|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **学号** | | 21820236 | | **姓名** | | 于烨泳 | | |
| **评分项** | | | | | | | | |
| 充分性（20%） | 准确性（20%） | | 逻辑性（20%） | | 深刻性（20%） | | 规范性（20%） | **总分** |
|  |  | |  | |  | |  |  |

**深入剖析反向传播算法**

1. **前馈神经网络（Feedforward Neural Network, FNN）**

从机器学习的角度来看，神经网络一般可以看作一个非线性模型，其基本组成单元为具有非线性激活函数的神经元，激活函数可以增强网络的表示能力和学习能力的，通过大量神经元之间的连接，使得神经网络成为一种高度非线性的模型。神经元之间的连接权重就是需要学习的参数，可以在机器学习的框架下通过梯度下降方法来进行学习[1]。

前馈神经网络（FNN）是一种极其重要的神经网络结构，其他网络结构如记忆网络，图网络的前馈思想是以FNN为基础的，且FNN是由多层的Logistic回归模型（连续的非线性函数）组成，而不是由多层的感知器（不连续的非线性函数）组成[2]，比起多层感知器，FNN具有更强大的表示学习能力。根据通用近似定理[3]，前馈神经网络具有很强的拟合能力，常见的连续非线性函数都可以用前馈神经网络来近似[1]。

在前馈神经网络中，各神经元分别属于不同的层。每一层的神经元可以接收前一层神经元的信号，并产生信号输出到下一层。第0层称为输入层，最后一层称为输出层，中间层称为隐藏层。整个网络中无反馈，信号从输入层向输出层单向传播，如图1所示。

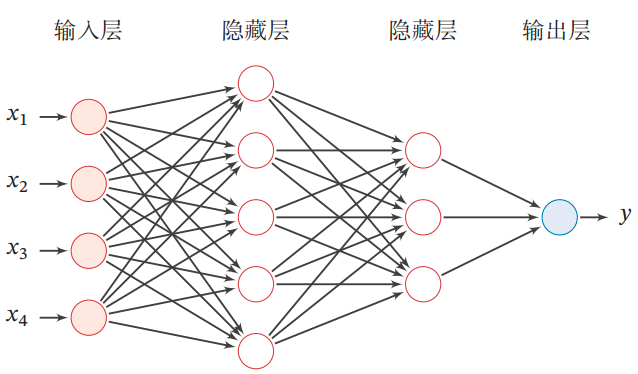


图 1 多层前馈神经网络[1]

上图中，令（**x**代表(x1,x2,x3,x4)向量），前馈神经网络通过公式(1)进行信息前向传播：

|  |  |
| --- | --- |
|  | (1) |

其中 代表 层的净活性值（未被激活）， 为激活函数， 则是经过 激活 的活性值，通过这种方式前馈神经网络逐层传递信息，信息传递到输出层就得到了网络的预测值 ，通过特定的损失函数可以计算出此次前向传播的损失（未考虑正则化项）：

|  |  |
| --- | --- |
|  | (2) |

在传统的机器学习任务中，目标是最小化迭代训练误差，网络参数和就可以通过梯度下降方法公式(3)，(4)（也有机器学习任务是梯度上升，将公式(3)，(4)中的“-”替换成“+”）来进行更新学习：

|  |  |
| --- | --- |
|  | (3) |
|  | (4) |

其中 是梯度上升最快的方向，乘以-1则可以得到梯度下降最快的方向，是学习率，两者相乘即可得到参数和参数更新的幅度，同时我们需要注意到是当前梯度下降最快的方向，并不一定是最优的方向，这就是导致模型陷入局部最优的问题根源。但是此梯度下降法需要计算损失函数对参数的偏导数，如果通过链式法逐一对每一层每一个参数进行求偏导，更新的效率非常低，所以在神经网络的参数训练过程中常使用反向传播算法（Backpropagation, BP）来更新参数。

1. **反向传播算法（Backpropagation, BP）**

梯度下降法是一种通用的优化算法，中心思想是沿着目标函数梯度的方向更新参数值以希望达到目标函数最小（或最大），反向传播法是梯度下降法在深度网络上的具体实现方式。由于深度学习网络按层深入，层层嵌套的特点，且现有网络结构大多为多输入少输出，所以对深度网络目标函数计算梯度的时候，使用用反向传播的方式由深到浅倒着计算以及更新参数将会更高效。

