赋予一个初始值，检查与最优值的差值，然后迭代更新权重

D.以上都不正确

24.下列哪个不属于激活函数？（ A ）

A．math B．sigmoid C．tanh D．relu

25.BP算法每一轮迭代中的更新步长是由（ D ）控制的。

A．梯度 B．目标函数 C．损失函数 D．学习率

26.下列BP算法流程描述中，正确的执行顺序为（ B ）

① 将输入示例提供给输入层神经元

② 逐层将信号前传，直到产生输出层的结果

③ 计算输出层的误差

④ 将误差逆向传播至隐层神经元

⑤ 根据隐层神经元的误差对权值和偏置进行调整

A．②③①④⑤ B．①②③④⑤ C．①②④③⑤ D．①②③⑤④

27.下列有关神经网络的说法中，正确的是（ A ）

A．增加层数可能扩大测试误差 B．减少层数一定缩小测试误差

C．增加层数一定减少训练误差 D．增加层数一定增加测试误差

28.创建一个含有20个元素且符合正态分布N(0,1)的张量（ D ）。

A.a = torch.Tensor(20,1) B.a = torch.tensor(1,20)

C.a = torch.rand(1,20) D.a = torch.randn(1,20)

29.深层神经网络在感知机的模型上做了扩展，加入了一个或多个（ B ）来增强模型的表达能力。

A、 输入层

B、 隐藏层

C、 输出层

D、 对抗网络

30.反向传播算法的主要目的是什么？（ A ）

A、 训练神经网络模型的参数

B、 优化神经网络模型的结构

C、 加速神经网络模型的训练

D、 选择最优的神经网络模型

31.在深度学习中，权重更新公式为？（ A ）

A、 新权重=当前权重-学习率\*梯度

B、 新权重=学习率\*当前权重-学习率\*梯度

C、 新权重=学习率\*梯度

D、 新权重=当前权重-梯度

32.下列关于模型能力（model capacity）的描述正确的是（ D ）。

A.隐藏层层数增加，模型能力一定增加

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.梯度截断可以缓解梯度爆炸

33.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

34.在PyTorch中，哪一项是专门用于神经网络构建的模块（ A ）。

A.nn B.optim

C.Autograd D.Variable

35.下面的优化方法中属于自适应学习率的优化方法的是（ D ）。

A.SGD B.Momentum

C.Nesterov Momentum D.AdaGrad

36.对于发生过拟合的原因，下列叙述正确的是（ B ）。

A.训练数据集量太大 B.训练数据集量太小

C.模型拥有少量参数 D.模型表现力差

37.以下关于CNN中池化的说法，不正确的是（ A ）。

A.经过池化运算，输入数据和输出数据的通道数会发生变化

B.池化对输入数据的微小偏差具有鲁棒性

C.池化只是从目标区域中取最大值或平均值，所以不存在要学习的参数

D.通常池化的窗口大小会和上一步卷积操作的步幅设定成相同的值

38.为解决序列模型长期依赖的问题，经常使用的结构是（ A ）。

A.LSTM B.AutoEncoder

C.VGG D.PixelRNN

39.下列场景中适合采用一对多结构RNN的是（ A ）。

A.生成图片说明 B.情感分析

C.机器翻译 D.基于帧粒度的视频分类

40.为解决序列模型长期依赖的问题，经常使用的结构是（ A ）。

A.LSTM B.AutoEncoder

C.VGG D.PixelRNN

41.以下（ C ）不是深度学习在近几年获得飞速发展的原因。

A.更好的非线性激活函数和新的正则化技术的发展

B.更多稳健的优化算法和深度学习平台的发展

C.感知机模型的提出

D.更多优秀的网络结构和并行运算的发展

42.以下关于深度学习和神经网络的描述，正确的是（ C ）。

A.神经网络等同于深度学习

B.神经网络是一种浅层学习

C.目前深度学习以神经网络为主要模型

D.神经网络不可以解决贡献度分配问题

43.AlphaGo是DeepMind研发的围棋人工智能系统，2016年以4:1击败韩国选手李世乭引起了整个社会对人工智能的极大关注，下列关于AlphaGo说法错误的是（ D ）。

