# 《人工智能基础》

# 实验报告

项目名称： 手写体数字识别

项目地址链接：https://aistudio.baidu.com/aistudio/projectdetail/4103894?contributionType=1&shared=1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 成员 | 组长 | 组员1 | 组员2 | 组员3 | 组员4 |
| 学号 | 20301164 | 20301155 | 20301172 | 20301133 | 20301154 |
| 姓名 | 柳萱莹 | 邹佳琪 | 王诗涵 | 李昊晴 | 周昱辰 |

1. 实验数据集
   1. 数据预处理
2. 说明对数据进行了哪些预处理工作，比如：裁剪、改变通道、改变格式、人脸检测等。说明处理的目的以及处理的结果（没有可以不写）。如果自己加入了新的数据，需要介绍新加入的数据（给出数量、尺寸、特点、如何采集的），以及加入新数据的目的。
3. 介绍最后所用数据集的规模、训练集中的样本数、测试集中的样本数、输入图像的尺寸。

答：（1）

预处理：

1. 读出所有训练数据和测试数据到train\_reader和test\_reader列表，将它们转换为数组形式分为图片数组和标签数组，数据类型转为double型和int型；
2. 通过给训练数据和测试数据的图片数组整体加负号，即将原黑底白字转化为白底黑字（新数据，即数量、尺寸与原数据集相同，特点就是变为白底黑字的了。原训练数据集数量60000，测试集数量10000）。接着将黑底白字图片数组与白底黑字图片数组拼接起来生成新数组，随机以相同状态打乱图片数组和标签数组，使它们打乱后依然对应；

处理的目的：

将数据集转化为numpy数组便于从测试数据和训练数据中分出图片和标签，也便于将黑底白字图转为白底黑字图并把二者做拼接并打乱作为新的训练数据和测试数据。

处理的结果：

得到混合有白底黑字和黑底白字图以及对应标签的四个数组all\_train\_image, all\_test\_image, all\_train\_label, all\_test\_label

（2）最后所用数据集规模：140000

训练集中样本数：120000

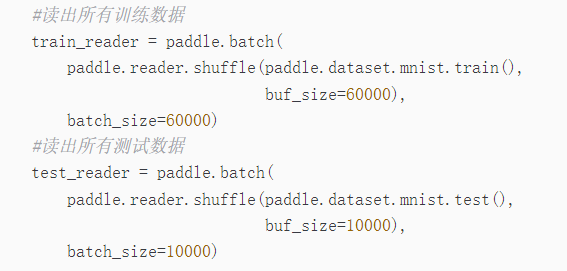
测试集中样本数：20000

输入图像的尺寸：[-1,1]（归一化处理后）

* 1. 数据加载

描述如何将训练数据和测试数据加载进入PaddlePaddle框架中的。说明加载的具体细节，比如：使用了PaddlePaddle里面哪种数据加载器，参数是如何设置的，这样设置的理由是什么。

答：（1）通过paddle框架获取mnist官方数据集的训练集和测试集



读出训练数据部分：

使用paddle.reader.shuffle()修改数据集大小，第一个参数是mnist训练数据集的creator，它返回一个reader creator，reader中的每个样本的图像像素范围是[-1,1]，标签范围是[0,9]；第二个参数是需要数据集的大小，这里设置为整个训练集的大小60000，先读出所有训练数据，为的是方便作后续的转白底黑字图的操作。

使用paddle.batch()将训练数据打包，每个batch的大小为60000，这里是先读出所有训练数据，为的是方便作后续的转白底黑字图的操作。

读出测试数据部分：

与读训练数据同理。

1. 处理成批处理数据应用于后面的模型

定义一个**reader\_creator**(image\_array,label\_array)其中有一个带有yield的reader（）函数，返回一个关于image的数组和此image数组第一个元素图片对应的label生成器，用于后面训练数据的提供器，每次从缓存中随机读取批次大小的数据。





使用paddle.reader.shuffle()修改数据集大小，第二个参数是需要数据集的大小，每次从**reader\_creator**中随机读取512大小的数据集，将指定大小为512的数据集按照batch大小为128打包成批处理数据，每批处理128条数据.这样有利于提高算法的执行效率，以小单位来处理。