BP算法将和梯度通过链式法则公式(5)，(6)分解：

|  |  |
| --- | --- |
|  | (5) |
|  | (6) |

其中是 层净活性值，根据公式 计算得到，所以可以根据上一层的活性值计算得出，而因为系数为*I*，所以求偏导结果就是单位矩阵，是第 层神经元的个数，具体计算见公式(7)，(8)：

|  |  |
| --- | --- |
|  | (7) |
|  | (8) |

因为 层的第i个神经元的净活性值是由 层所有神经元的活性值经过仿射变换得到的，为权重矩阵第i行，即第 层中第i个神经元中的所有权重，其中的每一个权重只与 层第j个神经元的激活值相乘，其余的偏导都为0，所以具体到每一层的参数的偏导上，得到的结果是，表示第i个元素为，其余都为0的行向量。

对比公式(5)，(6)，可以发现所需计算的仅剩，代表第 层神经元对最终损失的影响，也反映了最终损失对第 层神经元的敏感程度[1]，[1]中定义其为误差项：

|  |  |
| --- | --- |
|  | (9) |

对比直接计算，误差项可以找到递归关系，即可以通过下一层的来直接计算，见公式(10)：

|  |  |
| --- | --- |
|  | (10) |

其中是向量的Hadamard积运算符，表示每个元素相乘，diag表示的对角线元素。反向传播的含义就是第 层的一个神经元的误差项是所有与该神经元相连的第 层的神经元的误差项的权重和再乘上该神经元激活函数的梯度[1]。简而言之就是可以根据下一层的计算得出，综合之前计算得到的和就可以得到参数更新的方向和，与学习率α相乘即可更新参数，无需逐层通过链式法则计算参数更新的方向，大大提升了参数的更新效率。

反向传播算法的思想我感觉与前馈神经网络有“异曲同工”之妙，不过FNN传播的是信息，BP传播的是误差，两者传播的信息都可以基于本层的信息直接传递，而不用依赖之前层（反向传播是之后层）的信息；且FNN前馈的和激活函数与BP算法的 和激活函数的导数都是镜像的逆操作；不过BP误差的传递还需要前馈时保留的每一层的激活值和净激活值。

1. **反向传播与动态规划**

**报告内容为：以教材内容为出发点进行深入阐述，并进行相关拓展**

**必须独立完成**。不得进行简单拷贝。**以拷贝为主的报告将被认为不合格**。

引用的图表和主要公式应标明文献出处。

**提交说明：**

1. 期末报告电子版以word格式提交，命名规则：学号\_姓名\_报告.doc，举例：20721234\_孙悟空\_报告.doc
2. 历次的课后作业电子版汇总成单文档提交
3. 以上两个文档的电子版分开在超星学习通提交，提交截止时间为11月25日晚24点之前
4. 其中报告的纸质版，A4双面打印，11月24和11月25日两天，上午9-11点，下午13-16点，提交到计算机楼904实验室，盛鑫宇 助教处，逾期成绩不计。
   1. **文献清单**

列出所阅读文献的清单。给出完整信息：作者，题目，期刊，卷期，页码，年份，会议名称，会议地点，网址，学位论文出处，技术报告来源，专利号等。

**在报告正文中标明文献引用情况**。

* 1. **评分依据**

充分性（20%），内容是否充实

准确性（20%），内容是否准确无误

逻辑性（20%），阐述是否有条理，逻辑清晰

深刻性（20%），问题分析得是否深入，不浮于表面

规范性（20%），是否按模板要求撰写，格式是否符合规范

**参考文献**

1. 邱锡鹏，神经网络与深度学习，机械工业出版社，https://nndl.github.io/, 2020.
2. Bishop C M, 2007. Pattern recognition and machine learning[M]. 5th edition. Springer
3. Hornik K, Stinchcombe M, White H, 1989. Multilayer feedforward networks are universal approximators[J]. Neural networks, 2(5):359-366.