A.AlphaGo的技术涉及强化学习

B.AlphaGo使用蒙特卡罗树搜索提高稳定性

C.AlphaGo训练模型过程中通过价值网络和策略网络选择最优下子方法

D.训练AlphaGo没有用到人类玩家的棋局数据

44.以下关于感知机模型的说法，正确的是（ D ）。

A.单层感知机只由单层神经元组成

B.多层感知机由两层神经元组成，分别为输入层和输出层

C.损失函数是感知机模型优化的目标函数，拟合越差损失函数值越小

D.Sigmoid、tanh、ReLU都是感知机模型中常见的激活函数

45.下列说法中，不属于PyTorch优点的是（ A ）。

A.几乎不需要自行编写训练代码

B.有很多预训练模型

C.很容易编写自己的层类型并在GPU上运行

D.大量模块化组件，容易组合

46.下列叙述正确的是（ A ）。

A.GPU功耗大；擅长并行任务 B.GPU功耗大；擅长串行任务

C.CPU功耗小；擅长并行任务 D.CPU功耗大；擅长串行任务

47.下列关于多层感知机MLP的说法，不正确的是（ B ）。

A.MLP各层之间的神经元连接方式为全连接

B.增加MLP的隐藏层层数会增强MLP的表示能力，因此测试误差便一定会减小，预测能力也会更强

C.MLP通常使用反向传播算法来进行监督学习

D.MLP的常用激活函数有Sigmoid函数，tanh函数，ReLU函数等

48.相比于传统的神经网络激活函数Sigmoid函数和tanh函数，使用ReLU函数作为激活函数的优势不包括的是（ D ）。

A.ReLU函数计算简单，并且能加快网络的训练速度

B.可以有效缓解梯度消失的问题

C.ReLU的单侧抑制提供了网络的稀疏表达能力，模拟了人类大脑神经元的实际活动

D.ReLU函数具有连续，可导等比较好的数学性质

49.在一个神经网络里，知道每一个神经元的权重和偏差是最重要的一步。如果以某种方法知道了神经元准确的权重和偏差，就可以近似任何函数。实现这个最佳的办法是（ C ）。