1. 网络结构
2. 共使用了两种网络模型——lenet和VGG
3. **神经网络模型-lenet**

**详细介绍所采用的网络结构，给出网络结构图，说明**

1. 每个卷积层所采用的滤波器个数、卷积尺寸、滤波器深度、激活函数

答：共有5个卷积-池化层。

第1个卷积-池化层的滤波器个数：20，卷积尺寸：5，滤波器深度：20，激活函数：Relu；

第2个卷积-池化层的滤波器个数：50，卷积尺寸：5，滤波器深度：50，激活函数：Relu；

第3个卷积-池化层的滤波器个数：50，卷积尺寸：5，滤波器深度：50，激活函数：Relu；

第4个卷积-池化层的滤波器个数：50，卷积尺寸：5，滤波器深度：50，激活函数：Relu；

1. 池化层的个数、位置、池化窗口的尺寸及深度

个数：5个；

位置：在5个卷积层后；

尺寸：2；

深度：第一个池化层：20，第二个池化层：50，第三个池化层:50,第四个池化层：50.

1. 全连接层的个数、每个全连接层的滤波器个数、尺寸、深度

个数：1个；

尺寸：10

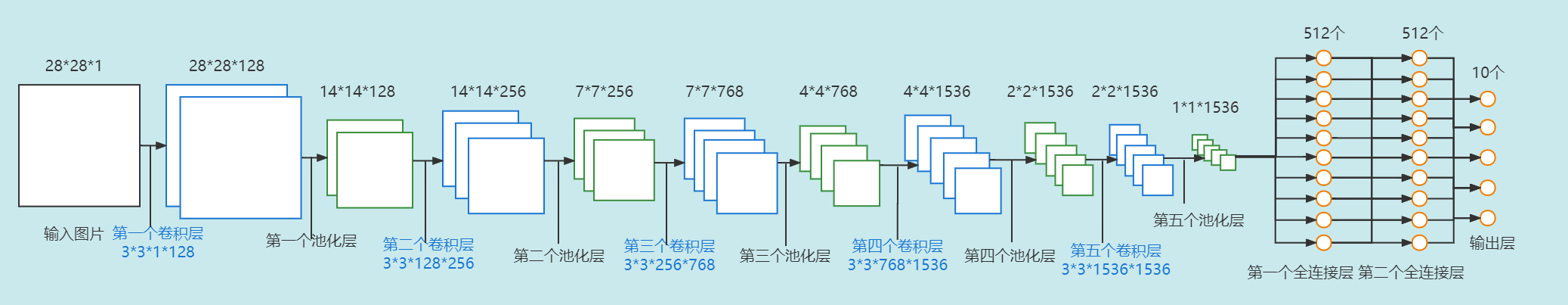
1. 输出层的神经元个数、激活函数

神经元个数：10

激活函数：softmax

（2）**神经网络模型-VGG**

**详细介绍所采用的网络结构，给出网络结构图，说明：**

****

1. 每个卷积层所采用的滤波器个数、卷积尺寸、滤波器深度、激活函数

VGG一共使用了5个卷积层

第1个卷积层：滤波器64\*2=128个，卷积尺寸：3\*3，滤波器深度：1，激活函数：Relu，输出大小：28\*28\*128

第2个卷积层：滤波器128\*2=256个，卷积尺寸：3\*3，滤波器深度：128，激活函数：Relu，输出大小：14\*14\*256

第3个卷积层：滤波器256\*3=768个，卷积尺寸：3\*3，滤波器深度：256，激活函数：Relu，输出大小：7\*7\*768

第4个卷积层：滤波器512\*3=1536个，卷积尺寸：3\*3，滤波器深度：768，激活函数：Relu，输出大小：4\*4\*1536

第5个卷积层：滤波器512\*3=1536个，卷积尺寸：3\*3，滤波器深度：1536，激活函数：Relu，输出大小：2\*2\*1536

1. 池化层的个数、位置、池化窗口的尺寸及深度

VGG池化层个数：5

位置：在每个卷积层后

池化窗口的尺寸：2\*2

深度：第1个池化层：128；第2个池化层：256；第3个池化层：768；第4个池化层：1536；第5个池化层：1536

第1个池化层输出大小：14\*14\*128

第2个池化层输出大小：7\*7\*256

第3个池化层输出大小：4\*4\*768

第4个池化层输出大小：2\*2\*1536

第5个池化层输出大小：1\*1\*1536

池化类型：max

1. 全连接层的个数、每个全连接层的滤波器个数、尺寸、深度

VGG全连接层的个数：2个

第1个全连接层的滤波器个数：1536\*512，尺寸：1\*1，深度：512

第2个全连接层的滤波器个数：512\*512，尺寸：1\*1，深度：512

4）输出层的神经元个数、激活函数

神经元个数：10

激活函数：softmax

