中国地质大学(武汉)

机器学习第一次作业

姓 名：叶宇涛

专 业：计算机科学与技术

学 号：20191000595

指导老师：刘超

目录

[中国地质大学(武汉) 1](#_Toc100503128)

[机器学习第一次作业 1](#_Toc100503129)

[5 神经网络 3](#_Toc100503130)

[5.5 BP算法 3](#_Toc100503131)

[5.5.1 编程实现标准BP算法，在西瓜数据集3.0上训练一个单隐层网络，并做数据分析和结果评价。 3](#_Toc100503132)

[5.5.2 编程实现累积BP算法，在西瓜数据集3.0上训练一个单隐层网络，做数据分析和结果评价，并和标准BP算法进行比较 3](#_Toc100503133)

# 神经网络

## BP算法

### 编程实现标准BP算法，在西瓜数据集3.0上训练一个单隐层网络，并做数据分析和结果评价。

#### 编程题目理解

题目中要求我们利用书上面的伪代码以及西瓜数据集3.0，实现一个标准BP算法，其中，只包含一个隐层结构，并且为全连接前馈网络。根据题目要求，需要设计1个输入层，1个隐层，1个输出层，其中，节点的个数任选，再利用书上面的步骤，对连接权和阈值进行更新。最后，利用测试集测试精度。

#### 标准BP算法原理阐述

标准BP算法是一种经典的算法，把学习过程分为信号的正向传播和反向传播。该模型的训练经常采用误差反向传播算法，来更新神经元的连接权和阈值。

正向传播的时候，输入样本从输入层进入网络，经过隐层到输出层输出。而反向传播的的过程，就是计算出网络的均方误差后，求均方误差的累计平均值最小值的过程。其中，均方误差公式为

我们的目标，就是将优化上面的，直到收敛到一定精度。其中，标准BP算法在一次迭代的时候，分别带入不同的样本进行训练，因此，需要训练的连接权和阈值次数会非常多，其中，可能会有不同的样本相互抵消，导致训练权和阈值没有变化。这就演化出了累计BP算法。

具体的迭代算法，可以根据书上面公式推导得出。其中，主要的公式如下：

上面的公式中，最主要的是、分别代表隐层的输入，输出层的输入，代表输出层的最后输出，给出函数为sigmoid函数。输入层到隐层连接权、隐层到输出层连接权分别为、，阈值分别为、。这四个参数是需要主要更新的参数，而是用来简化计算的参数，也就是梯度下降中的中间量。根据这些公式，可以写出算法的主要设计思路。

#### 算法设计思路

根据上面的原理，可以设计出算法的思路如下：

1. 导入西瓜数据集，处理数据，将类别转化为数字编码、将特征转化为数字编码。
2. 设计类Data，用来初始化参数：迭代次数、训练集、测试集的划分、连接权和阈值。
3. 进入标准BP算法的训练步骤。
4. 随机初始化两个连接权矩阵和阈值矩阵，学习率设定为0.1.
5. 当迭代次数小于指定的迭代次数，进行如下循环：
6. 对训练集的每一个样本，进行如下循环：
7. 计算，计算输出层的预估输出，计算出此时的均方误差，添加到中，根据正向传播的参数，更新反向传播参数：连接权、阈值、梯度下降用到的两个中间量。
8. 如果样本没有用完，返回第六步。
9. 如果迭代次数小于指定的迭代次数，返回第五步。
10. 将训练完成的神经网络参数记录，将迭代次数作为横坐标，作为纵坐标，画出收敛图。
11. 将测试集输入至神经网络中，如果最后的预测值大于0.5，表示好瓜，小于0.5表示坏瓜。将其与真正的测试结果进行对比，计算精度。

#### 实验流程、测试结果及分析

实验流程如上面算法设计思路一般进行。其中，各类超参数设置不同会导致不同的结果，一般来说，隐层节点设置为7个为最佳。在这里，输入层节点数设置为8，也就是西瓜数据集3.0的特征数量，而输出层指定为1，隐层指定为9，也就是特征数量+1。最后的测试结果如下：

