# 模式识别第4次作业

周强 202128019427002 电子学院

1. **Consider a three-layer network for classification with nodes in hidden layer, and nodes in output layer. The patterns (also say samples) are in dimensional space. The activation function (or transfer function) for the nodes in the hidden layer is the sigmoid function. Differently, the nodes in the output layer will employ the following softmax operation as their activation function:**

**where stands for the weighted sum at the -th node in the output layer.**

* 1. **Derive the learning rule under the back propagation framework if the criterion function for each sample is the sum of the squared errors, that is （即分析每一层权重的更新方法）:**

**Where is the known target value for the sample at the -th node in the output layer.**

答：先计算隐层到输出层的权重的更新量。记为隐层的第个节点到输出层第个节点的权重，其中。

其中

计算输入层到隐层的权重更新量。记为输入层的第个节点到隐层第个节点的权重，其中.

其中

总结：参数的更新量取决于学习率和参数连接的节点，为边的指向节点收集到的误差的加权和经过导数缩放，乘以边的起始节点的输入再乘以学习率。

* 1. **结合课堂所学知识, 对反向传播算法进行总结。  
     注意: 本题只需要推导出单个样本对权重更新的贡献即可 (因为多个样本只是简单 地相加）**

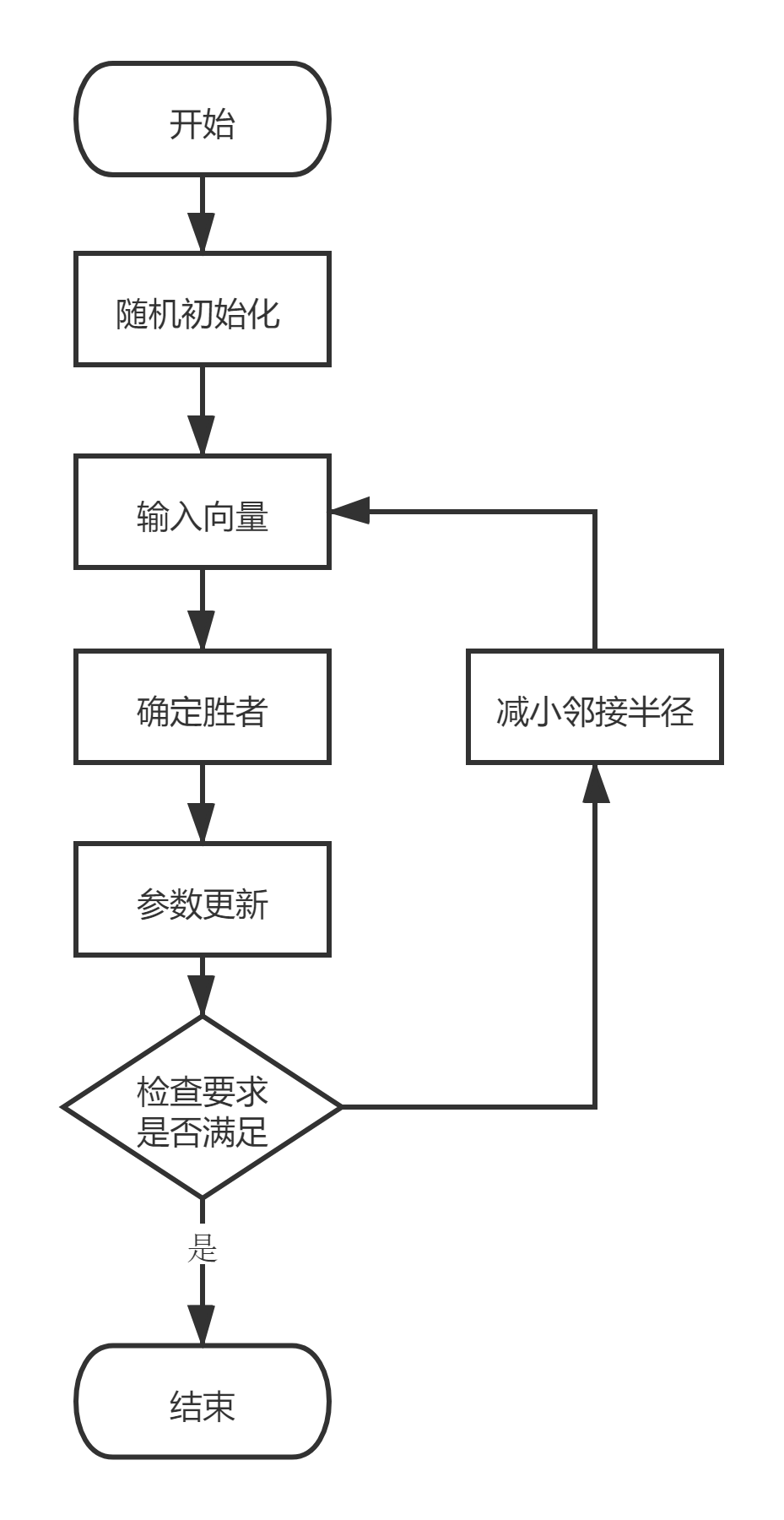
答：反向传播的过程如下：

1. 前向传播：将训练数据输入神经网络，从前向后计算每一层神经网络的输出。
2. 误差收集：计算神经网络的输出与给定目标值之间的误差。