**（3）简述选择该网络结构的理由。**

VGG通过堆叠2个3\*3的卷积核来代替1个5\*5的卷积核，通过堆叠3个3\*3的卷积核来代替1个7\*7的卷积核，结构非常简洁，整个网络都使用了同样大小的卷积核尺寸（3x3）和最大池化尺寸（2x2）。

**（4）说明尝试了几种网络模型，分别是什么模型。**

尝试使用了两种网络模型，Lenet和VGG神经网络模型。

1）LeNet

1998年的诞生的LeNet(LeCunetal.)可谓是现在各种卷积神经网络的始祖了，其网络结构虽然只有5层，却包含了卷积神经网络的基本组件(卷积层，池化展，全连接层)。

使用的问题：

①训练难。图像的数据由一个个的像素组成，如果将每个像素都输入到一个的单独的神经元中，那神经网络的尺寸，权值参数将是非常的大的。这么多的参数，需要消耗非常多的计算资源，以及需要非常大的数据集。

②无法应对图像内容位置的变化。对干手写体的数字图像，其每个字符的大小，倾斜度，书写的位置都会导致特征位置的变化。如果将

每个像素输入到一个单独的神经元中，则需要大量的不同位置的训练数据输入到神经元中。

③忽略了图像的局部相关性。图像的相邻像素有很强的相关性。单独一个像素输入到神经元中则丢失了这种相关性。

2）VGG

优点：

①VGG的结构非常简洁，整个网络都使用了同样大小的卷积核尺寸（3x3）和最大池化尺寸（2x2）。

②几个小滤波器（3x3）卷积层的组合相当于一个大滤波器（5x5或7x7）

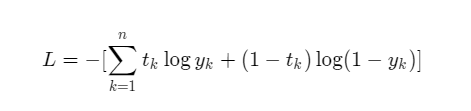
③卷积层好：验证了通过不断加深网络结构可以提升性能。

缺点：

VGG耗费更多计算资源，并且使用了更多的参数（大部分的参数集中在全连接层中），导致更多的内存占用。

1. 损失函数及学习率
2. 给出网络所使用的损失函数公式，说明选择该损失函数的理由，说明反向传播是如何作用在权重上的（可以简写或不写）以及其他相关的内容。

交叉熵的公式为：

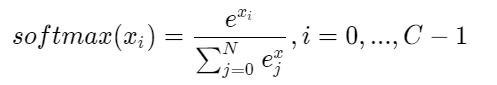


其中，log表示以e为底数的自然对数。yk​代表模型输出，tk​代表各个标签。tk中只有正确解的标签为1，其余均为0（one-hot表示）。

因此，交叉熵只计算对应着“正确解”标签的输出的自然对数。比如，假设正确标签的索引是“2”，与之对应的神经网络的输出是0.6，则交叉熵误差是−log0.6=0.51；若“2”对应的输出是0.1，则交叉熵误差为−log0.1=2.30。由此可见，交叉熵误差的值是由正确标签所对应的输出结果决定的。

对于分类问题，真实结果是标签，而模型输出是实数值，如果模型能输出十个标签的概率，对应真实标签的概率输出尽可能接近100%，而其他标签的概率输出尽可能接近0%，且所有输出概率之和为1。与此对应，真实的标签值可以转变成一个10维度的one-hot向量，在对应数字的位置上为1，其余位置为0，比如标签“6”可以转变成[0,0,0,0,0,1,0,0,0,0]。

为了实现上述假设，需要引入Softmax函数。它可以将原始输出转变成对应标签的概率，公式如下：



C是标签类别个数。 从公式的形式可见，每个输出的范围均在0~1之间，且所有输出之和等于1，这是这种变换后可被解释成概率的基本前提。对应到代码上，我们需要在网络定义部分修改输出层：self.fc = FC(name\_scope, size=10, act='softmax')，即是对全连接层FC的输出加一个softmax运算。