A.随机赋值，祈祷它们是正确的

B.搜索所有权重和偏差的组合，直到得到最优值

C.赋予一个初始值，检查与最优值的差值，然后迭代更新权重

D.以上都不正确

50.下面关于随机梯度下降法（SGD）的叙述，不正确的是（ B ）。

A.使用参数的梯度，沿梯度方向更新参数，并重复这个步骤多次

B.参数更新公式中的学习率需要随机选取

C.SGD形式简单，并且容易实现

D.SGD在收敛过程中波动较大

12.在强化学习中，策略（Policy）通常用来描述智能体在给定状态下应该采取的行动。以下关于策略的描述（ B ）是正确的。

A. 策略是一个确定性函数，它为每个状态返回一个确定的动作。

B. 策略是一个概率函数，它为每个状态返回所有可能动作的概率分布。

C. 策略仅依赖于先前的奖励，而不是当前的状态。

D. 策略在强化学习中不是必须的，智能体可以直接从状态到动作的映射。

13. 强化学习的主要目标是最大化（ C ）？

A. 状态值 B. 动作值 C. 总奖励 D. 损失函数

14. 在神经网络的反向传播过程中，以下步骤的正确执行顺序是（A ）。

① 计算每一层的输出

② 根据损失函数计算输出层的误差

③ 将误差信号反向传播到各层

④ 更新网络中的权重和偏置

⑤ 提供输入数据并进行前向传播

A. ⑤①②③④ B. ⑤②①③④ C. ①⑤③②④ D. ②⑤①③④

15. 下列关于深度学习优化器的描述，正确的是（ C ）。

A. 使用Adam优化器时，学习率越大，模型越容易收敛

B. 使用SGD优化器时，动量（Momentum）参数越大，模型越容易陷入局部最小值

C. AdaGrad优化器通过为每个参数调整学习率来提高学习效率

D. RMSprop优化器不考虑历史梯度的影响

16.下列关于模型能力（model capacity）的描述正确的是（ D ）。

A.隐藏层层数增加，模型能力一定增加

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.梯度截断可以缓解梯度爆炸

17. 输入图片大小为256×256，依次经过一层卷积（kernel size 3×3，padding 1，stride 1），一层卷积（kernel size 3×3，padding 1，stride 2），再经过一层卷积（kernel size 3×3，padding 1，stride 1），输出特征图大小为（ A ）。

A. 128×128 B. 127×127

C. 126×126 D. 125×125

18. 在PyTorch框架中，若要实现神经网络的参数优化，应该使用以下哪个模块？（B ） A. torch.nn B. torch.optim C. torch.tensor D. torch.cuda

19. 以下关于循环神经网络（RNN）的说法，不正确的是（ D ）。

A. RNN可以处理变长的序列数据

B. 梯度消失问题是RNN训练中的一个常见问题

C. LSTM是RNN的一种变体，可以有效解决梯度消失问题

D. RNN的输出只依赖于当前时刻的输入和前一个时刻的隐藏状态

20. 为了提高卷积神经网络的泛化能力，通常采用以下哪种技术？（ A ）

A. 数据增强 B. 提高学习率 C. 减少卷积核数量 D. 增加网络层数

三、判断题

1.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。 （ √ ）

2.提升卷积核(convolutional kernel)的大小会显著提升卷积神经网络的性能。 （ × ）

3.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。 （ × ）

4.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。 （ × ）

5.当在卷积神经网络中加入池化层(pooling layer)时，平移变换的不变性会被保留。 （ √ ）

6.卷积神经网络结构具有权重共享的特性。 （ √ ）

7.全连接神经网络结构具有权重共享的特性。 （ × ）

8.深度学习模型在训练集上的误差称为训练误差。 （ √ ）

9.深度学习模型在测试集上的误差称为泛化误差。 （ √ ）

10.使用深度学习模型拟合训练集，训练误差大，泛化误差小，称为过度拟合现象。 （ × ）

# 四、概念题

1. 过拟合

过拟合是指训练误差和测试误差之间的差距太大，模型复杂度高于实际问题，模型在训练集上表现很好，但在测试集上却表现很差。没有理解数据背后的规律，泛化能力差。(2分)

2. 随机梯度下降

不使用全量的样本来计算梯度，而使用单一样本来近似估计梯度，可以极大地减少计算量，提高计算效率。每次从训练集中随机选择一个样本，计算其对应的损失和梯度，进行参数更新，反复迭代。