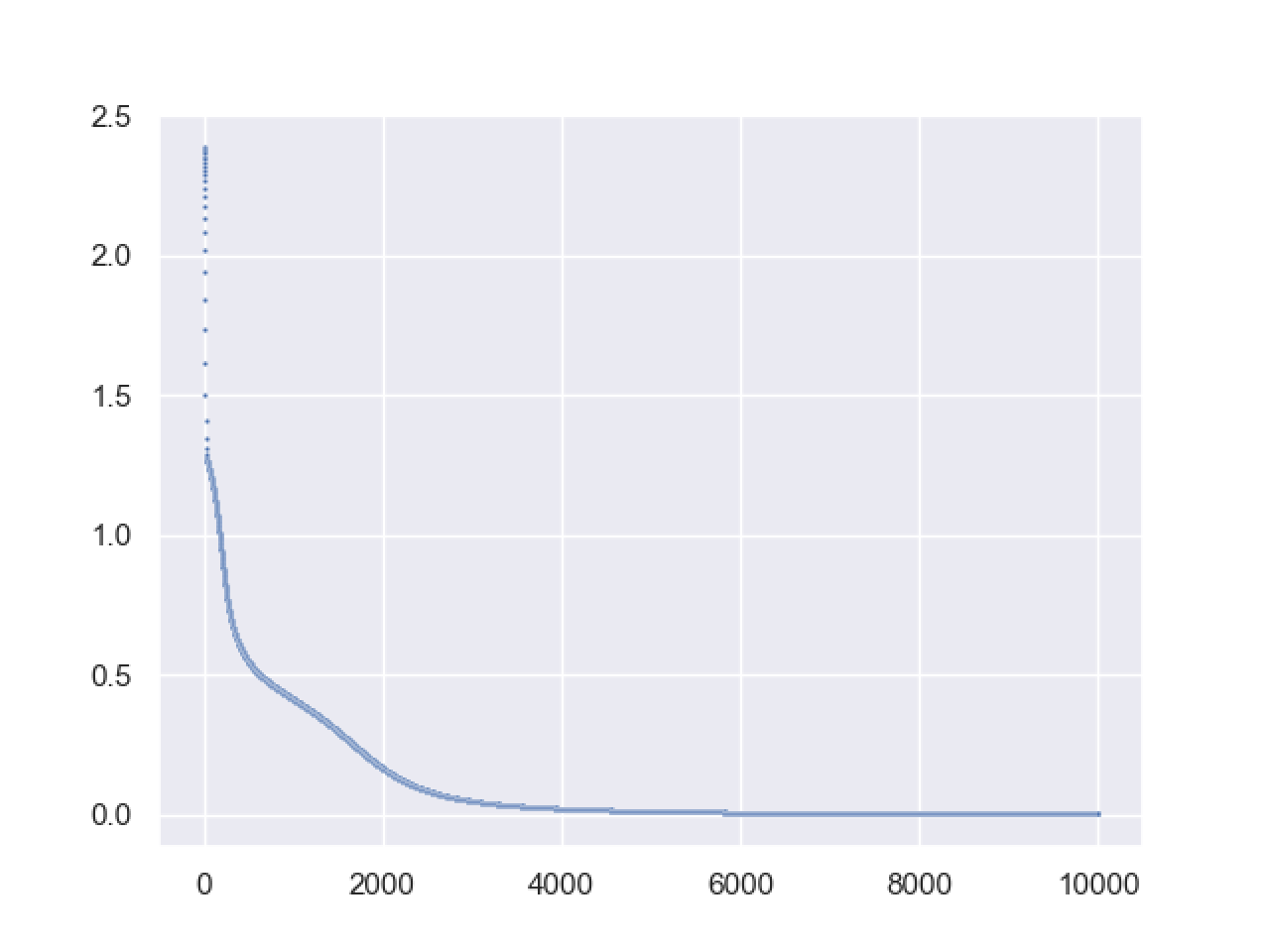


图 ‑1 标准BP算法迭代10000次累计误差变化

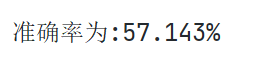


图 ‑2 精确率

**数据分析及结果分析：**

由于西瓜数据集的样本数量太少，对其训练出的曲线看不出来是过拟合还是欠拟合。但是标准BP算法一般会导致过拟合，所以会产生早停和正则两种修正方法。准确率较低的原因是对于训练集的拟合效果太好，也就是过拟合，并且样本数量太少，无法真正体现出神经网络的强大表达能力。并且，测试集的数据也不是特别多，随机性比较大，该结果是正常表现。

