

一、计算分析题（共 1 题，每题 30 分，共 30 分。）

1.考虑一个卷积神经网络实现彩色图像分类任务，输入数据的维数为 $(32, 32, 3)$ ，分别表示高、宽和通道数。符号表示如下：

CONV-K-N-S-P: 每个卷积核的大小为 $K \times K$ ， N 个卷积核， S 表示步长， P 表示填充的长度；

POOL-K: 表示一个 $K \times K$ 的最大池化层，步长为 K ，零填充；

FLATTEN: 对输入进行展开操作；

FC-N: 表示带 N 个神经元的全连接层。

请依次计算并回答下列问题：

（1）若第一层使用 **CONV-5-6-1-0**，请计算第一层输出的维数和本层使用的参数个数。（3 分）

（2）若第一层使用 **CONV-5-6-1-P**，在一些场景下（如 **Res-Net**）我们要求等宽卷积时，请计算 P 。（2 分）

（3）在第（1）基础上，第二层使用 **POOL-2**，请计算第二层输出的维数和本层使用的参数个数。（5 分）

（4）若第一层使用 **FLATTEN + FC-100**，请计算第一层输出的维数和本层使用的参数个数。（5 分）

（5）若我们设计了一个如下表所示的卷积神经网络，请计算后填表：
（15 分）

层	输出数据维数（三维表示）	参数个数
INPUT	$32 \times 32 \times 3$	0
CONV-5-6-1-0	$32 \times 32 \times 3$	0

ReLU		
POOL-2		
CONV-5-60-2-1		
ReLU		
POOL-2		
FLATTEN		
FC-10		

一、计算题 (共 1 题, 每题 15 分, 共 15 分。)

设矩阵 A 为: $A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 1 & 2 \\ 1 & 1 & 2 & 2 \\ 2 & 2 & 1 & 1 \end{pmatrix}$, 设矩阵 B 为: $B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$,

- (1). 在步长 (Stride)=1 和零填充 (Padding)=1 时, 计算它们的卷积 $B \otimes A$;
- (2). 进一步计算在 2×2 , 步长为 2 时的 Maxpooling($B \otimes A$)。

二、证明题 (共 1 题, 每题 10 分, 共 10 分。)

前馈神经网络使用公式 $z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)}$, $z^{(l)} \in \mathbb{R}^{M_l}$, $a^{(l-1)} \in \mathbb{R}^{M_{l-1}}$, $W^{(l)} \in \mathbb{R}^{M_l \times M_{l-1}}$ 进行前馈计算, 试证: $\frac{\partial z^{(l)}}{\partial w_{i,j}^{(l)}} = l_i(a_j^{(l-1)})$, 其中 $w_{i,j}^{(l)}$ 是 $W^{(l)}$ 第 i 行第 j 列的元素。

2. 主观题 (30 分)

二、综合题 (共 1 题, 每题 30 分, 共 30 分。)

2. 循环神经网络往往可以用来建模文本、语音等时序数据。请依次回答以下问题:

- (1) 请画出简单循环神经网络的网络结构图, 并写出隐状态更新的迭代公式。(10 分)
- (2) 试证随时间反向传播算法 (BPTT):

$$\frac{\partial L}{\partial U} = \sum_{t=1}^T \sum_{k=1}^t \delta_{t,k} h_{k-1}^T, \text{ 其中 } \delta_{t,k} = \prod_{z=k}^{t-1} (\text{diag}(f'(z_t)) \cdot U^T) \delta_{t,t}. \text{ (10 分)}$$

- (3) 请描述上述简单循环神经网络的缺陷, 画出基于门控的 GRU (Gated Recurrent Unit) 的网络结构图, 并详细分析每个门控的功能。(10 分)

3. 主观题 (20 分)

三、证明题 (共 1 题, 每题 20 分, 共 20 分。)

3. 神经网络有非常强大的拟合能力, 但是其 90 年代的预测效果并不好。其中一个重要的原因是网络权重的初始化对结果影响很大, 接近最优解的初始化往往能很快收敛到最优解, 但是相距甚远的初始化收敛很慢甚至不收敛。现代理论往往采用给定概率分布下随机初始化权重, 最具代表性的是 Xavier 初始化和 He 初始化。

(1) 若使用正态分布 $N(0, \sigma^2)$ 初始化, 使用 Logistic 激活函数, 试证 Xavier 初始化: $\sigma^2 = 16 \times \frac{2}{M_{l-1} + M_l}$, 其中 M_l 为第 l 层神经元个数。(10 分)

(2) 若使用均匀分布 $U(-r, r)$ 初始化, 使用 ReLU 激活函数, 试证 He 初始化: $r = \sqrt{\frac{6}{M_{l-1}}}$, 其中 M_l 为第 l 层神经元个数。(10 分)

4. 主观题 (20 分)

四、简答题 (共 2 题, 每题 10 分, 共 20 分。)

请简要回答训练误差和泛化误差两者之间的关系, 并列举你认为可有效降低泛化误差的深度学习技术。(10 分)

简单描述损失函数的地貌概念, 并简要回答如何从动态学习率调整和梯度估计修正两方面来进行更有效的优化, 使之能收敛到损失函数地貌的最小值。(10 分)