3. 语义分割

语义分割旨在将图像中的每个像素分配给预定义的语义类别之一。语义分割不仅需要标识图像中存在的对象，还需要为每个像素赋予一个语义标签。

4. 多层感知器

多层感知器是一种前向结构的人工神经网络，映射一组输入向量到一组输出向量。可以被看作是一个有向图，由多个的节点层所组成，每一层都全连接到下一层。

5. 梯度消失

梯度消失是指当神经网络层数增多时，越靠近输入层的层之间梯度可能变得非常小，接近于零。这会导致网络难以学习和更新这些较早层次的参数，从而影响网络的性能。

6. 优化器

优化器是在深度学习的反向传播过程中，指引损失函数（目标函数）的各个参数往正确的方向更新合适的大小，使得更新后的各个参数让目标函数不断逼近全局最小点。

7. 批量归一化

批量归一化是一种方法，通过对层的输入进行重新居中和重新缩放，使人工神经网络的训练更快、更稳定。

9. 循环神经网络

循环神经网络是一类以序列数据为输入，在序列的演进方向进行递归且所有节点按链式连接的递归神经网络。

10. 欠拟合

欠拟合是指模型的复杂度低于实际问题，导致模型在训练集和测试集上的表现都不够好。这通常是因为模型未能捕捉到数据中的关键特征，或者模型太过简单，无法很好地拟合数据。

11. 小批量梯度下降

小批量梯度下降是介于批量梯度下降和随机梯度下降之间的一种方法。它将训练数据分为多个小批量，每个批量包含一定数量的样本，然后计算每个批量的梯度并更新参数。

12. Deep Q-Network

Deep Q-Network是一种结合了深度学习和强化学习的算法，用于解决具有高维输入空间的决策问题。使用深度神经网络来近似Q函数，并通过经验回放和目标网络技术来稳定训练过程。

# 五、简答题

1. 有一个LeNet的结构，输入数据大小为28×28大小单通道图片，在C1卷积层使用8个大小为3×3的卷积核进行卷积，padding=1，步长为1。在S2池化层使用2×2的卷积核进行步长为2的最大值池化。在C3卷积层使用12个大小为3×3的卷积核进行卷积，padding=1，步长为1。在S4池化层使用2×2的卷积核进行步长为2的最大值池化。扁平化后接入C5全连接层，其中C5全连接层的输出大小为84，请问C5全连接层的输入是多大。

原始输入为28×28

C1的输出为(28+2×1-3)/1+1=28 特征图大小为28×28×8

S2的输出为28/2=14 特征图大小为14×14×8

C3的输出为(14+2×1-3)/1+1=14 特征图大小为14×14×12

S4的输出为14/2=7 特征图大小为7×7×12

C5的输入为7×7×12=588

2. 考虑一个简化版的VGG网络结构，输入数据大小为32×32大小三通道图片。在第一个卷积层C1使用16个大小为3×3的卷积核进行卷积，padding=1，步长为1。在第一个池化层S2使用2×2的核进行步长为2的最大值池化。在第二个卷积层C3使用32个大小为3×3的卷积核进行卷积，padding=1，步长为1。在第二个池化层S4使用2×2的核进行步长为2的平均值池化。扁平化后接入全连接层F5，其中F5全连接层的输出大小为128。请问F5全连接层的输入是多大