3. 反向传播：从后向前依次计算各层权重的更新量。
4. 权重更新：根据上一步的更新量，计算新的权重。
5. 重复上面1-4步，直到权重收敛或者达到最大迭代次数。
6. **请描述自组织算法的计算步骤，给出训练算法的框图。**

答：自组织算法是一种用于聚类的无监督算法，这意味着该方法无需任何有标签的数据即可自动将具有相似特征的样本归为一类。其基本步骤如下：

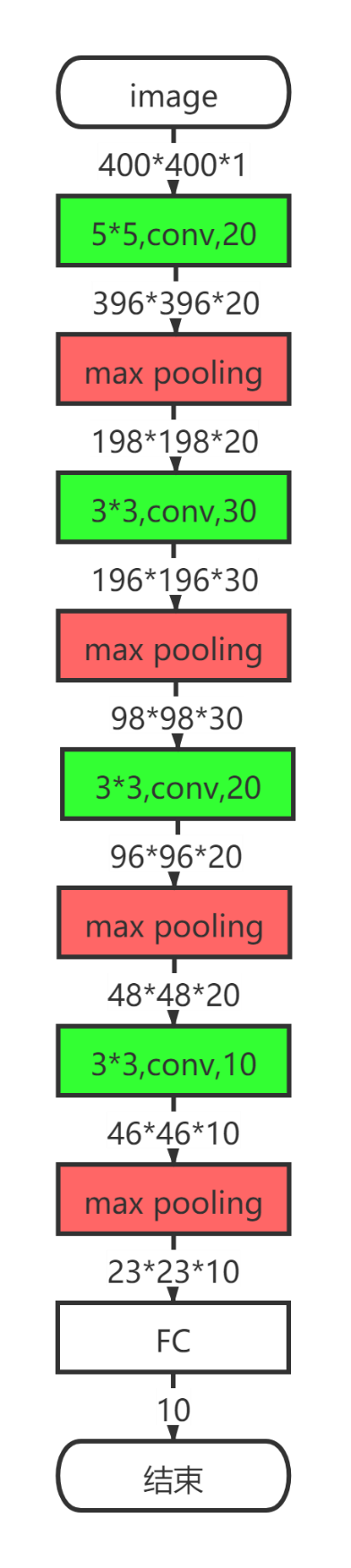
1. 随机初始化网络
2. 输入向量：计算输入向量与映射层权重的欧式距离。
3. 确定胜者：选择输入向量与权重向量距离最小的输出节点为胜者，并给出其邻接神经元集合。
4. 参数更新：更新胜者神经元及其邻接神经元的权重。
5. 检查要求：检查是否达到预先设定的要求。如果达到要求则算法结束；否则减小邻接半径，重复上述2-4步。

算法的流程图如下



1. **拟考虑对大小的图像数据集运用卷积神经网络。假定共有4个卷积层。第一隐含层采用大小的滤波器, 其它卷积层采用大小的滤波器。第1至第4卷积层分别包含的图像数目为 。Pooling操作均采用在大小的局部窗口内取最大值(即max pooling),同时假定在max pooling之后完成激励操作。假定连接到最后一个pooling操作之后的前向神经网络为单层前向神经网络,且输出层的结点个数为10。**
   1. **请按层指出该网络需要计算的权重数量; 相对于全连接和非权值共享, 请指出当 同时采用权值共享和局部连接时所减少的权重数量;**

答：绘制此神经网络的结构图如下



假设输入的的图像为单通道图像，则输入层到第1隐层的单个卷积核需要的参数量为,则20个卷积核的参数量为，加上20个偏置权重，则共需要。

假设不对图像进行填充，卷积核移动的步长为1，则第1隐层的节点个数为个，若使用全连接神经网络，输入层到第1隐层需要的权重个数为。

其他各层的计算类似，不再赘述计算过程，直接将结果列在下表中（这里忘记考虑池化对节点数量的减少，需要修改）。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 输入通道 | 卷积核尺寸 | 输出通道 | 参数数量 | 全连接参数数量 |
| 输入层到第1隐层 | 1 |  | 20 | 520 |  |
| 第1隐层到第2隐层 | 20 |  | 30 | 5430 |  |
| 第2隐层到第3隐层 | 30 |  | 20 | 5420 |  |
| 第3隐层到第4隐层 | 20 |  | 10 | 1810 |  |

* 1. 请指出在遇到max pooling操作时,在进行反向传播时错误如何传?