手写数字识别是分类任务，使用均方误差作为分类任务的损失函数存在逻辑和效果上的缺欠。手写数字识别的输出只可能是0-9之间的10个整数，相当于一种标签。

1. 说明尝试了几种损失函数，分别是什么函数。

均方误差（是在房价预测模型中使用的损失函数，适用于线性回归模型，实数输出与标签相减，而在分类问题中不太合适）

SoftMax函数（所有输出和为1。网络结构变化：单一输出（实数值）->每个分类一个输出（概率））

交叉熵（在模型输出为分类标签的概率时，直接以标签和概率做比较也不够合理，人们更习惯使用交叉熵误差作为分类问题的损失衡量。基于最大似然思想：最大概率得到观察结果的假设是真的）

1. 学习率是多少？简述确定该学习率的过程。

学习率为0.005

我们无法分析地计算给定数据集上给定模型的最佳学习速率。 相反，必须通过反复试验发现良好（或足够好）的学习率。要考虑的学习率的值范围小于1.0且大于10 ^ -6。

如果学习率太小，神经网络可能收敛过慢，过快的话，又会找不到最小值，甚至导致发散。因此过快或者过慢的学习率都是不好的。

1. 优化器

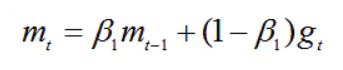
（1）介绍所采用的优化器的原理、特点，及选择它的原因。

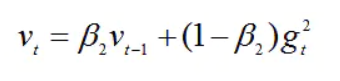
（2）说明尝试了几种优化器，分别是什么优化器。

1. Adam优化器。

原理：作为梯度下降的优化方法之一，对梯度的一阶矩估计和二阶矩估计进行综合考虑，计算出更新的步长。过程为：

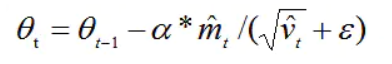
首先计算t时刻的梯度；

然后计算梯度的指数移动平均数：*β1* 默认取0.9（或一个接近1的数字，此处取0.9）；

接着计算梯度平方的指数移动平均数：*β2*默认取0.999；

由于m0初始化为0，因此会导致mt偏向0（尤其是训练初期），因此对mt进行偏差纠正：；

对vt也需要纠正：；

总公式为：，其中α为学习率，*ε=*10的-8次方。

特点：适用于梯度存在很大噪声的情况，适合于大规模数据集。

测试数据：

Test:0, Cost:0.09224, Accuracy:0.97248

Test:1, Cost:0.08534, Accuracy:0.97781

Test:2, Cost:0.03183, Accuracy:0.99074

Test:3, Cost:0.03365, Accuracy:0.99005

Test:4, Cost:0.04285, Accuracy:0.98885

Test:5, Cost:0.03997, Accuracy:0.98960

Test:6, Cost:0.03551, Accuracy:0.98980

Test:7, Cost:0.04199, Accuracy:0.99030

Test:8, Cost:0.03217, Accuracy:0.99179

Test:9, Cost:0.04600, Accuracy:0.98980



1. SGD优化器。

原理：在得到梯度的基础上，得到输出参数=输入参数-学习率\*梯度。

特点：实现简单，适合大规模数据，比较快。

测试数据：

Test:0, Cost:0.06184, Accuracy:0.97940

Test:1, Cost:0.04457, Accuracy:0.98507

Test:2, Cost:0.04364, Accuracy:0.98602

Test:3, Cost:0.03565, Accuracy:0.98875

Test:4, Cost:0.03359, Accuracy:0.98985

Test:5, Cost:0.03378, Accuracy:0.98975

Test:6, Cost:0.03241, Accuracy:0.99025

Test:7, Cost:0.03303, Accuracy:0.99074

Test:8, Cost:0.03316, Accuracy:0.99045

Test:9, Cost:0.03251, Accuracy:0.99040



1. Adagrad

原理：计算梯度指数移动平均数；

随后，输出参数为：；

特点：他能对每个不同的参数调整不同的学习率，对频繁变化的参数以更小步长进行更新，而稀疏的参数以更大的步长进行更新。更适合有分布稀疏的场景。

测试数据为：