答：

原始输入为32×32×3

C1的输出为(32+2×1-3)/1+1=32 特征图大小为32×32×16

S2的输出为32/2=16 特征图大小为16×16×16

C3的输出为(16+2×1-3)/1+1=16 特征图大小为16×16×32

S4的输出为16/2=8 特征图大小为8×8×32

F5的输入为8×8×32=2048

2. LSTM结构图如图1所示，请写出遗忘门、输入门、候选状态、输出门、细胞状态以及隐含状态的计算表达式。

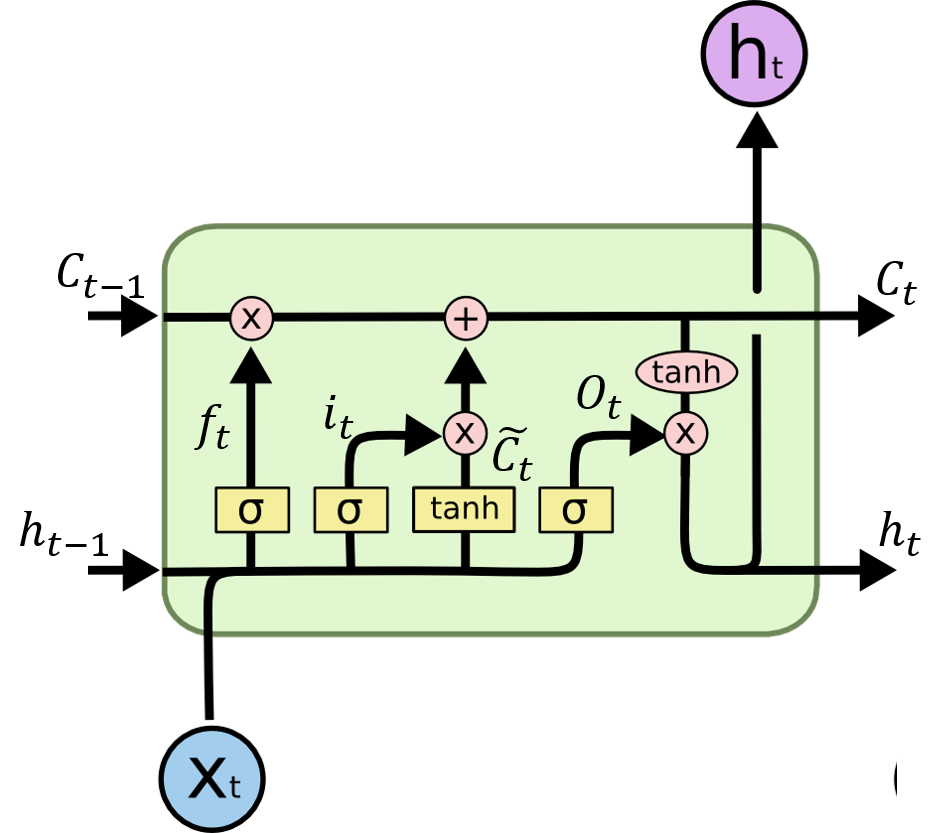


图1













1. 对函数应用梯度下降法。假定初始值，学习率固定为α=0.01，计算两次梯度下降法更新后的和。

第一步计算梯度：



第二步计算梯度：



3. 对函数应用梯度下降法。假定初始值，学习率固定为α=0.01，计算两次梯度下降法更新后的和。（8分）

答：

第一步计算梯度：



第二步计算梯度：



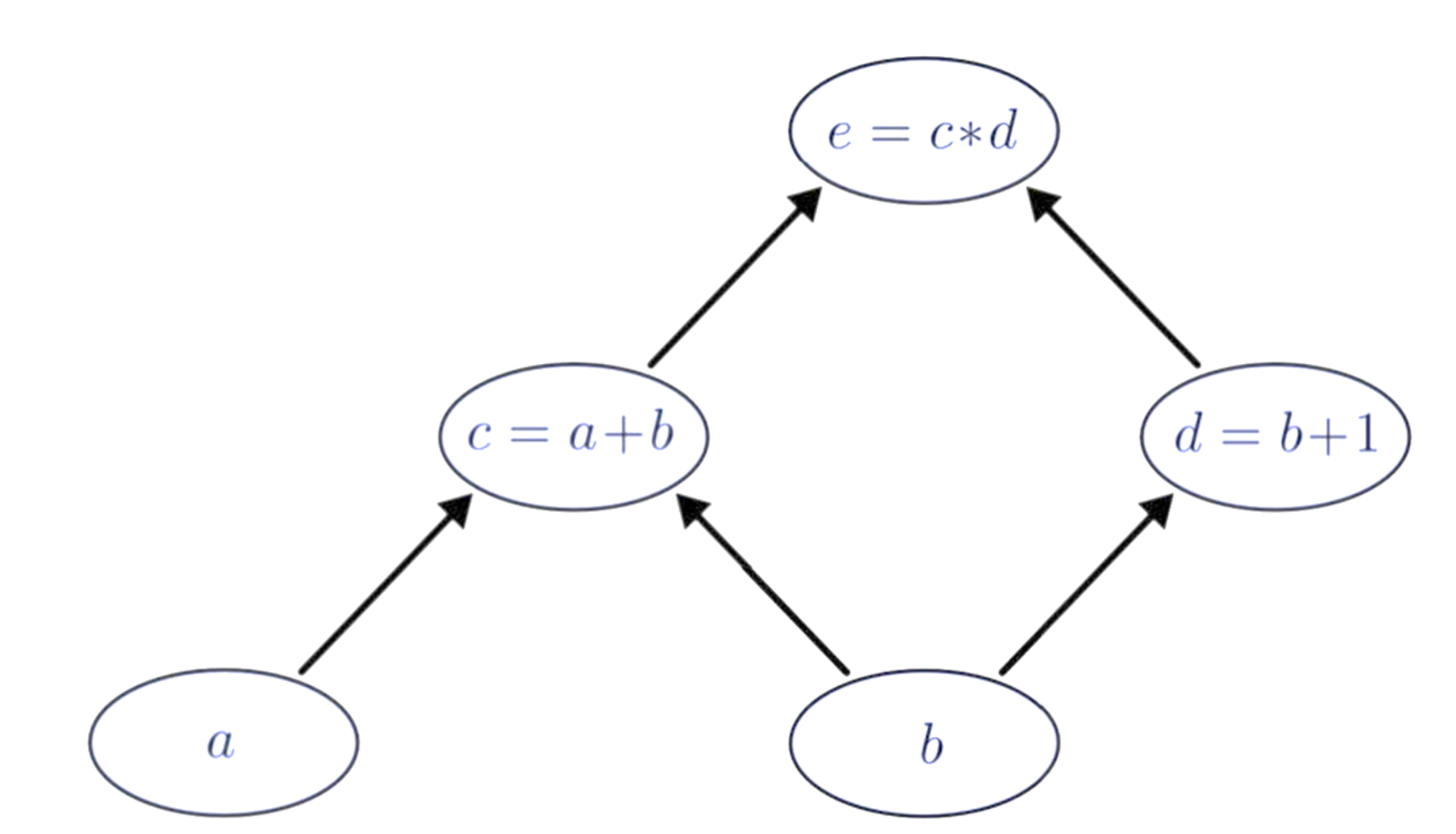
4.复合函数e=(a+b)\*(b+1),c=a+b,d=b+1,其中，a=2，b=1计算图如图2所示，请计算，,同时列出反向传播路径上的所有偏导值。。

