

PRML 第五次作业

神经网络分类器

学生： 槐 泽 鹏

导师： X X X

学号:xxxxxxxxxxxxxxxxxx

联系方式: xxxxxxxxxxxxxx

2021 年 1 月 11 日

目 录

第 1 节 神经网络	1
第 2 节 实验	2
2.1 数据集	3
2.2 结果	3
2.3 分析	3
第 3 节 小结	5
第 4 节 参考文献	5
第 5 节 代码	6

第 1 节 神经网络

本次大作业旨在完成博士机器学习及模式识别课程第五次编程大作业。

MLP(多层感知器, Multi-layer perceptron) 和 CNN(Convolutional Neural Networks) 是两类常见的神经网络基本模型, 前者常用于特征提取之后作分类器, 后者常用于对图像等数据进行特征提取。

MLP^[1] 由一个输入层, 一个输出层和一个或多个隐藏层组成。MLP 中的所有神经元结构类似, 每个神经元都有几个输入(连接前一层)神经元和输出(连接后一层)神经元, 该神经元会将相同值传递给与之相连的多个输出神经元, 如图 1.1所示。

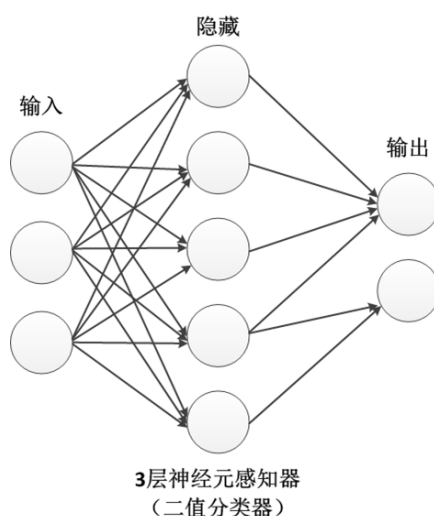


图 1.1 MLP 示意图

CNN 作为深度学习的开山之作, 是谷歌的 ImageNet 比赛上所有经典模型^[2-5]的基本组成部分, 也是深度学习的最基本又很重要的基本模型。其旨在对空间规律分布数据进行特征提取, 不同 CNN 组合成的神经网络模型极大地促进了深度学习乃至人工智能的发展。

卷积神经网络通常包含以下几种层:

卷积层 (Convolutional layer), 卷积神经网络中每层卷积层由若干卷积单元组成, 每个卷积单元的参数都是通过反向传播算法优化得到的。卷积运算的目的是提取输入的不同特征, 第一层卷积层可能只能提取一些低级的特征如边缘、线条和角等层级, 更多层的网络能从低级特征中迭代提取更复杂的特征。

池化层 (Pooling layer), 通常在卷积层之后会得到维度很大的特征, 将特征切成几个区域, 常见池化有最大池化或平均池化, 得到新的、维度较小的特征,