答：池化层中没有需要学习的参数，因此进行反向传播时无需进行参数更新，仅仅是将误差传递到前一层即可。

以的池化为例考察误差传递方式，设,则池化结果为

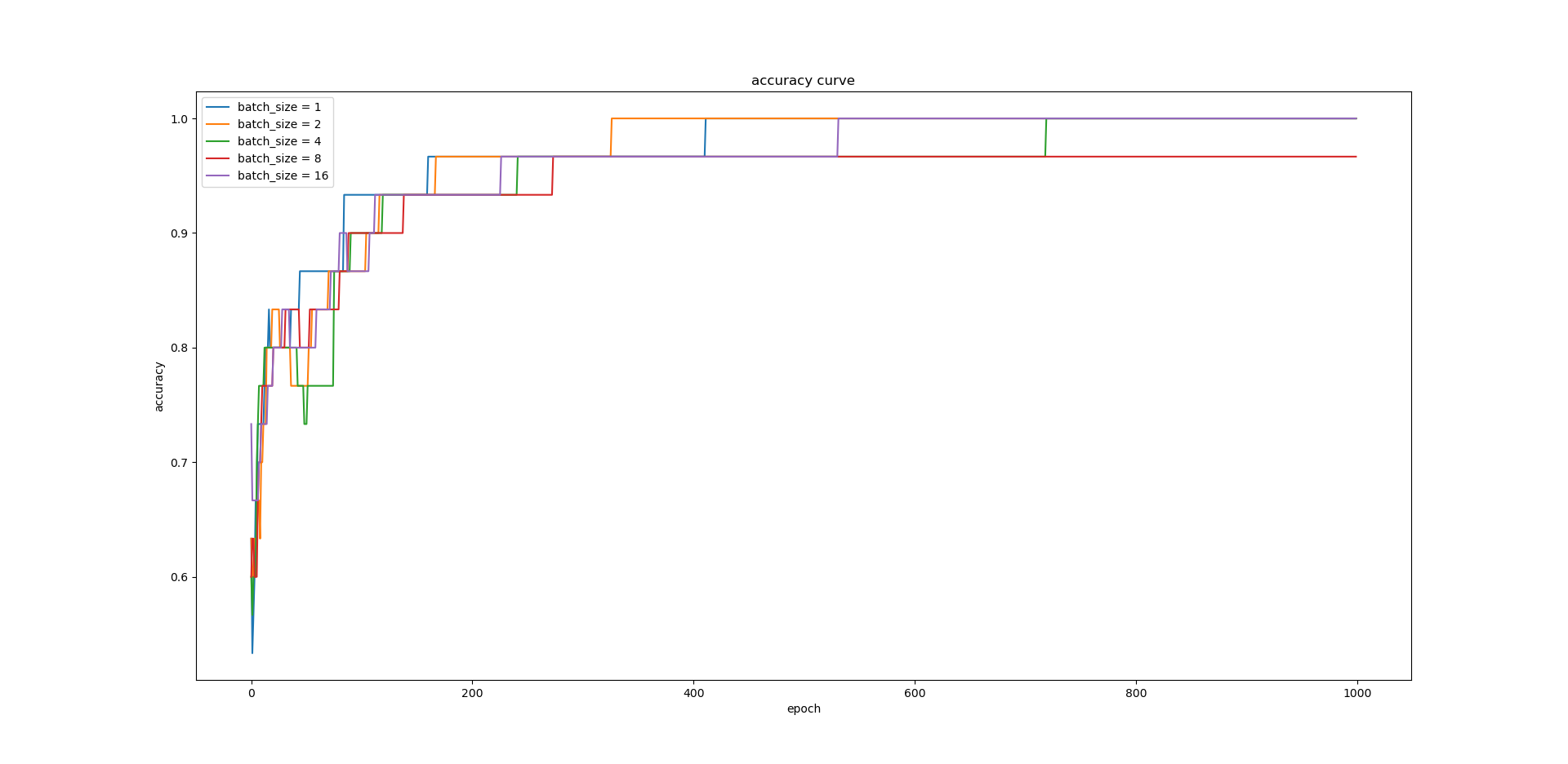
则下一层的误差会原封不动的传递到到上一层中输出最大的节点，别的节点都为0。

* 1. 请讨论你所能想到的对网络结构的改变。 (本题可在下一章讲完之后再做, 本次作业先布置)