图2



（1分）

,,

,（1分）

（2分）

（2分）

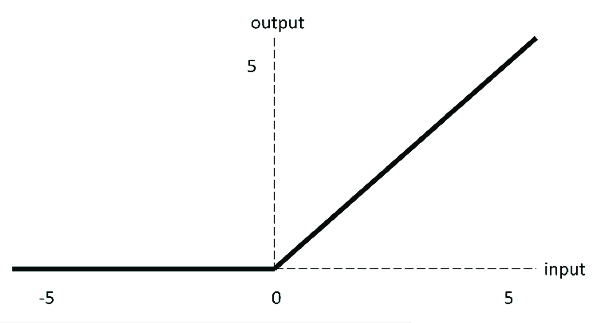
5. 简述ReLU激活函数的数学表达式、优点、缺点以及画出其图像。



优点：没有梯度消失问题，收敛速度快，计算速度快。

缺点：不以0为中心，部分数据提供梯度。

图像：

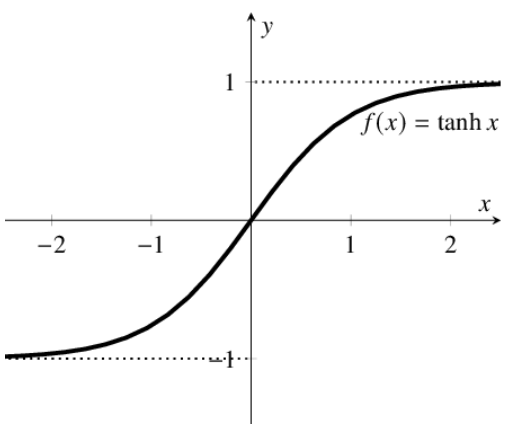


1. 简述Tanh激活函数的数学表达式、优点、缺点以及画出其图像。

优点：可导，以0为中心。

缺点：存在幂运算，容易梯度消失。

图像：



六、编程题

3. 搭建一个具有长短时记忆网络（LSTM）结构的神经网络类，用于处理输入大小为 1 的时间序列数据。该网络结构包括以下层次：

1. LSTM层：输入维度为1，隐藏层维度为100。
2. 线性层：用于将LSTM输出映射到预测结果，输出维度为1。
3. 初始化隐藏状态和细胞状态为零的操作。

请完成以下代码：

import torch

import torch.nn as nn

class LSTM(nn.Module):

def \_\_init\_\_(self, input\_size=1, hidden\_layer\_size=100, output\_size=1):

super().\_\_init\_\_()

