



# 人工智能导论

----连接学派例题



华中科技大学

HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

- 不定项选择题

1、卷积神经网络（CNN）中的卷积层的主要作用是什么？（ B ）

- A. 降低数据维度
- B. 提取局部特征
- C. 减少过拟合
- D. 增强数据噪声

- 不定项选择题

2、卷积神经网络中的池化层通常用于什么？（ C ）

- A. 增加数据的分辨率
- B. 提高模型的计算复杂度
- C. 进行特征选择和降维
- D. 生成卷积核

- 不定项选择题

3、循环神经网络（RNN）最适合处理哪种类型的数据？（ B ）

- A. 静态图像数据
- B. 序列数据
- C. 无标签数据
- D. 随机噪声数据



- 不定项选择题

4. 在RNN中，哪个变量负责在每个时间步之间传递信息？（ A ）
- A. 隐状态
  - B. 权重矩阵
  - C. 输入向量
  - D. 激活函数

- 不定项选择题

5. 在Transformers中，注意力机制的主要功能是？（ B ）

- A. 提高模型的收敛速度
- B. 捕获输入序列中不同位置的依赖关系
- C. 增加网络的非线性表达能力
- D. 降低模型的计算复杂度

- 不定项选择题

6、以下哪些是卷积神经网络（CNN）的关键组件？（ ABC ）

- A. 卷积层
- B. 池化层
- C. 全连接层
- D. 循环层



- 不定项选择题

7. 图神经网络适合处理以下哪些数据？（ ABCD ）

- A. 社交网络
- B. 化学分子
- C. 交通网络
- D. 电路设计图



4、M-P模型被认为是最早的神经元模型，其拥有布尔输入和输出以及阶跃响应函数。前馈神经网络引入了反向传播的机制和S型函数解决了样本的线性不可分问题。

（ADALINE模型、M-P模型、感知器模型、循环神经网络、前馈神经网络、Hopfield网络、链式法则、线性不可分、通用近似、序列建模、联想记忆）

5、一个最朴素的卷积神经网络可仅由\_\_\_\_\_卷积层\_\_\_\_\_和  
\_\_\_\_\_非线性激活单元\_\_\_\_\_构成。如果用卷积神经网络去解决一个图像  
分类问题，对最终性能起决定性因素的除了网络的深度，还有  
\_\_\_\_\_数据集的规模\_\_\_\_\_和\_\_\_\_\_网络优化的难度\_\_\_\_\_。

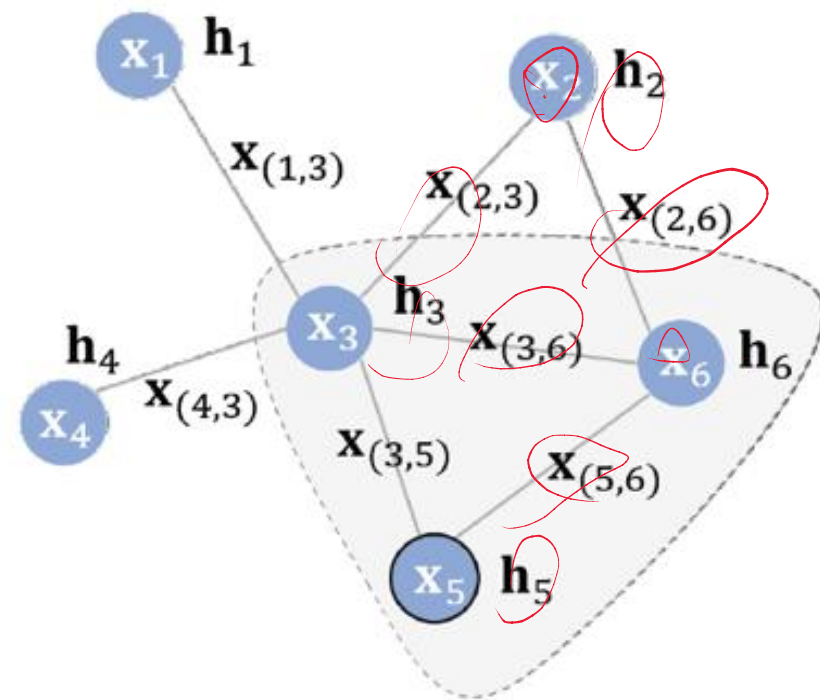
（卷积层、池化层、归一化层、非线性激活单元、全连接层、特征  
表达的能力、数据集的规模、网络的参数量、网络的计算量、网络  
优化的难度、残差连接、Dropout、网络的GPU使用率）

6、GAN网络通过引入推土机（Wasserstein）距离，解决了网络训练的不稳定性，引入条件约束增加了输出的可控性，利用循环一致性减少了对成对数据的依赖，并采用渐进式训练缓解训练高分辨率图像时的不稳定问题。