同时可以降低对像素的敏感性。

全连接层（Fully-Connected layer），常用于最后作分类器。

典型 CNN 结构如图 1.2所示。

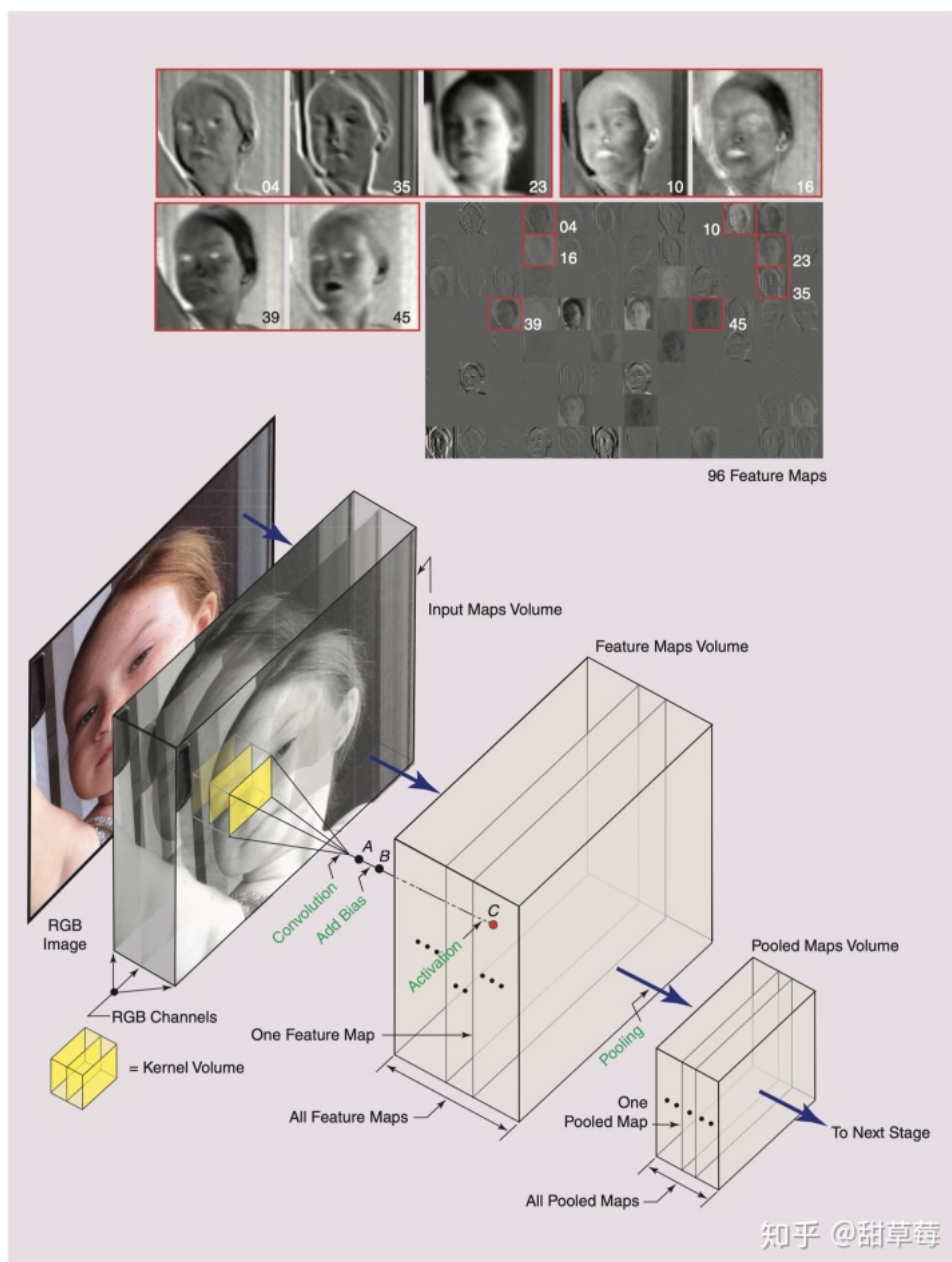


图 1.2 CNN 示意图

第 2 节 实验

本节我们在 3 个数据集上测试 MLP，在 MNIST 上测试 CNN。

2.1 数据集

多层感知器采用 3 个数据集：HCV、Abalone 和 Iris，他们的静态参数如表 1 所示。CNN 网络采取 MNIST 数据集，由于全部选取所需内存过大，本次大作业选取该数据集一部分作为训练集，其静态参数如表 2 所示。

同时，为分析不同层数和隐含层单元数量对神经网络模型的影响，这里选取不同层数和不同隐含层单元数量的网络模型。详细结果见下节。

表 1: 数据集参数

数据集名称	参数		
	所有/代码使用属性数量	分类名称/数量	样本数量
HCV	14/10	病症类别/4	615
Abalone	7/7	性别/3	4176
Iris	4/4	植物类别/3	150