# 定义LSTM的输入维度、隐藏层维度和输出维度

# 创建LSTM层

# 创建线性层，用于将LSTM输出映射到预测结果

# 初始化隐藏状态和细胞状态

def forward(self, input\_seq):

# 将输入序列传递给LSTM层，并获取输出和新的隐藏状态/细胞状态

# 将LSTM输出传递给线性层，得到预测结果

# 返回最后一个时间步的预测结果

答：

import torch

import torch.nn as nn

class MyLSTM(nn.Module):

def \_\_init\_\_(self, input\_size=1, hidden\_layer\_size=100, output\_size=1):

super().\_\_init\_\_()

# 定义LSTM的输入维度、隐藏层维度和输出维度

**self.hidden\_layer\_size = hidden\_layer\_size**

# 创建LSTM层

**self.lstm = nn.LSTM(input\_size, hidden\_layer\_size)**

# 创建线性层，用于将LSTM输出映射到预测结果

**self.linear = nn.Linear(hidden\_layer\_size, output\_size)**

# 初始化隐藏状态和细胞状态

**self.hidden\_cell = (torch.zeros(1, 1, self.hidden\_layer\_size),**

**torch.zeros(1, 1, self.hidden\_layer\_size))**

def forward(self, input\_seq):

**# 将输入序列传递给LSTM层，并获取输出和新的隐藏状态/细胞状态**

**lstm\_out, self.hidden\_cell = self.lstm(input\_seq.view(len(input\_seq), 1, -1), self.hidden\_cell)**

**# 将LSTM输出传递给线性层，得到预测结果**

**predictions = self.linear(lstm\_out.view(len(input\_seq), -1))**

**# 返回最后一个时间步的预测结果**

**return predictions[-1]**

5. 搭建一个卷积神经网络（CNN）类，用于处理输入大小为 32×32 的三通道图像数据。该网络结构包括以下层次：

第一层卷积：通道数从3增加到16，卷积核大小为3，步长为1，填充为1，采用ReLU激活函数。接着使用最大池化层，核大小为2，步长为2。

第二层卷积：通道数保持为16，卷积核大小为3，步长为1，填充为1，同样采用ReLU激活函数。随后使用平均池化层，核大小为2，步长为2。

将输出展平为一维向量。

第一个全连接层：输出大小为128，采用ReLU激活函数。

第二个全连接层：输入大小为128，输出大小为10，最后采用Softmax激活函数。

结果要包括\_\_init\_\_()函数和forward()函数。

import torch

import torch.nn as nn

import torch.nn.functional as F

class Net(nn.Module):

def \_\_init\_\_(self):

super().\_\_init\_\_()

**self.conv1 = nn.Conv2d(3,16, kernel\_size=3, stride=1, padding=1)**

**self.conv2 = nn.Conv2d(16,16, kernel\_size=3, stride=1, padding=1)**

**self.fc1 = nn.Linear(16\*8\*8, 128)**

**self.fc2 = nn.Linear(128, 10)**

def forward(self, x):

**x = F.max\_pool2d(F.relu(self.conv1(x)), kernel\_size=2, stride=2)**

**x = F.avg\_pool2d (F.relu(self.conv2(x)), kernel\_size=2, stride=2)**

**x = x.view(-1, 256)**

**x = F.relu(self.fc1(x))**

**x = self.fc2(x)**

**return F.softmax(x, dim=1)**