（卷积网络、批归一化层、推土机（Wasserstein）距离、欧式距离、KL散度、条件约束、随机噪声、监督式学习、成对的数据、循环一致性、渐进式训练、预训练、无监督训练）

7、在图神经网络的节点更新过程中，  
请问隐藏状态 $h_6$ 的获取与图中哪些状态有关？请写出所有相关状态（空白不必写满）。

\_\_\_ $x_2$ \_\_\_、\_\_\_ $x_3$ \_\_\_、\_\_\_ $x_5$ \_\_\_、\_\_\_ $x_6$ \_\_\_、  
\_\_\_ $x(2,6)$ \_\_\_、\_\_\_ $x(3,6)$ \_\_\_、\_\_\_ $x(5,6)$ \_\_\_、  
\_\_\_ $h_2$ \_\_\_、\_\_\_ $h_3$ \_\_\_、\_\_\_ $h_5$ \_\_\_



### question1: 验证感知器的计算

请用如下多层感知器模型（如图1）验证异或（满足如表1所示的规则）问题，其中激活函数  $f(x) = \begin{cases} 1, x \geq \theta \\ 0, x < \theta \end{cases}$ 。请写出详细计算过程。

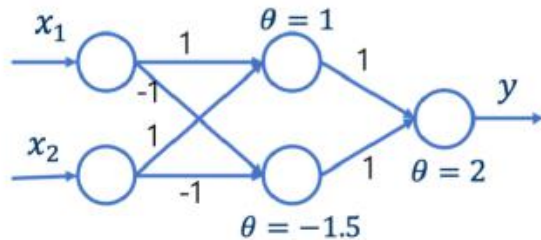


图 1: 多层感知器模型

$x_1$	$x_2$	$y$
0	0	0
0	1	1
1	0	1
1	1	0

表 1: 运算规律

解：根据感知器的工作原理，将图 1 的多重感知器分解为图 2 所示的三个单层感知器。则根据图 2 中 (a)、(b)、(c)，我们可以得到： $y_1 = f(x_1 \times 1 + x_2 \times 1), \theta = 1$ ， $y_2 = f(x_1 \times -1 + x_2 \times -1), \theta = -1.5$  和  $y = f(y_1 \times 1 + y_2 \times 1), \theta = 2$ 。则：

当  $x_1 = 0, x_2 = 0$  时， $y_1 = 0, y_2 = 1, y = 0$ 。

当  $x_1 = 0, x_2 = 1$  时， $y_1 = 1, y_2 = 1, y = 1$ 。

当  $x_1 = 1, x_2 = 0$  时， $y_1 = 1, y_2 = 1, y = 1$ 。

当  $x_1 = 1, x_2 = 1$  时， $y_1 = 1, y_2 = 0, y = 0$ 。

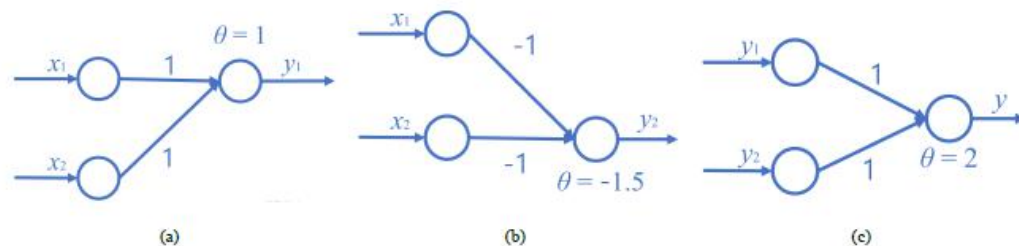


图 2

2. 请判断以下CNN说法的正误，并解释原因：

- 全连接层通常只出现在CNN的最后几层。
- 只有池化层才能起到下采样的作用。
- 残差网络采用跳跃连接计算一个非线性函数使得输入能传递到更深的层。

正确，全连接层通常作为分类器，只出现在CNN的最后输出部分。

错误，下采样是指降低特征图或者输入图片分辨率的操作，除了池化以外，步长大于1的卷积，插值等等也可以实现下采样。

错误，残差网络中的跳跃连接是一个恒等映射，不是非线性函数。

有同学把分析重心落在了能否把输入传递到更深的层，  
其实这里只是一个计算方式的简单问题。



## RNN 中为什么会出现梯度消失？如何解决 RNN 中的梯度消失问题？

RNN 中梯度消失从本质上来讲是因为梯度反向传播中的连乘效应。

假设 RNN 模型如下：

隐藏状态：  $h^{(t)} = \sigma(z^{(t)}) = \sigma(Ux^{(t)} + Wh^{(t-1)} + b)$ ，此处激活函数一般为  $\tanh$ ；