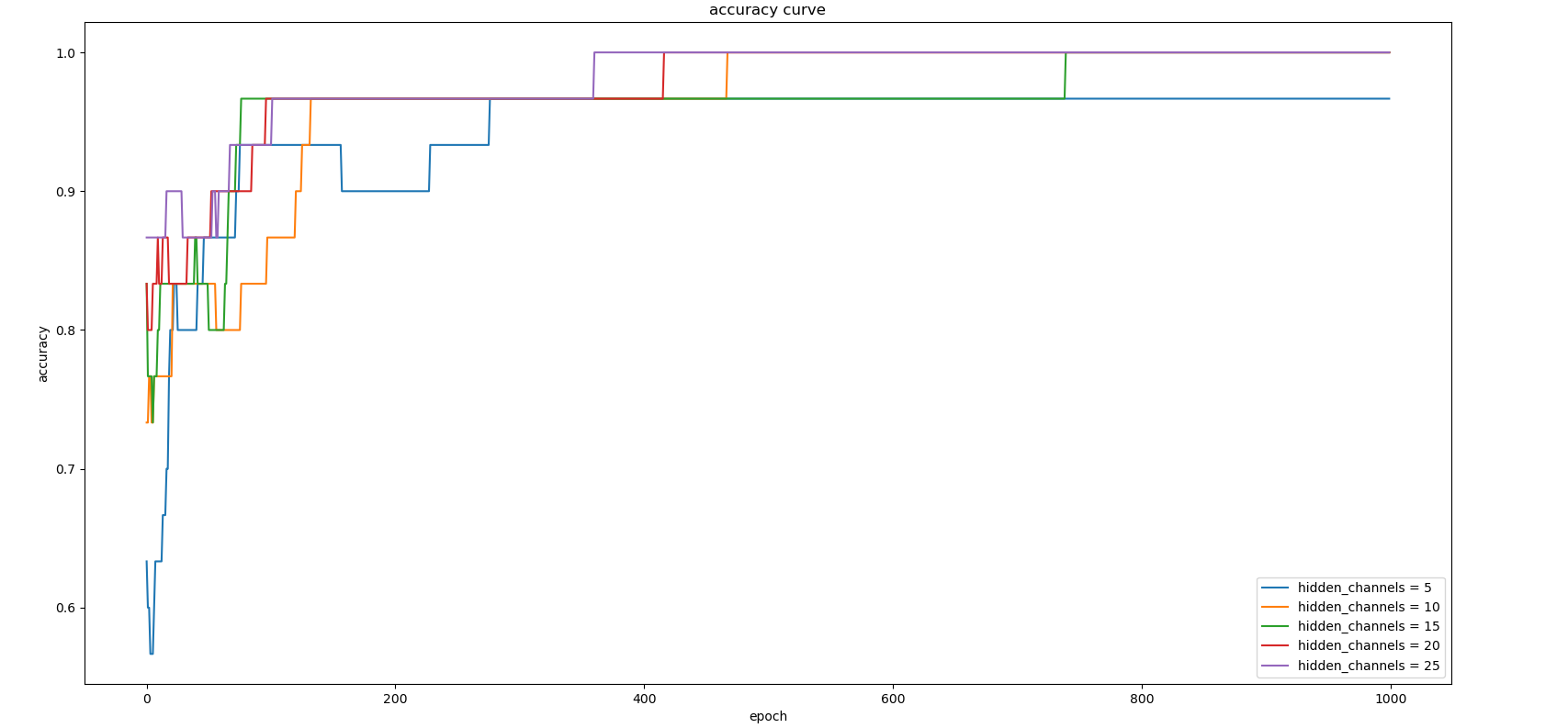
答：可能的网络结构的改变如下：

1. 采用不同的激活函数
2. 将最大值池化改为平均值池化
3. 修改各个隐层卷积核个数
4. 修改隐层个数
5. 编程题
   1. 采用批量方式更新权重和单样本更新权重的对比如下图所示。

单样本更新权重时，收敛速度较快，但是不稳定，因为单个样本的梯度和全体样本的梯度可能有较大偏差；批量更新权重时，收敛速度较慢，而且比较稳定。实验表明，二者的收敛结果基本一致。