Test:0, Cost:0.06400, Accuracy:0.97920

Test:1, Cost:0.03215, Accuracy:0.98980

Test:2, Cost:0.02421, Accuracy:0.99169

Test:3, Cost:0.10416, Accuracy:0.97223

Test:4, Cost:0.03637, Accuracy:0.98895

Test:5, Cost:0.03253, Accuracy:0.99084

Test:6, Cost:0.04779, Accuracy:0.98761

Test:7, Cost:0.03042, Accuracy:0.99144

Test:8, Cost:0.02986, Accuracy:0.99189

Test:9, Cost:0.04259, Accuracy:0.98890



对比以上三种优化器，发现他们测试准确性几乎持平。在运行时间上，Adam和Adagrad更短，但Adagrad更简单，计算量也更少，因此选择Adagrad优化器。

1. 实验结果
2. 用表格或图的形式，对比各种尝试（使用不同的网络模型、损失函数、优化器）的实验结果（recall，precision，accuracy，error-rate）；
3. 分析、解释实验过程中出现的一些现象。
4. 总结与收获（必写）

谈谈学习这门课的心得体会。每个小组成员分别写自己的心得（不得少于100字），汇总于此。

学生1 柳萱莹：

通过这学期的学习，对人工智能有了更进一步的了解和认识。

个人觉得研究人工智能的目的，一方面是要创造出具有智能的机器，另方面是要弄清人类智能的本质，因此，人工智能既属于工程的范畴，又属于科学的范畴。通过研究和开发人工智能，可以辅助，可以辅助，部分替代甚至拓宽人类的智能，使计算机更好的造福人类。人工智能研究的近期目标是使现有的计算机不仅能做般的数值计算及非数值信息的数据处理，而且能运用知识处理问题，能模拟人类的部分智能行为。按照这一目标，根据先行的计算机的特点研究实现智能的的有关理论、技术和方法，建立相应的智能系统。例如目前研究开发的专家系统，机器翻译系统、机器人等。随着社会的发展，技术的进步，人工智能的发展是任何人都无法想象的。

在这门课后，将会自主的对人工智能进行更多的探索和学习，感谢这门课程为我打开了人工智能的窗户，让我有幻想和愿望去探索更加丰富的人工智能世界。

学生2姓名：心得。。。。

学生3姓名：王诗涵

通过本学期的学习，我了解了人工智能的概念及其发展历史，学习了知识表示、各种搜索策略以及机器学习、神经网络的相关知识，使我对人工智能产生了浓厚的兴趣。在做大作业——手写数字体识别的过程中，感受到了人工智能的神奇之处。在新一轮科技革命和产业变革的浪潮中，人工智能从感知和认知两方面模拟人类智慧，赋予机器学习以及推断能力。发展至今，人工智能已经成为新一轮科技革命和产业变革的核心驱动力，正在对世界经济、社会进步和人民生活产生极其深刻的影响。在未来的发展中，人工智能必将占据重要地位，带给人们生活无限的可能性。

这门课使我懂得了人工智能的重要性，是我接触人工智能的启蒙课程，让我较为全面的了解到相关知识。我在以后的学习中也会更加关注人工智能领域的发展，希望学习到更深层次的知识，能够在这个领域有所贡献。

学生4姓名：李昊晴：

人工智能基础这门课，让我对于人工智能的概念不再只是停留于已经研发出的人工智能产品，而是可以深入的了解人工智能背后的原理。现在的人工智能对于我们来说，并不是遥不可及，其背后也是由算法原理支撑的。到现在为止，AI已经出现过三次热潮，一是图灵测试，二是语言识别，三是深度学习与大数据。目前，我们正处于第三次热潮。学习与应用人工智能的知识，才可以抓住与把握这次热潮，使得人类社会进入一个更加便捷、多元、发展的新时代。

人工智能的未来是无限的，也是充满创造力的。我也将坚定地学习与探索人工智能领域的未知与精彩。

学生5姓名：周昱辰

通过这段时间的学习，我对人工智能有了一定程度的认识。人工智能是经过多年发展起来的一门综合性学科，它旨在研究如何利用计算机等现代工具模拟人类智能行为的系统。人工智能，最早是由著名计算机科学家图灵在20世纪50年代提出的，就是著名的“图灵测试”。最近几年，随着深度学习发展，人工智能被运用在各行各业，因此有人把人工智能称为第四次科技革命，他将给人们的生活带来翻天覆地的变化。随着社会的发展，技术的进步，人工智能的发展是任何人都无法想象的，在今后的学习生活中我也会主动学习相关知识，也定将对我们今后的知识水平的进一步提升起到较大帮助。