模型输出：  $o^{(t)} = Vh^{(t)} + c$ ；

预测输出：  $\hat{y}^{(t)} = \sigma(o^{(t)})$ ；

模型损失：  $L = \sum_{i=1}^T L^{(i)}$ 。

$W$  的梯度表达式为：

$$\frac{\partial L}{\partial W} = \sum_{t=1}^T \frac{\partial L}{\partial y^{(T)}} \frac{\partial y^{(T)}}{\partial o^{(T)}} \frac{\partial o^{(T)}}{\partial h^{(T)}} \frac{\partial h^{(T)}}{\partial h^{(t)}} \frac{\partial h^{(t)}}{\partial W}$$

式中的  $\frac{\partial h^{(T)}}{\partial h^{(t)}}$  可展开为：

$$\frac{\partial h^{(T)}}{\partial h^{(t)}} = \prod_{k=t+1}^T \frac{\partial h^{(k)}}{\partial h^{(k-1)}} = \prod_{k=t+1}^T \tanh'(z^{(k)})W$$

进一步可以得到  $W$  的梯度表达式为：

$$\frac{\partial L}{\partial W} = \sum_{t=1}^T \frac{\partial L}{\partial y^{(T)}} \frac{\partial y^{(T)}}{\partial o^{(T)}} \frac{\partial o^{(T)}}{\partial h^{(T)}} \left[ \prod_{k=t+1}^T \tanh'(z^{(k)})W \right] \frac{\partial h^{(t)}}{\partial W}$$

其中  $\tanh'(z^{(k)}) \leq 1$ ，而  $W$  又可以进行特征值分解：

$$W = V \text{diag}(\lambda) V^{-1}$$

如果  $W$  的主特征值小于 1（这是大部分情况），则会出现梯度消失的情况。

解决方法：

(1) 使用 ReLU 作为激活函数；ReLU 函数的左侧导数为 0，右侧导数恒为 1，这就避免了“梯度消失”的发生。但恒为 1 的导数容易导致“梯度爆炸”，设定合适的阈值可以解决这个问题。

(2) 更改内部结构来解决梯度消失问题，使用 LSTM。

这里是最完整的解答。比较简便的说法可以直接设权重矩阵  $W$  可分解为  $W = V \text{diag}(\lambda) V^{-1}$ ，一般而言的特征值小于 1，经过序列的多次操作会导致值趋于 0，也会让梯度消失。

有一些同学的叙述太模糊，或者不全面。比如“重复执行小于 1 的操作”，或者认为是激活函数导致了梯度越来越小。

## 考虑偏置:

假设输入是一张 $128 \times 128$ 的RGB彩色图像,如果网络第一个卷积层的卷积核为 $3 \times 3$ 的大小,步长为1,共有64个卷积核后接一个步长为2的 $2 \times 2$ 平均池化层,随后再接两层核大小为 $5 \times 5$ 、步长为1、卷积核个数为128的卷积层以及一层步长为2的 $2 \times 2$ 的平均池化层,最后接含有128个神经元的全连接层和一个含有10个神经元的输出层。试回答:

- (1) 通过平均池化层后特征图的大小( $H \times W \times C$ )为多少?
- (2) 计算该网络的参数量。
- (3) 网络的参数量主要由哪部分贡献?贡献了百分之多少?( $3 \times 3$ 卷积的padding为1,  $5 \times 5$ 卷积核的padding为2)。

### 主要问题:

- 1、答案可以有两种,一种考虑偏置,一种不考虑。
- 2、部分同学没有注意中间是再接2层 $5 \times 5$ 的卷积,导致少考虑一层卷积。

1. 通过平均池化层后特征图的大小( $H \times W \times C$ )为多少?

解:  $64 \times 64 \times 64$

2. 计算该网络的参数量

解: conv-1 参数:  $3^2 \times 3 \times 64 + 64 = 1792$

conv-2 参数:  $5^2 \times 64 \times 128 + 128 = 204,928$

conv-3 参数:  $5^2 \times 128 \times 128 + 128 = 409,728$

FC-1 参数:  $32^2 \times 128 \times 128 + 128 = 16,777,344$

FC-2 参数:  $128 \times 10 + 10 = 1,290$

总参数:  $1792 + 204928 + 409728 + 16777344 + 12190 = 17,395,082$

3. 网络的参数量主要有由哪部分贡献?贡献了百分之几?