表 2: 数据集参数

数据集名称	参数	
	标签类别数量	样本数量
MNIST	10	2000

2.2 结果

表 3 给出 MLP 在不同数据集上的测试结果。表 4 给出 CNN 在不同数据集上的测试结果。

2.3 分析

对于 MLP，实验结果有以下规律：

1. 随着层数加深，特征提取和分类效果越好，这代表越深的模型特征提取能力越强。
2. 但是进一步加深网络深度，效果反而衰减，这是由于过拟合，这也是残差、dropout 等结构的优越所在。
3. 隐含层神经元越多的确会提升效果，但是达到一定程度后就饱和，效果无法再提升。例如，HCV 数据集、3 层网络下的测试结果随着神经元数量增加而提升，但是，Abalone 数据集、5 层网络下的测试结果就已经饱和。

对于 CNN，实验结果有以下规律：

表 3: MLP 测试结果

数据集	隐含层节点数	层数				
		3	4	5	6	7
HCV	5	0.900	0.885	0.881	0.891	0.878
	16	0.912	0.894	0.924	0.906	0.904
	32	0.933	0.937	0.959	0.933	0.956
Abalone	5	0.512	0.534	0.547	0.543	0.538
	16	0.540	0.525	0.542	0.533	0.542
	32	0.535	0.536	0.545	0.537	0.536
Iris	5	0.993	0.987	0.993	0.980	0.980
	16	0.986	0.986	0.986	0.986	0.993
	32	0.993	0.986	0.993	0.986	0.993
备注：学习率 0.01，权重衰减系数 0.001，优化算法 Adam						

表 4: CNN 测试结果

隐含层通道数	层数			
	3	4	5	6
[1,32,64,32]	0.758	None	None	None
[1,32,128,32]	0.783	None	None	None
[1,32,256,32]	0.774	None	None	None
[1,64,128,64]	0.637	None	None	None
[1,32,64,128,64]	None	0.807	None	None
[1,64,128,256,64]	None	0.535	None	None
[1,32,64,128,64,32]	None	None	0.693	None
[1,32,64,256,64,32]	None	None	0.725	None
[1,32,64,128,256,64,32]	None	None	None	0.676
备注： 卷积核 3*3 填充为 1(保证卷积后尺寸不变) 采取平均池化 学习率 0.01 权重衰减系数 0.001 优化算法 Adam 卷积层之后跟两个全连接层 最后一层输出维度为 10				

- 1.随着卷积层层数加深，特征提取和分类效果越好，这是由于感受野扩大以及模型特征提取能力增强导致。
- 2.但是进一步加深网络深度，效果反而衰减，这是由于过拟合。
- 3.卷积层通道数的确影响实验结果，在未过拟合情况下，增加通道数有利于提升模型分类精度，但是过分增加通道数也会带来效果衰减。

综合以上两个实验，有一些普遍共性规律：一是模型“越大越粗越深”的确会带来更强的特征提取和分类能力，但是过分堆积网络规模一定会带来过拟合和效果衰减。二是模型的超参数有时很影响分类结果，作者在实验中大致尝试了几个参数就有明显区别，但是这里没有采用循环调参，因为本次大作业旨在加深对神经网络的理解，因此没有过分追求精度。三是模型初始化的种子也会对分类结果带来几个点的影响。

第 3 节 小结

本次大作业旨在对 MLP 和 CNN 两种经典神经网络模型进行实验和分析。首先在 3 组数据集 HCV、Abalone 和 Iris 上测试了 MLP 分类结果，并分析了网络层数及隐含层维数对分类结果的影响；再在 MNIST 数据集上测试了 CNN 分类结果，并分析了卷积深度和通道数对分类结果的影响。最终实验结果符合预期，较好得完成了本次大作业。

第 4 节 参考文献

- [1] M. W Gardners. R Dorling. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 1998.
- [2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.
- [3] A Krizhevsky, I Sutskever, and G Hinton. Imagenet classification with deep convolutional neural networks. 2012.
- [4] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *Computer Science*, 2014.
- [5] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, and Andrew Rabinovich. Going deeper with convolutions. 2014.

第 5 节 代码

代码已经上传至本人 GitHub 仓库，详见

https://github.com/huaizepeng2020/PRML_HW5.git。

其中，*MLP_hcv.py*、*MLP_abalone.py* 和 *MLP_Iris.py* 是 MLP 测试代码。
CNN_MNIST.py 是 CNN 测试代码。