* 1. 隐含层不同结点数目对训练精度的影响

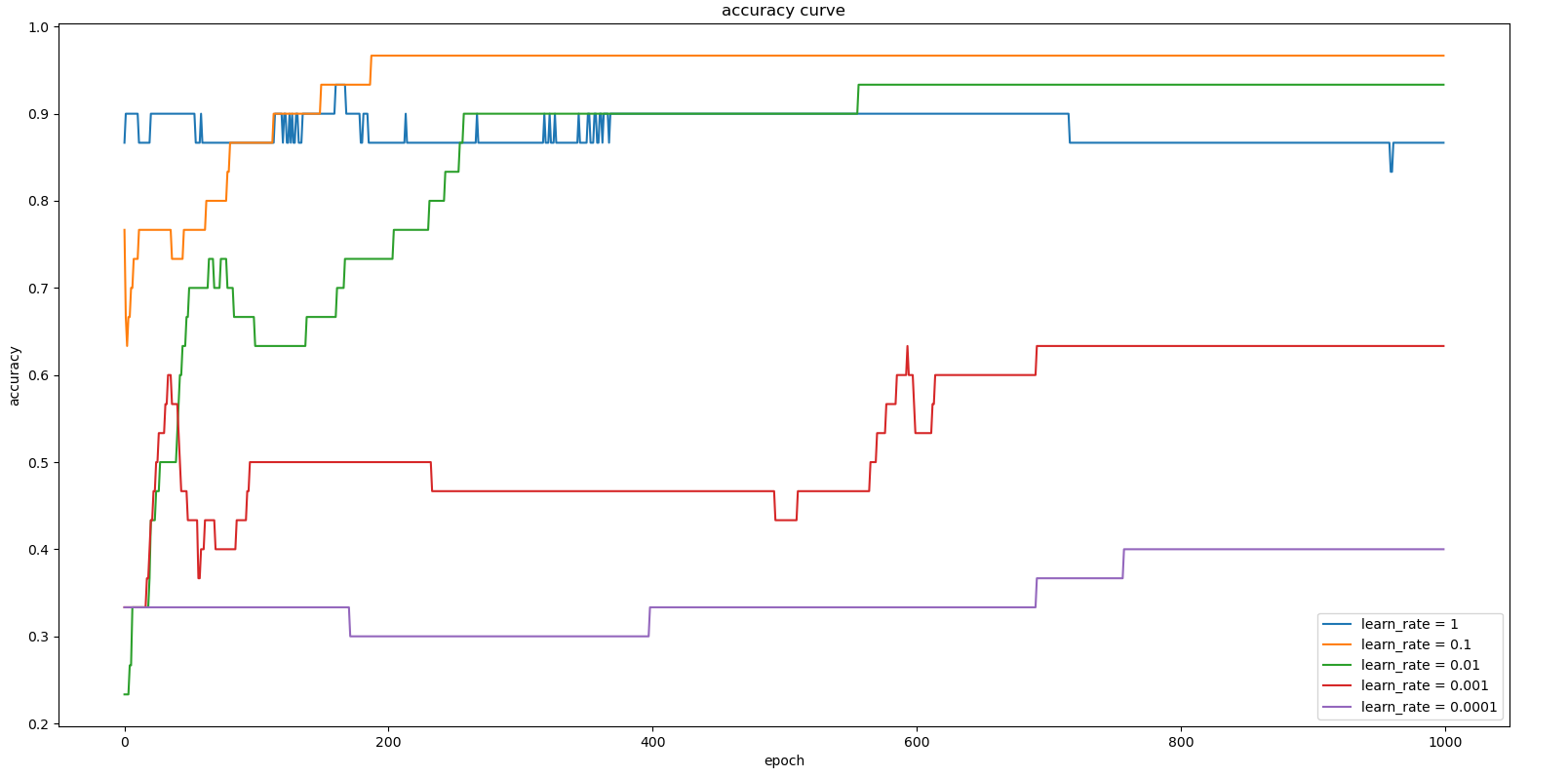
隐藏层节点较少时，网络的表达能力较差，隐藏层节点过多时，网络的表达能力小，容易过拟合。由下图可知，训练集数据上的分类准确率达到，很有可能是过拟合。

* 1. 观察不同的梯度更新步长对训练的影响，并给出一些描述或解释；

学习率较小时，参数更新较慢，学习效率很低。

学习率较大时，准确率可能卡主不动，甚至是变差。

因此合适的方法是最开始时学习率较大，快速学习；当准确率长时间没有进步时，逐渐减小准确率，以平衡性能和效率。



* 1. 在网络结构固定的情况下，绘制出目标函数随着迭代步数增加的变化曲线

超参数设定如下的情况下，准确率随着迭代步数的增加的变化曲线如图所示。