#### 代码结构，核心代码简要分析

**代码结构:**

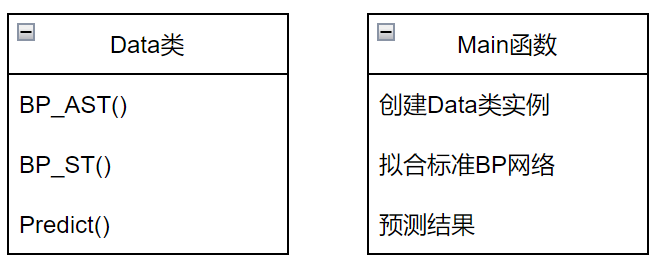


图 ‑3 标准BP算法代码结构

核心代码分析及注释：



图 ‑4 标准BP算法核心代码及注释

#### 本次实验解决的主要问题，主要收获

本次实验主要遇到的问题是对于numpy库的使用不够熟练，矩阵的乘除需要查阅相关资料才能实现。并且，对于矩阵的维度不是很熟悉，需要反复调试。对于公式的实现能力还有待加强。

本次实验还让我深入了解了标准BP算法的实现，可以从零开始实现一个标准BP算法，了解其中的原理，清楚了梯度下降算法的具体推导步骤和实现，清楚了阈值和对应激活函数的关系，以及神经网络强大的表达能力，过拟合该如何修正等等。了解了BP算法的局限性，例如梯度消失等等问题。

#### 编码及内容撰写中的参考来源

1. <https://blog.csdn.net/weixin_42398658/article/details/83929474>，深度学习 --- BP算法详解（流程图、BP主要功能、BP算法的局限性）
2. 机器学习 周志华 P101-P105

### 编程实现累积BP算法，在西瓜数据集3.0上训练一个单隐层网络，做数据分析和结果评价，并和标准BP算法进行比较

#### 编程题目理解

累积BP算法和标准BP算法差不多的算法，题目要求我们利用累积BP算法对西瓜数据集3.0进行神经网络的训练，其中，节点的个数任选，再利用书上面的步骤，对连接权和阈值进行更新。最后，利用测试集测试精度。

#### 累积BP算法原理阐述

和标准BP算法原理相似，都是对神经网络的反向传播过程中的参数进行训练，将其中的均方误差最小化。但是和标准BP算法不同的是，累积BP算法输入的不再是一个个样本，而是对训练集扫描一次后，将训练集整个输入进去。因此，标准BP算法的参数更新效率会十分频繁，时间复杂度达到了，而累积BP算法的参数更新频率少，时间复杂度为。累积BP算法直接对累积误差最小化，标准BP算法可能会因为不同样例的更新，导致参数相互“抵消”的情况。

但是在很多情况下，累积BP算法的累计误差下降到一定程度后，进一步的下降会非常慢，这时候标准BP算法会更快获得更好的解，在训练集非常大的时候尤为明显。

**原理和标准BP算法相同：**

累积误差公式为

我们的目标，就是将优化上面的，直到收敛到一定精度。累积BP算法的更新就更为简单，具体的迭代算法，可以根据书上面公式推导得出。其中，主要的公式如下：

上面的公式中，最主要的是、分别代表隐层的输入，输出层的输入，代表输出层的最后输出，给出函数为sigmoid函数。输入层到隐层连接权、隐层到输出层连接权分别为、，阈值分别为、。这四个参数是需要主要更新的参数，而是用来简化计算的参数，也就是梯度下降中的中间量。根据这些公式，可以写出算法的主要设计思路。

#### 算法设计思路

#### 实验流程、测试结果及分析

#### 代码结构，核心代码简要分析

#### 本次实验解决的主要问题，主要收获

#### 编码及内容撰写中的参考来源