解: 主要由全连接层贡献,贡献了:

$$\frac{16,777,344}{17,395,082} \times 100\% = 96.45\%$$



## 不考虑偏置：

假设输入是一张 $128 \times 128$ 的RGB彩色图像，如果网络第一个卷积层的卷积核为 $3 \times 3$ 的大小，步长为1，共有64个卷积核后接一个步长为2的 $2 \times 2$ 平均池化层，随后再接两层核大小为 $5 \times 5$ 、步长为1、卷积核个数为128的卷积层以及一层步长为2的 $2 \times 2$ 的平均池化层，最后接含有128个神经元的全连接层和一个含有10个神经元的输出层。试回答：

- (1) 通过平均池化层后特征图的大小( $H \times W \times C$ )为多少？
- (2) 计算该网络的参数量。
- (3) 网络的参数量主要由哪部分贡献？贡献了百分之多少？（ $3 \times 3$ 卷积的padding为1， $5 \times 5$ 卷积核的padding为2）。

### 主要问题：

- 1、答案可以有两种，一种考虑偏置，一种不考虑。
- 2、部分同学没有注意中间是再接2层 $5 \times 5$ 的卷积，导致少考虑一层卷积。

- 经过第一个平均池化层后，特征图大小为 $64 \times 64 \times 64$ ；  
经过第二个平均池化层后，特征图大小为 $32 \times 32 \times 128$ 。
- 该网络结构如下表所示。在不考虑 bias 的情况下，对于第一个卷积层，参数量为

$$3 \times 3 \times 3 \times 64 = 1,728 \quad (25)$$

同理，之后层参数量为

$$5 \times 5 \times 64 \times 128 = 204,800 \quad (26)$$

$$5 \times 5 \times 128 \times 128 = 409,600 \quad (27)$$

$$32 \times 32 \times 128 \times 128 = 16,777,216 \quad (28)$$

$$128 \times 10 = 1,280 \quad (29)$$

则参数总个数为 17,394,624。

Type/stride	Filter Shape	Input Size
Conv/s1	$3 \times 3 \times 3 \times 64$	$128 \times 128 \times 3$
Avg Pool/s2	Pool $2 \times 2$	$128 \times 128 \times 64$
Conv/s1	$5 \times 5 \times 64 \times 128$	$64 \times 64 \times 64$
Conv/s1	$5 \times 5 \times 128 \times 128$	$64 \times 64 \times 128$
Avg Pool/s2	Pool $2 \times 2$	$64 \times 64 \times 128$
FC/s1	$32 \times 32 \times 128 \times 128$	$32 \times 32 \times 128$
FC/s1		128
sigmoid/s1	Classifier	$1 \times 10$

- 参数量主要由全连接层贡献，贡献计算如下：

$$(16,777,216 + 1280) \div 17,394,624 = 0.9446 \quad (30)$$

2.1 GCN中主要靠\_\_\_\_\_的作用使得卷积能够方便的在不规则的图数据结构上运算。

邻接矩阵

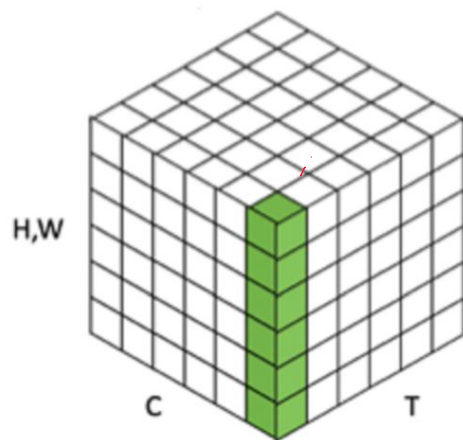
注：拉普拉斯矩阵也是正确答案，但注意拉普拉斯矩阵和邻接矩阵有区别。且GCN中涉及的是归一化后的拉普拉斯矩阵。

2.2 请问常规卷积和图卷积有哪些相同点和不同点？试各回答三点。

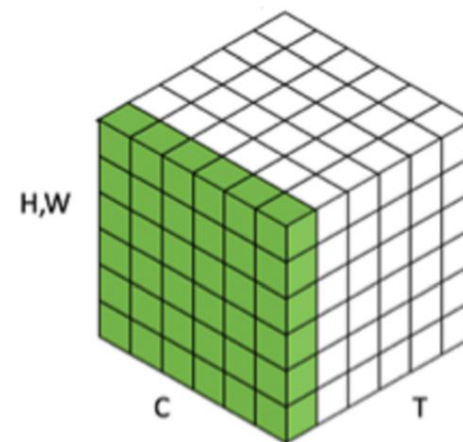
相同点：1) 局部连接； 2) 权值共享； 3) 多层网络。

常规卷积	图卷积
定义在欧氏空间	定义在非欧氏空间
节点的邻域数量固定	节点的邻域不固定
卷积核大小固定	卷积核大小不固定

# 1. 以下两个图分别表示什么类型的注意力机制？



该图代表空间上的注意力机制，利用注意力机制调整空间上某些区域的重要性。



该图代表空间和通道上的共同注意力机制，利用注意力机制同时调整空间上特定区域和特征图特定通道的重要性。